Linear scaling second-order Moller–Plesset theory in the atomic orbital basis for large molecular systems

Philippe Y. Ayala and Gustavo E. Scuseria

Citation: The Journal of Chemical Physics **110**, 3660 (1999); doi: 10.1063/1.478256
View online: http://dx.doi.org/10.1063/1.478256
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/110/8?ver=pdfcov
Published by the AIP Publishing

Linear scaling second-order Moller–Plesset theory in the atomic orbital basis for large molecular systems

Philippe Y. Ayala and Gustavo E. Scuseria[a)]
Center for Nanoscale Science and Technology, Rice Quantum Institute, and Department of Chemistry, Mail Stop 60, Rice University, Houston, Texas 77005-1892

(Received 21 September 1998; accepted 16 November 1998)

We have used Almlöf and Häser's Laplace transform idea to eliminate the energy denominator in second-order perturbation theory (MP2) and obtain an energy expression in the atomic orbital basis. We show that the asymptotic computational cost of this method scales quadratically with molecular size. We then define atomic orbital domains such that selective pairwise interactions can be neglected using well-defined thresholding criteria based on the power law decay properties of the long-range contributions. For large molecules, our scheme yields linear scaling computational cost as a function of molecular size. The errors can be controlled in a precise manner and our method reproduces canonical MP2 energies. We present benchmark calculations of polyglycine chains and water clusters containing up to 3040 basis functions. © *1999 American Institute of Physics.* [S0021-9606(99)30108-2]

I. INTRODUCTION

Second-order Moller–Plesset (MP2) perturbation theory has been for many years a cornerstone of *ab initio* molecular orbital studies and is an important part of many reliable models of chemistry. Indeed, MP2 theory is one of the least computationally demanding ways of including electron correlation at the *ab initio* level. Despite being computationally less expensive than more sophisticated methods like coupled-cluster, MP2 in its conventional formulation requires the dedication of formidable computational resources for the study of large molecular systems. For the most part, the largest MP2 calculations carried out to date have been limited to molecules with high symmetry.[1–4] Recent advances in linear-scaling density functional theory (DFT) and Hartree–Fock (HF) methods[5–13] make this problem even more acute. In this paper, our objective is to develop an MP2 approach that makes it feasible to study very large molecular systems like biomolecules, i.e., molecules with hundreds of atoms.

Throughout this paper, we denote occupied molecular orbitals by i and j and virtual (or unoccupied) molecular orbitals by a and b. Atomic orbitals are denoted by greek letters. In the conventional approach (see Taylor[14] and Head-Gordon *et al.*[15] for algorithmic examples), the four-center two-electron integrals in the atomic orbital (AO) basis, $(\mu\nu|\lambda\sigma)$, are transformed into the canonical molecular orbital (MO) basis,

$$(ia|jb) = \int i(1)j(2)\frac{1}{r_{12}}a(1)b(2)dr_1dr_2$$

$$= \sum_{\mu\nu\lambda\sigma} (\mu\nu|\lambda\sigma)C_{\mu i}C_{\nu a}C_{\lambda j}C_{\sigma b}. \tag{1.1}$$

In the present work, we only consider the closed-shell case. The generalization to open-shell unrestricted MP2 is straightforward. The closed-shell MP2 energy is obtained from

$$E_2 = -\sum_{ij,ab}^{\text{occ,vir}} \frac{(ia|jb)[2(ia|jb)-(ib|ja)]}{\epsilon_a+\epsilon_b-\epsilon_i-\epsilon_j}, \tag{1.2}$$

where ϵ_i and ϵ_a are the eigenvalues of the occupied and virtual blocks of the Fock matrix, respectively.

Due to the delocalized nature of the canonical MOs, the computational effort involved in the integral transformation scales asymptotically as $\mathcal{O}(N^5)$ with system size. *Local correlation space*[16–22] and *resolution of the identity*[23–27] techniques have been proven in recent years to be much less computationally intensive as well as reliable alternatives to the canonical formulation. The correlation energy given by such techniques differ from the canonical MP2 energy by typically 1% to 2% for small and moderate basis set sizes. For larger basis sets, the deviation can be expected to be much smaller.[20,23,24]

In this paper, we show that the canonical MP2 energy can be reproduced within a given accuracy with computational work scaling linearly with system size. In order to achieve this goal, we have used Almlöf and Häser's seminal work on the Laplace MP2 ansatz[28–31] as a starting point. The Laplace MP2 method is briefly described in Secs. II and III and we refer the interested reader to the original papers for details. Our linear scaling MP2 method as well as benchmark calculations are presented in the remaining sections.

II. LAPLACE MP2

In 1991, Almlöf showed that the exact MP2 energy can be obtained using noncanonical MOs while retaining the simplicity of the conventional formulation.[28] He noted that the Laplace transform of the energy denominator in Eq. (1.2)

a)Electronic mail: guscus@katzo.rice.edu

yields an energy expression invariant with respect to unitary transformation of weighted molecular orbitals (indicated with primes in the equation below).

$$\begin{aligned} E_2 &= -\sum_{iajb} \int_0^\infty (ia|jb)[2(ia|jb)-(ib|ja)]e^{-\Delta_{iajb}t}dt \\ &= -\int_0^\infty \sum_{i'a'j'b'} (i'a'|j'b')[2(i'a'|j'b') \\ &\quad -(i'b'|j'a')]dt, \\ \Delta_{iajb} &= \epsilon_a+\epsilon_b-\epsilon_i-\epsilon_j, \\ i' &= ie^{\epsilon_i t/2}; \quad a' = ae^{-\epsilon_a t/2}. \end{aligned} \tag{2.1}$$

Just as important was showing that the t-integration can be accurately carried out by Gaussian quadrature. Almlöf and Häser showed that micro-Hartree accuracy can be obtained with 8–10 quadrature points and that milli-Hartree accuracy is obtained with only 3–5 quadrature points.[29] Given τ quadrature weights and abscissa $\{w_\alpha, t_\alpha\}$, an implementation of the Laplace-MP2 method thus consists in the trivial modification of an existing MP2 algorithm,

$$\begin{aligned} E_2 &= -\sum_\alpha^\tau w_\alpha e_2^\alpha, \\ e_2^\alpha &= \sum_{i^\alpha a^\alpha j^\alpha b^\alpha} (i^\alpha a^\alpha|j^\alpha b^\alpha)[2(i^\alpha a^\alpha|j^\alpha b^\alpha)-(i^\alpha b^\alpha|j^\alpha a^\alpha)], \\ i^\alpha &= ie^{\epsilon_i t_\alpha/2}, \quad a^\alpha = ae^{-\epsilon_a t_\alpha/2}. \end{aligned} \tag{2.2}$$

We also note that the Laplace MP2 energy is invariant if one uses

$$i^\alpha = ie^{(\epsilon_i-\epsilon_F)t_\alpha/2}, \quad a^\alpha = ae^{-(\epsilon_a-\epsilon_F)t_\alpha/2}. \tag{2.3}$$

By setting ϵ_F to a Fermi level between the highest occupied molecular orbital (HOMO) and lowest unocupied molecular orbital (LUMO) energies, the exponential weights in Eq. (2.3) are always smaller than unity, irrespective of the sign of the MO energy.

There are several ways to determine the quadrature parameters.[29] The most simple consist in a least-squares fit of the $1/x$ function over the interval of values spanned by Δ_{iajb}. Five quadrature points are usually required to fit $1/x$ with accuracy greater than three decimals. Häser and Almlöf showed that there is a strong correlation between the error in the fit and the error in the Laplace-MP2 energy.[29] The quality of the fit improves for systems with larger HOMO–LUMO gaps and smaller eigenvalue scales. For the polyglycines and water-clusters used as benchmarks in this paper, Δ_{iajb} typically ranges from 0.9 to 55 Hartrees. Because of their smaller range of Δ_{iajb}, frozen-core calculations require fewer quadrature points to achieve the same accuracy. Nevertheless, all electrons have been correlated in calculations throughout this work. Table I shows the energy deviation from the exact answer using the least-square Gaussian quadrature. It is worth noting that the error introduced by this quadrature is largely systematic when a minimum of five quadrature points are used. As shown in Table I, the quadrature error for $(H_2O)_{10}$ is roughly ten times that of H_2O for $\tau=5$.

TABLE I. Total energy errors (in Hartrees) from exact answer (obtained using GAUSSIAN) for the Laplace-MP2 method as a function of the discretization order τ.

		τ		
Molecule	Basis set	3	5	7
H_2O	3-21G	1.6×10^{-4}	3.1×10^{-6}	8.9×10^{-8}
	6-31G*	2.1×10^{-4}	1.4×10^{-5}	6.6×10^{-7}
$(H_2O)_{10}$	3-21G	3.9×10^{-3}	2.9×10^{-5}	6.2×0^{-7}
	6-31G*	3.8×10^{-3}	1.6×10^{-4}	1.5×10^{-6}
2-glycine	3-21G	4.6×10^{-4}	1.7×10^{-5}	7.6×10^{-7}
4-glycine	3-21G	7.7×10^{-4}	3.0×10^{-5}	3.7×10^{-6}

The accuracy of the Gaussian quadrature improves with the discretization order. However, a monotonic improvement is not assured with increasing τ. This is due to two reasons. First, the distribution of the Δ_{iajb} is not taken into account in the least-square fit and, second, the stiffness of the least-square problem determining the optimum quadrature parameters increases with τ. Equivalent solutions to the least-square problem may yield different energy deviations.

A systematic improvement in the Laplace-MP2 energy can be obtained using an Euler–McLaurin formula, very much like it is done in radial quadrature schemes for density functional theory[32]

$$\begin{aligned} E_2 &= \int_0^\infty e_2(t)dt \\ &= \int_0^1 e_2(t)\frac{dt}{dr}dr \\ &= \int_0^1 f_2(r)dr \\ &\approx \frac{1}{\tau+1}\left[\sum_{k=1}^{\tau} f_2\left(\frac{k}{\tau+1}\right)+\frac{1}{2}(f_2(0)+f_2(1))\right]. \end{aligned} \tag{2.4}$$

In principle, $e_2(t)$ and all higher derivatives are zero at $t=\infty$, $r=1$. Evaluating $e_2(t)$ is increasingly time consuming as t (and r) approaches 0, therefore, it is best to choose a change of variable for which its Jacobian and a number of its derivatives are zero for $r=0$. To insure that the nth derivative of $f_2(r)$ is zero for $r=0$, one can choose the following change of variable:

$$t = \frac{\sum_{k=n+2}^{K} a_k r^k}{(1-r)^m}. \tag{2.5}$$

For numerical stability around $r=1$, it is best to choose m as low as possible. Knowing the values taken by the derivatives of $f_2(r)$ at $r=0$ and $r=\infty$ can also help fitting unevenly spaced data points.

In the Euler–McLaurin scheme, the choice of change of variable is of course arbitrary and the quality of the integration will depend on it. After experimentation with various forms for Eq. (2.5), we have found that the least-square

based quadrature is superior to the Euler–McLaurin scheme for low discretization order (typically up to $\tau=5$). If an accuracy greater than five or six decimals is desired, our experience has been that the Euler–McLaurin scheme with eight or more energy points is preferable over the least-square quadrature.

III. AO-LAPLACE MP2

The atomic orbital representation that follows naturally from Almlöf's Laplace MP2 was first discussed, in great detail, by Häser.[30] Using the following transformations

$$\underline{\mu}^{\alpha}=\sum_{\nu} X^{\alpha}_{\mu\nu}\nu; \quad \overline{\mu}^{\alpha}=\sum_{\nu} Y^{\alpha}_{\mu\nu}\nu,$$
$$X^{\alpha}_{\mu\nu}=\sum_{i}^{\mathrm{occ}} C_{\mu i}C_{\nu i}e^{\epsilon_i t_\alpha}, \quad Y^{\alpha}_{\mu\nu}=\sum_{a}^{\mathrm{vir}} C_{\mu a}C_{\nu a}e^{-\epsilon_a t_\alpha}, \tag{3.1}$$

the energy expression can be written in a number of ways,[30] one of which is

$$e_2^{\alpha}=-\sum_{\mu\nu\lambda\sigma}(\underline{\mu}^{\alpha}\overline{\nu}^{\alpha}|\underline{\lambda}^{\alpha}\overline{\sigma}^{\alpha})[2(\mu\nu|\lambda\sigma)-(\mu\sigma|\lambda\nu)], \tag{3.2}$$

where

$$(\underline{\mu}^{\alpha}\overline{\nu}^{\alpha}|\underline{\lambda}^{\alpha}\overline{\sigma}^{\alpha})=\sum_{\gamma\delta\kappa\epsilon} X^{\alpha}_{\mu\gamma}Y^{\alpha}_{\nu\delta}(\gamma\delta|\kappa\epsilon)X^{\alpha}_{\kappa\lambda}Y^{\alpha}_{\epsilon\sigma}. \tag{3.3}$$

For t_α equal to zero, the $\boldsymbol{X}^{\alpha}$ and $\boldsymbol{Y}^{\alpha}$ transformation matrices are the Hartree–Fock density matrix ($\boldsymbol{P}$) and its complimentary projector ($\boldsymbol{Q}$). In this paper, we will loosely refer to $\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{X}^{\alpha}$ and $\boldsymbol{Y}^{\alpha}$ as ''density matrices.'' The Laplace density matrices ($\boldsymbol{X}^{\alpha}$ and $\boldsymbol{Y}^{\alpha}$) and the standard HF density matrices ($\boldsymbol{P}$ and $\boldsymbol{Q}$) have similar properties. For the standard density matrices, the orthogonality of the occupied and virtual spaces translates into

$$\boldsymbol{PSQ}=\boldsymbol{0}, \tag{3.4}$$

$$\boldsymbol{PS}+\boldsymbol{QS}=\boldsymbol{I}, \tag{3.5}$$

where $\boldsymbol{S}$ denotes the AO overlap matrix, and $\boldsymbol{I}$ the identity matrix. Similarly, for the Laplace density matrices, we have

$$\boldsymbol{X}^{\alpha}\boldsymbol{S}\boldsymbol{Y}^{\alpha}=\boldsymbol{0}, \tag{3.6}$$

$$\boldsymbol{X}^{\alpha}\boldsymbol{S}+\boldsymbol{Y}^{\alpha}\boldsymbol{S}=\boldsymbol{I}_{\mathbf{exp}}, \tag{3.7}$$

where $\boldsymbol{I}_{\mathbf{exp}}$ is a diagonal matrix with trace

$$\mathrm{Trace}[\boldsymbol{I}_{\mathbf{exp}}]=\sum_{i}^{\mathrm{occ}} e^{\epsilon_i t_\alpha}+\sum_{a}^{\mathrm{vir}} e^{-\epsilon_a t_\alpha}. \tag{3.8}$$

For large t_α, only molecular orbitals with an energy close to the Fermi-level contribute to the $\boldsymbol{X}^{\alpha}$ and $\boldsymbol{Y}^{\alpha}$ matrices. For example, a typical core-level MO energy for oxygen is approximately -20 Hartrees; thus, for t_α greater than 1.15 a.u., the exponential weight for this MO would then be 10^{-10} or less. This energy level is, therefore, progressively projected out with increasing t_α during the integration. The Laplace quadrature, in effect, correlates energy levels falling within a specific energy window that shrinks with increasing t. The valence energy levels contribute to each discrete Laplace quadrature point, whereas the deepest occupied energy levels contribute only to the energy points with the smaller t_α, typically 1 to 2 energy points. The absolute error in the quadrature is thus likely to be concentrated on the correlation of the core energy levels and can be expected to be fairly systematic.

Our implementation of the AO-Laplace MP2 makes use of the Schwarz based screening advocated by Häser.[30] Omitting the α superscript from now on, negligible contributions to the energy are screened out using four two-index quantities $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$

$$A_{\mu\nu}=|(\mu\nu|\mu\nu)|^{1/2}, \quad B_{\mu\nu}=|(\mu\nu|\underline{\mu}\nu)|^{1/2},$$
$$C_{\mu\nu}=|(\mu\overline{\nu}|\mu\overline{\nu})|^{1/2},$$
$$D_{\mu\nu}=\mathrm{Min}\left(\sum_{\sigma} B_{\mu\sigma}|Y_{\sigma\nu}|,\sum_{\sigma} C_{\mu\sigma}|X_{\sigma\nu}|\right)$$
$$\geqslant|(\underline{\mu}\overline{\nu}|\underline{\mu}\overline{\nu})|^{1/2}. \tag{3.9}$$

By virtue of the Schwarz inequality, we have, for example,

$$|(\mu\nu|\lambda\sigma)|\leqslant A_{\mu\nu}A_{\lambda\sigma}, \quad |(\underline{\mu}\overline{\nu}|\underline{\lambda}\overline{\sigma})|\leqslant D_{\mu\nu}D_{\lambda\sigma}. \tag{3.10}$$

The fact that the energy expression can take many forms besides Eq. (3.2) gives rise to a formidable screening protocol (see Häser[30]). Using $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$, one can *a priori* screen out partially transformed integrals on the basis of their contribution to the final MP2 energy. For example, given an energy contribution threshold, θ, the integral $(\mu\nu|\lambda\sigma)$ needed in the first quarter transformation will be generated if

$$A_{\mu\nu}A_{\lambda\sigma}(2D_{\mu\nu}D_{\lambda\sigma}+D_{\mu\sigma}D_{\lambda\nu})\geqslant\theta. \tag{3.11}$$

The $(\mu\nu|\lambda\sigma)$ integral will be discarded, irrespective of its magnitude, if

$$|(\mu\nu|\lambda\sigma)|(2D_{\mu\nu}D_{\lambda\sigma}+D_{\mu\sigma}D_{\lambda\nu})\leqslant\theta. \tag{3.12}$$

Our computational implementation of the AO-Laplace MP2 is based on a development version of the GAUSSIAN suite of programs.[33] We obtain the MP2 energy by performing all four quarter transformations of the AO integrals and then contracting the transformed integrals with the AO integrals as in Eq. (3.2) for each Laplace quadrature point. In our implementation of this multipass semidirect AO-Laplace MP2, we have chosen the following sequence of operations in order to enable the study of very large systems while limiting memory usage and book-keeping.

- Step 1
  - Generate $(\gamma\delta|\kappa\epsilon)$ needed for first quarter transformation.
  - Perform quarter transformation for a subset of $\underline{\mu}$.
  - Store $(\underline{\mu}\delta|\kappa\epsilon)$ to disk in appropriate order and free memory.
  - Generate $(\mu\nu|\lambda\sigma)$ needed for contraction.
  - Store $(\mu\nu|\lambda\sigma)$ to disk in appropriate order and free memory.
- Step 2
  - Retrieve $(\underline{\mu}\delta|\kappa\epsilon)$.
  - Perform three remaining quarter transformations.

Retrieve AO integrals and contract with $(\mu\bar{\nu}|\underline{\lambda}\bar{\sigma})$.

Even though each transformation is formally more computationally expensive in the AO basis than in the MO basis, Schwarz screening makes carrying the four quarter transformations in the AO basis rapidly competitive with state-of-the-art conventional MP2 algorithms, both in terms of CPU time and physical resource usage. For this reason, the AO-Laplace MP2 is the method of choice for the study of very large molecular systems. This point has even more value, of course, if the scaling of the computational work involved in the AO-Laplace MP2 method with respect to molecular size is significantly below that of the conventional MP2 in the MO basis. This issue is addressed in the next sections.

IV. QUADRATIC SCALING AO-LAPLACE MP2

We would like to start this section reminding the reader of some well-known results that are important to the discussion below. First, let us reiterate that the atomic orbital basis constitutes a localized basis set and the overlap between atomic orbitals decays exponentially with atomic separations. The radial overlap for Gaussian primitive functions is given by

$$S_{\mu\nu}\sim\left(\frac{\pi}{\zeta_\mu+\zeta_\nu}\right)^{3/2}\exp\left[-\frac{\zeta_\mu\zeta_\nu}{\zeta_\mu+\zeta_\nu}R_{\mu\nu}^2\right], \tag{4.1}$$

where ζ_μ and ζ_ν are orbital exponents and $R_{\mu\nu}$ is their center separation. For large enough systems, the overlap matrix has thus only $\mathcal{O}(N)$ significant elements. We also recall that, the prefactor of the four-center two-electron integral $(\mu\nu|\lambda\sigma)$ is bound, within a multiplicative constant, by $S_{\mu\nu}S_{\lambda\sigma}$. The number of significant two-electron integrals in the AO basis therefore grows asymptotically as $\mathcal{O}(N^2)$.[34–36]

One of the main frustrations in implementing and using conventional MP2 algorithms for large molecular systems is that even though the number of significant $(\mu\nu|\lambda\sigma)$ integrals grows as $\mathcal{O}(N^2)$, the number of significant $(ia|jb)$ transformed integrals grows as $\mathcal{O}(N^4)$ due to the delocalized nature of the canonical molecular orbitals.[31] Since the density matrices are invariant with respect to unitary transformation of the MOs, whether the MOs are localized or not is irrelevant in the AO-Laplace MP2. In the AO basis, attention should be paid to the decay behavior — that is, the sparsity — of $\boldsymbol{X}$ and $\boldsymbol{Y}$. This decay is specific to the molecule under study. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are sparse [i.e., if they have only $\mathcal{O}(N)$ significant elements], from Eq. (3.3) and the decay behavior of the AO integrals mentioned above,[34–36] it is clear that there are only $\mathcal{O}(N^2)$ surviving $(\mu\bar{\nu}|\underline{\lambda}\bar{\sigma})$ — namely $(\mu\bar{\mu}|\underline{\lambda}\bar{\lambda})$, yielding an $\mathcal{O}(N^2)$ Laplace MP2 algorithm.

Kohn conjectured that, for finite systems, the electron density decays exponentially with distance and that insulator-type systems exhibit the fastest decay.[37] Insulator-type molecular systems are precisely the systems for which MP2 gives better results as their HOMO–LUMO gap is relatively large. Calculations have shown (see Schwegler *et al.*[11] and Millam *et al.*[7] for examples) that HF and DFT density matrix elements can be approximately bound by an exponential. Of course, the same conclusion applies to $\boldsymbol{X}$. Although

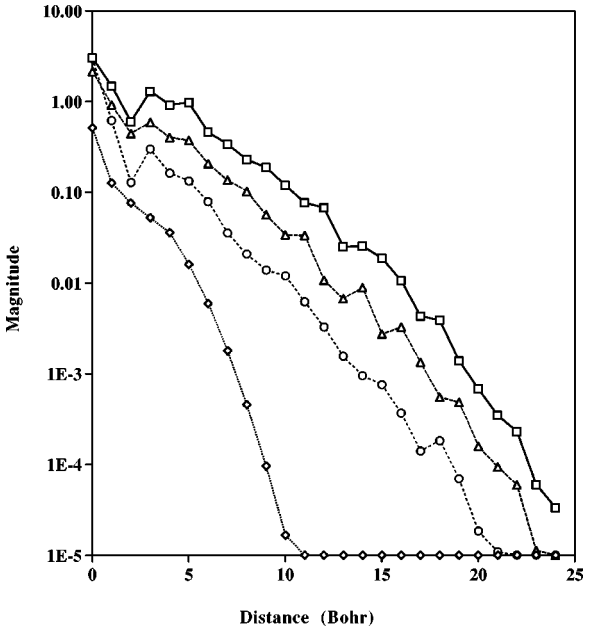

FIG. 1. Behavior of $A_{\mu\nu}$ (diamonds), $B_{\mu\nu}$ (circles), $C_{\mu\nu}$ (triangles), and $D_{\mu\nu}$ (squares), see text for definition, as a function of the μ,ν center separation for three-dimensional $(H_2O)_{30}$ using the 3-21G basis set.

in all likelihood very similar to that of $\boldsymbol{X}$,[30] the behavior of $\boldsymbol{Y}$ remains unclear. Nevertheless, we point out that by shifting the MO energy by a Fermi level as in Eq. (2.3), $\boldsymbol{X}$ and $\boldsymbol{Y}$ are always sparser than $\boldsymbol{P}$ and $\boldsymbol{Q}$, respectively.

As a matter of fact, one does not need to rely on the assumed sparsity of $\boldsymbol{Y}$ to show that the AO-Laplace MP2 method is inherently an asymptotically quadratic scaling method for molecules with relatively large HOMO–LUMO gaps. Thanks to the Schwarz inequality, Eq. (3.10), the quadratic scaling can be demonstrated by simply showing that $D_{\mu\nu}$ decays with increasing μ, ν separation. If $D_{\mu\nu}$ decays, then for a large enough system, $(\mu\bar{\nu}|\underline{\lambda}\bar{\sigma})$ will have only $\mathcal{O}(N^2)$ elements above threshold.

Given the relationship between $\boldsymbol{D}$ and $\boldsymbol{C}$ [Eq. (3.9)] along with the sparsity of $\boldsymbol{X}$, it is sufficient to show that $\boldsymbol{C}$ decays. The half-transformed integral from which $C_{\mu\nu}$ is constructed is given by

$$C_{\mu\nu}^2=|(\mu\bar{\nu}|\mu\bar{\nu})|=\left|\sum_{\gamma\delta}Y_{\nu\gamma}(\mu\gamma|\mu\delta)Y_{\delta\nu}\right|. \tag{4.2}$$

Because $(\mu\gamma|\mu\delta)$ constitutes a positive definite matrix, one can rigourously (within a multiplicative constant) bound the value of the half-transformed integral by

$$|(\mu\bar{\nu}|\mu\bar{\nu})|\leqslant\left|\sum_{\gamma\delta}Y_{\nu\gamma}S_{\mu\gamma}S_{\mu\delta}Y_{\delta\nu}\right|, \tag{4.3}$$

$$=(YS)_{\mu\nu}^2. \tag{4.4}$$

We know that $\boldsymbol{XS}$ is sparse, because $\boldsymbol{X}$ is sparse for systems with large HOMO–LUMO gaps and $\boldsymbol{S}$ is sparse for any sufficiently large system. Therefore, since $\boldsymbol{XS}+\boldsymbol{YS}$ is diagonal [see Eq. (3.6)], the decay properties of $\boldsymbol{YS}$ and, in turn, that of $\boldsymbol{C}$ and $\boldsymbol{D}$ are guaranteed. Figure 1 shows the decay of $\boldsymbol{A}$, $\boldsymbol{B}$,

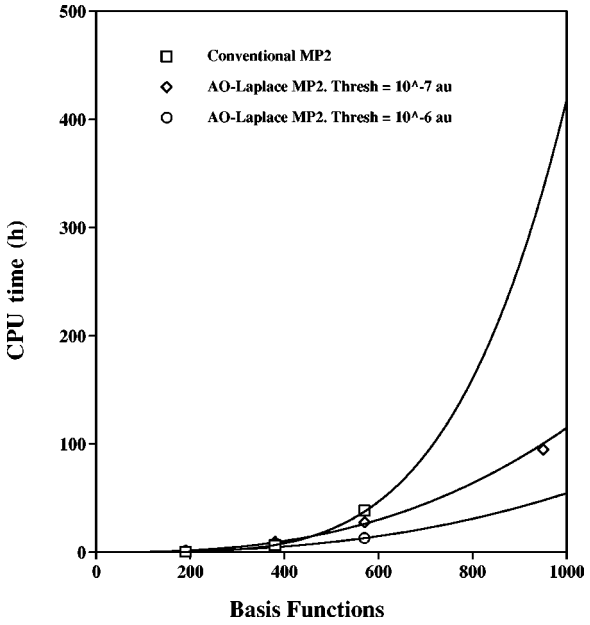

FIG. 2. CPU times (Ref. 38) for MP2 energy calculations of three-dimensional water clusters using the 6-31G* basis set. Five-point quadrature used in the AO-Laplace MP2 method.

C, and ***D*** for a three-dimensional cluster containing 30 water molecules.

Having established that, in the asymptotic regime, only the diagonal transformed integrals $(\mu\bar{\mu}|\underline{\lambda}\bar{\lambda})$ are significant, it can be immediately seen that the surviving exchange contribution to the MP2 energy is given by

$$(\mu\bar{\mu}|\underline{\lambda}\bar{\lambda})(\mu\lambda|\lambda\mu)\approx(\mu\bar{\mu}|\underline{\lambda}\bar{\lambda})\frac{S_{\mu\lambda}S_{\mu\lambda}}{R_{\mu\lambda}}$$

$$=(\mu\bar{\mu}|\underline{\lambda}\bar{\lambda})\frac{e^{-2\zeta R_{\mu\lambda}^2}}{R_{\mu\lambda}}, \qquad (4.5)$$

and that the surviving Coulomb contribution is given by $(\mu\bar{\mu}|\underline{\lambda}\bar{\lambda})(\mu\mu|\lambda\lambda)$.

Clearly, the exchange contribution to the MP2 energy decays exponentially whereas the Coulomb contribution decays as a power law. Based on the above arguments, we conclude that the long-range MP2 contributions are solely due to $(\mu\bar{\mu}|\underline{\lambda}\bar{\lambda})(\mu\mu|\lambda\lambda)$. In practice, we observe that the exchange contributions vanish for μ, λ separations greater than ~10–15 Bohrs and that significant Coulomb contributions are still present for distances of ~30 Bohrs.

Although never before clearly pointed out, these arguments demonstrate that the AO-MP2 method is inherently a quadratic scaling method for molecules with relatively large HOMO–LUMO gaps and that a quadratic scaling MP2 algorithm can be obtained using Häser's screening protocol. Figure 2 shows CPU times[38] for three-dimensional water clusters using the 6-31G* basis set. Three-dimensional water clusters are relatively compact systems with a fairly broad magnitude distribution for the matrix elements of ***X***, ***Y***, ***A***, ***B***, ***C***, and ***D***. They are actually challenging cases for the AO-Laplace MP2. Using the 6-31G* basis set and a threshold of $\theta=10^{-7}$ a.u. for screening energy contributions [see Eqs. (3.11) and (3.12)], the crossover with the conventional MO algorithm occurs for $(H_2O)_{30}$. For certain applications, a looser threshold could be appropriate, and in such instances, the crossover occurs for much smaller systems [for example, $(H_2O)_{20}$ for $\theta=10^{-6}$ a.u. in Fig. 2]. A similar situation occurs for nonpolarized basis sets where we obtain earlier crossovers. The basis set size essentially affects the prefactor of the computational cost of the Laplace MP2 method but does not affect its asymptotic scaling. For the moderate size water-clusters shown in Fig. 2, the AO-Laplace MP2 already scales asymptotically as $N^{2.6}$.

The AO-Laplace MP2 becomes faster and less demanding than conventional semi-direct MP2 with increasing system size. For $(H_2O)_{50}$ using the 6-31G* basis set, the CPU time for the conventional MP2 algorithm is estimated to be in excess of two weeks,[38] whereas it takes a maximum of 1.2 days per AO-Laplace MP2 quadrature point in the AO basis using our current implementation of Häser screening protocol. Similarly, the disk requirements are much less demanding in the AO-Laplace MP2 than in the conventional MO algorithm.

The quadratic limit for the count of significant AO integrals is reached for different size systems depending on the neglect threshold and the dimensionality of the system. Reaching this limit is a necessary condition for the AO-Laplace MP2 algorithm to start exhibiting quadratic scaling character. This is to be put in contrast with local-correlation space MP2 methods for which near quadratic scaling is reported for small and medium size molecules thanks to the *a priori* neglect of certain excitations.[18,21]

V. ACCURACY

In order to isolate the error due to Häser's screening protocol in our implementation from that due to the Laplace quadrature, we compare the Laplace AO-MP2 energy with the Laplace MO energy. As mentioned earlier, the implementation of the Laplace MP2 method in the MO basis consists of a trivial modification of any existing MP2 algorithm. In our Laplace MO-MP2 code, the only screening taking place is the screening of insignificant $(\mu\nu|\lambda\sigma)$ integrals on the basis of their prefactor (accuracy threshold on integral evaluation set to 10^{-10} a.u.). Unless mentioned otherwise, the results presented in this work correspond to the quadrature point with the smallest t_α of a five-point Laplace quadrature. For both water clusters and polyglycine chains, the first point parameters are approximately $t_\alpha=0.02$ Hartree^{-1} and $w_\alpha=0.05$ Hartree^{-1}. In every respect (CPU timings, resource usage and accuracy), this one calculation is representative of the five to seven quadrature points needed to obtain the AO-Laplace MP2 energy with micro-Hartree accuracy. In fact, due to the increasing sparsity of ***X*** and ***Y*** with increasing t_α, the energy calculation with the smallest t_α is the most computationally demanding, and that is why we selected it for benchmark purposes. An estimate of the total CPU time for the total Laplace MP2 calculation can be obtained by applying a factor of ~3.5–5 to the CPU times for the quadrature point with the smallest t_α.

Table II shows the error induced by the Schwarz-based screening for $\theta=10^{-7}/w_\alpha$, 10^{-7} and 10^{-9} a.u. Because the

TABLE II. Energy errors (in Hartrees) from exact answer and CPU times (Ref. 38) (in min) for integral generation and transformation steps (see Sec. III) for one AO-Laplace MP2 quadrature point (smallest t_α) as a function of the neglect threshold θ.

Molecule	Basis set	θ	ΔE	Step 1	Step 2
$(H_2O)_{20}$[a]	3-21G	10^{-9}	1.0×10^{-7}	38.3	41.2
		10^{-7}	4.2×10^{-6}	18.6	13.3
		$10^{-7}/w$	5.2×10^{-6}	14.1	11.7
	6-31G*	$10^{-8}/w$	2.0×10^{-6}	146.2	127.2
		$10^{-7}/w$	7.2×10^{-6}	96.7	64.4
$(H_2O)_{30}$[a]	3-21G	10^{-9}	1.2×10^{-7}	146.0	104.3
		10^{-7}	4.1×10^{-6}	63.6	39.2
		$10^{-7}/w$	7.0×10^{-6}	39.7	33.2
	6-31G*	$10^{-7}/w$	2.5×10^{-5}	276.8	196.9
$(H_2O)_{20}$[b]	6-31G*	10^{-7}	9.0×10^{-7}	69.2	21.4
$(H_2O)_{50}$[b]	3-21G	10^{-7}	4.3×10^{-6}	48.4	40.7
8-glycine	3-21G	10^{-7}	2.9×10^{-6}	58.2	81.1
		$10^{-7}/w$	4.1×10^{-6}	50.1	66.1
	6-31G*	$10^{-7}/w$	1.6×10^{-5}	424.1	429.1
12-glycine	3-21G	10^{-7}	4.8×10^{-6}	158.2	165.1
		$10^{-7}/w$	5.9×10^{-6}	118.2	161.0

[a]Three-dimensional water cluster.
[b]Two-dimensional water cluster.

quadrature weights can span up to 2 orders of magnitude, it is appropriate to take them into account in the neglect threshold. For the smallest t_α, the neglect threshold $\theta=10^{-7}/w_\alpha$ is effectively $\sim\theta=210^{-6}$ a.u. There is a sharp constrast in the timings[38] depending on whether $\theta=10^{-7}/w_\alpha$ or $\theta=10^{-9}$ a.u. is used. Thanks to the strong correlation between the maximum element magnitude and the center separation for all four $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ matrices, the insignificant $\mathcal{O}(N^4)$ contributions can be efficiently screened out by discrimination over either relative spatial position or decreasing magnitude.

To close this section, we observe that reducing the scaling of the MP2 calculation from $\mathcal{O}(N^5)$ down to $\mathcal{O}(N^2)$ has many concepts and practical issues in common with obtaining the Hartree–Fock exchange energy with $\mathcal{O}(N)$ effort. The types of errors induced by the Schwarz screening in the AO-Laplace MP2 are also very similar to the ones in linear-scaling HF-exchange methods.[9–12]

VI. LONG-RANGE CORRELATION

We next analyze the nature of the interactions that survive in large molecules in the $\mathcal{O}(N^2)$ asymptote of the AO-Laplace MP2 method. For distant molecular fragments the interaction energy should decay as $1/R^6$, however, a slower decay could be expected for long-range intramolecular interactions in molecules. This slower decay may be rationalized considering the interaction energy of two fragments, A and B, separated by a distance R along the z axis. The Taylor expansion of the $1/r_{12}$ potential is

$$\frac{1}{r_{12}}=\frac{1}{R}+\frac{z_1-z_2}{R^2}+\frac{2z_1^2-x_1^2-y_1^2+2z_2^2-x_2^2-y_2^2}{2R^3}+\frac{2x_1x_2+2y_1y_2-z_1z_2}{2R^3}+\mathcal{O}\left(\frac{1}{R^4}\right). \tag{6.1}$$

Using Eq. (6.1), the two-electron integral becomes for $\mu,\ \nu\in A$ and $\lambda,\ \sigma\in B$

$$(\mu\nu|\lambda\sigma)=\frac{(\mu|\nu)(\lambda|\sigma)}{R}+\frac{(\mu|z_1|\nu)(\lambda|\sigma)-(\lambda|z_2|\sigma)(\mu|\nu)}{R^2}+\mathcal{O}\left(\frac{1}{R^3}\right). \tag{6.2}$$

The orthogonality constraint between the occupied and virtual spaces is given by [cf. Eq. (3.6)]

$$(\underline{\mu}|\overline{\nu})=\sum_{all\ \gamma\delta}X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\nu}=0. \tag{6.3}$$

In the distant molecular fragments model, the transformed AOs will be localized on each fragment. This implies that, for $\underline{\mu},\overline{\nu}\in A$, the orthogonality constraint is satisfied by only considering atomic orbitals located on A, namely

$$(\underline{\mu}|\overline{\nu})=\sum_{\gamma\delta\in A}X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\nu}=0,\qquad X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\nu}=0\quad\text{for}\quad\gamma\delta\notin A. \tag{6.4}$$

Therefore, the transformed AO integral $(\underline{\mu}\,\overline{\nu}|\underline{\lambda}\,\overline{\sigma})$ can be obtained by limiting the $\gamma\delta\kappa\sigma$ summation in Eq. (3.3) to $\gamma\delta\in A$ and $\kappa\epsilon\in B$. We have then

$$(\underline{\mu}\,\overline{\nu}|\underline{\lambda}\,\overline{\sigma})=\frac{(\underline{\mu}|\overline{\nu})(\underline{\lambda}|\overline{\sigma})}{R}+\frac{(\underline{\mu}|z_1|\overline{\nu})(\underline{\lambda}|\overline{\sigma})-(\underline{\lambda}|z_2|\overline{\sigma})(\underline{\mu}|\overline{\nu})}{R^2}+\mathcal{O}\left(\frac{1}{R^3}\right). \tag{6.5}$$

Due to the orthogonality constraint, the only surviving term in Eq. (6.5) is the $\mathcal{O}(1/R^3)$ term. The individual contributions to the interaction energy are obtained by contracting Eq. (6.2) with Eq. (6.5). Thus, the long-range Coulomb contribution to the MP2 energy is the collection of $\mathcal{O}(1/R^4)$ and higher order terms.[39]

$$\begin{aligned}(\mu\nu|\lambda\sigma)(\underline{\mu}\,\overline{\nu}|\underline{\lambda}\,\overline{\sigma})=&\sum_{\gamma\delta\in A}\sum_{\kappa\epsilon\in B}S_{\mu\nu}X_{\mu\gamma}Y_{\nu\delta}\frac{M_{\gamma\delta\kappa\epsilon}}{R^4}X_{\lambda\kappa}Y_{\sigma\epsilon}S_{\lambda\sigma}\\&+(\mu|z_1|\nu)X_{\mu\gamma}Y_{\nu\delta}\frac{M_{\gamma\delta\kappa\epsilon}}{R^5}X_{\lambda\kappa}Y_{\sigma\epsilon}S_{\lambda\sigma}\\&-S_{\mu\nu}X_{\mu\gamma}Y_{\nu\delta}\frac{M_{\gamma\delta\kappa\epsilon}}{R^5}X_{\lambda\kappa}Y_{\sigma\epsilon}(\lambda|z_2|\sigma)\\&+X_{\mu\gamma}Y_{\nu\delta}\frac{M_{\mu\nu\lambda\sigma}M_{\gamma\delta\kappa\epsilon}}{R^6}X_{\lambda\kappa}Y_{\sigma\epsilon}+\mathcal{O}\left(\frac{1}{R^7}\right),\end{aligned} \tag{6.6}$$

where

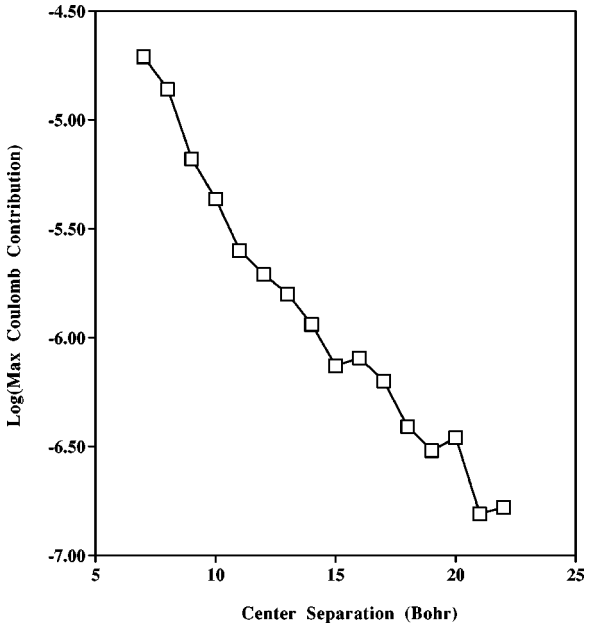

FIG. 3. Decay of $Max(\mu\bar{\nu}|\lambda\bar{\sigma})(\mu\nu|\lambda\sigma)$ as a function of the μ,λ center separation for three-dimensional $(H_2O)_{30}$ using the 3-21G basis set.

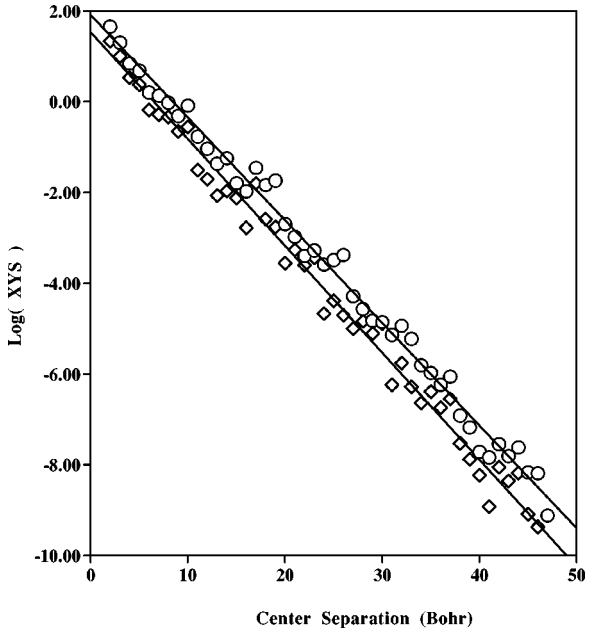

FIG. 4. Decay of $XY_{\mu p}S_p$ as a function of the μ, p distance for 8-glycine (see text for notation). Diamonds: 3-21G basis set; circles: 6-31G* basis set.

$$M_{\gamma\delta\kappa\epsilon}=(\gamma|x_1|\delta)(\kappa|x_2|\epsilon)+(\gamma|y_1|\delta)(\kappa|y_2|\epsilon)-\tfrac{1}{2}(\gamma|z_1|\delta)(\kappa|z_2|\epsilon).$$

The interaction energy is obtained by carrying out the summation over μ, $\nu\in A$ and λ, $\sigma\in B$. Thanks again to Eq. (6.4), the $1/R^4$ and $1/R^5$ terms add to zero giving rise to the familiar $1/R^6$ decay of the interaction energy.

In order to show that this approximate model is representative of the situation in large molecules, the behavior of $(\mu\bar{\nu}|\lambda\bar{\sigma})(\mu\nu|\lambda\sigma)$ as a function of the μ,λ center separation is shown in Fig. 3. In this figure, we have plotted the maximum magnitude of $(\mu\bar{\nu}|\lambda\bar{\sigma})(\mu\nu|\lambda\sigma)$ over all μ, ν, λ, σ as a function of the μ, λ distance. In agreement with the above discussion, the maximum Coulomb type contribution to the correlation energy decays as $\sim 1/R_{\mu\lambda}^4$ (exponent of the power fit: -4.12).

VII. LINEAR SCALING AO-LAPLACE MP2

Continuing our discussion on the quadratic scaling character of the AO-Laplace MP2 method and on the interaction energy of distant molecular fragments, we show below that linear scaling can be achieved by introducing interaction domains and neglecting selective domain–domain interactions.

For each atomic orbital μ, we define an interaction domain determined by a sphere centered on μ. The long-range contribution of the μ, λ pair, $(\mu\bar{\mu}|\lambda\bar{\lambda})(\mu\mu|\lambda\lambda)$, will be accumulated only if the edges of the μ and λ domains are within WS Bohrs from each other, that is, if the domains are not well separated. The interaction domain of μ, $\mathcal{D}(\mu)$, is defined as the sphere containing the collection of charge distributions $\gamma\delta$ that verify

$$(X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\mu})^2\geqslant\epsilon, \quad \text{for} \quad \gamma\delta\in\mathcal{D}(\mu). \tag{7.1}$$

The radius of the sphere is given by the maximum distance between μ and the $\gamma\delta$ charges within its domain. This definition for the interaction domain is validated by the results obtained in Sec. IV where we have shown that the matrix elements of $\boldsymbol{YS}$ decay with increasing center separation for molecules with large HOMO–LUMO gaps. Because the matrix elements of $\boldsymbol{X}$ and $\boldsymbol{S}$ also decay with distance, it is clear that the domain radius increases with decreasing ϵ. We have examined the decay of $X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\mu}$ for water clusters and polyglycine chains and found that it is similar for both systems. The decay is rapid, exponential in fact. An example is given in Fig. 4. Introducing here the notation that we use later in Sec. VII A, we denote by p the charge distribution $\gamma\delta$ and by $XY_{\mu p}S_p$ the quantity $X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\mu}$. In Fig. 4, the maximum magnitude of $(XY_{\mu p}S_p)$ as a function of the μ, p distance for 8-glycine is shown for two basis sets, namely, the double-zeta 3-21G basis set and the polarized 6-31G* basis set. The decay rate is similar in both cases, however, the prefactor of the exponential decay is larger in the 6-31G* case.

Because $X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\mu}$ decays, it is possible to bound (see below)

$$\left|\sum_{\gamma\delta\,\in\,\mathcal{D}(\mu)} X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\mu}\right|\leqslant\epsilon_2\sim\sqrt{\epsilon}, \tag{7.2}$$

and therefore, insure that the orthogonality constraint is satisfied within ϵ_2. Equations (7.1) and (7.2) put together thus constitute the intramolecular equivalent to Eq. (6.4) used in Sec. VI. In other words, ϵ yields a measure of how close from completeness the domain is, or, how good the distant molecular fragments model is. To a certain extent, our approach shares the philosophy involved in what Saebo, Pulay, and others proposed in their local-correlation space methods;[16,17] however, our objective of reproducing the canonical MP2 energy within a given accuracy remains un-

compromised. We argue that the total correlation energy obtained by our approach can be made reliably accurate by having ϵ suitably small and WS, the threshold distance for two domains to be well separated, suitably large. In Sec. VII A, we use the decay of $X_{\mu\gamma}S_{\gamma\delta}Y_{\delta\mu}$ to show that the neglected contributions are bound and that the ϵ and WS parameters can be automatically determined so as to preserve accuracy.

Our long-range contribution thresholding yields a linear scaling MP2 algorithm that enables the study of very large molecular systems with several hundreds of atoms. However, the study of such systems raises the issue of obtaining the $\boldsymbol{X}^{\alpha}$ and $\boldsymbol{Y}^{\alpha}$ transformation matrices [cf. Eq. (3.1)] with linear-scaling effort in cases where diagonalizing the Fock matrix to obtain the MO energies and coefficients becomes an $\mathcal{O}(N^3)$ bottleneck.[7,41–43] This issue is addressed in Sec. VII B.

A. Thresholding

The transformed integral $(\underline{\mu}\overline{\mu}|\underline{\lambda}\overline{\lambda})$ can be written as

$$(\underline{\mu}\overline{\mu}|\underline{\lambda}\overline{\lambda})=\sum_{\nu\sigma\gamma\delta} X_{\mu\nu}Y_{\mu\sigma}(\nu\sigma|\gamma\delta)X_{\lambda\gamma}Y_{\lambda\delta}. \tag{7.3}$$

As usual in the case of spherically symmetric Gaussian functions, the two-electron four-center integral $(\nu\sigma|\gamma\delta)$ can be expressed in terms of two charge distributions p and q lying in between the ν and σ centers and the γ and δ centers, respectively. Once again, it is convenient to express the prefactor of the integral in terms of the overlap matrix elements.

$$(\underline{\mu}\overline{\mu}|\underline{\lambda}\overline{\lambda})=\sum_{pq} XY_{\mu p}S_p \times\frac{\mathrm{Erf}(\sqrt{(\zeta_p\zeta_q)/(\zeta_p+\zeta_q)}R_{pq})}{R_{pq}}S_qXY_{\lambda q}, \tag{7.4}$$

where

$$S_p=S_{\nu\sigma},\quad S_q=S_{\gamma\delta},\qquad XY_{\mu p}=X_{\mu\nu}Y_{\mu\sigma};\quad XY_{\lambda q}=X_{\lambda\gamma}Y_{\lambda\delta}, \tag{7.5}$$

and ζ_p and ζ_q are Gaussian exponents for the charge distributions p and q, respectively. Using this notation, the orthogonality of the occupied and virtual spaces, Eq. (3.6), yields for all μ

$$\sum_p XY_{\mu p}S_p=0. \tag{7.6}$$

In our implementation, we are using a slight variation of Eq. (7.1) to define the domains. For each μ, we define the domain, $\mathcal{D}(\mu)$, by the subset of charge distributions $\{p\}$ that verify

$$(XY_{\mu p}S_p)^2\frac{\mathrm{Erf}(\sqrt{(\zeta_p)/2}R_{\mu p})}{R_{\mu p}}>\epsilon. \tag{7.7}$$

Provided that the domain is large enough to reliably have

$$\frac{\mathrm{Erf}(\sqrt{(\zeta_p)/2}R_{\mu p})}{R_{\mu p}}\approx\frac{1}{R_{\mu p}}\quad\text{for } p\notin\mathcal{D}(\mu), \tag{7.8}$$

we define the radius of the domain as R_0^{μ}, the largest $R_{\mu p}$ that verify Eq. (7.7)

Here, we show that it is possible to neglect domain–domain interactions in a consistent manner. Given ϵ, the selection threshold for the interaction domain, and θ, the threshold for energy contribution used for the Schwarz screening [see Eqs. (3.11) and (3.12)], we establish conditions on the domain well-separated parameter, WS. To do that, we need to consider the various cases resulting from the relative positions of μ, λ, p, and q.

In the case where the charge distributions p and q fall outside the domains of μ and λ, we have the following result:

$$XY_{\mu p}S_p\frac{\mathrm{Erf}(\sqrt{(\zeta_p\zeta_q)/(\zeta_p+\zeta_q)}R_{pq})}{R_{pq}}S_qXY_{\lambda q}(\mu\mu|\lambda\lambda) \leqslant\epsilon\sqrt{R_{\mu p}R_{\lambda q}}\frac{\mathrm{Erf}(\sqrt{(\zeta_p\zeta_q)/(\zeta_p+\zeta_q)}R_{pq})}{R_{pq}}(\mu\mu|\lambda\lambda),\quad\text{for } p,q\notin\mathcal{D}(\mu)\quad\text{and}\quad\mathcal{D}(\lambda). \tag{7.9}$$

Ignoring the prefactor of $(\mu\mu|\lambda\lambda)$, the right-hand side of Eq. (7.9) will take its maximum value for $p=q$. In such instance, we have then

$$\epsilon\sqrt{R_{\mu p}R_{\lambda q}}\frac{\mathrm{Erf}(\sqrt{(\zeta_p\zeta_q)/(\zeta_p+\zeta_q)}R_{pq})}{R_{pq}}(\mu\mu|\lambda\lambda) \leqslant\epsilon\frac{\sqrt{R_{\mu p}R_{\lambda q}}}{R_{\mu p}+R_{\lambda q}}\leqslant\frac{\epsilon}{2}. \tag{7.10}$$

For p and q such that each belong to one of the two domains, i.e., a dispersion-type contribution, we have

$$\sum_{pq} XY_{\mu p}S_p\frac{\mathrm{Erf}(\sqrt{(\zeta_p\zeta_q)/(\zeta_p+\zeta_q)}R_{pq})}{R_{pq}}S_qXY_{\lambda q}(\mu\mu|\lambda\lambda) \approx\sum_{p,q} XY_{\mu p}S_p\frac{1}{WS}S_qXY_{\lambda q}(\mu\mu|\lambda\lambda),\quad\text{for } p\in\mathcal{D}(\mu)\quad\text{and}\quad q\in\mathcal{D}(\lambda). \tag{7.11}$$

The orthogonality of the occupied and virtual spaces, Eq. (7.6), yields

$$\left|\sum_{p\in\mathcal{D}(\mu)} XY_{\mu p}S_p\right|=\left|\sum_{p\notin\mathcal{D}(\mu)} XY_{\mu p}S_p\right|=\epsilon_2. \tag{7.12}$$

We can introduce, here, the exponential decay $|XY_{\mu p}S_p|=A_{\mu}e^{-\tau_{\mu}R_{\mu p}}$. By doing so, a loose bound for ϵ_2 can be obtained by integrating over space. For a two-dimensional system, we have

$$\left|\sum_{p\notin\mathcal{D}(\mu)} XY_{\mu p}S_p\right|=\epsilon_2\leqslant A_{\mu}e^{-\tau_{\mu}R_0^{\mu}}\int_0^{\infty}2\pi re^{-\tau_{\mu}r}dr. \tag{7.13}$$

It is now possible to establish a relationship between the domain threshold, ϵ, and ϵ_2.

$$\epsilon_2 \leqslant \frac{2\pi}{\tau_\mu^2}\sqrt{\epsilon R_0^\mu}. \tag{7.14}$$

This relationship also makes it possible to establish a requirement on WS by combining Eq. (7.11) with Eq. (7.13). It is sufficient for WS to verify the following condition in order to insure that the sum of the screened-out contributions falls below θ

$$WS \geqslant \frac{4\pi^2}{\tau_\mu^2 \tau_\lambda^2}\frac{\epsilon}{\theta}\frac{\sqrt{R_0^\mu R_0^\lambda}}{R_0^\mu + R_0^\lambda}. \tag{7.15}$$

This result can easily be generalized to one and three-dimensional systems.

The last case that needs to be addressed is the instance where p belongs to the domain of μ and q does not belong to the domain of λ. It is sufficient to focus on the case where q is just outside the domain of λ. For this ionic-type contribution we have

$$\begin{aligned} XY_{\mu p} S_p \frac{\mathrm{Erf}(\sqrt{(\zeta_p\zeta_q)/(\zeta_p+\zeta_q)}R_{pq})}{R_{pq}} S_q XY_{\lambda q}(\mu\mu|\lambda\lambda) \\ \approx A_\mu e^{-\tau_\mu R_{\mu p}} A_\lambda e^{-\tau_\lambda R_0^\lambda}\frac{1}{R_{pq}}\frac{1}{R_{\mu\lambda}} \\ \leqslant A_\mu e^{-\tau_\mu R_{\mu p}} A_\lambda e^{-\tau_\lambda R_0^\lambda}\frac{1}{WS}\frac{1}{R_0^\lambda + R_0^\mu} \\ = A_\mu e^{-\tau_\mu R_{\mu p}}\sqrt{\epsilon R_0^\lambda}\frac{1}{WS}\frac{1}{R_0^\lambda + R_0^\mu}. \end{aligned} \tag{7.16}$$

Using Eqs. (7.12) and (7.13), the total contribution for $p \in \mathcal{D}(\mu)$ can be bound,

$$\sum_{p\in\mathcal{D}(\mu)} A_\mu e^{-\tau_\mu R_{\mu p}}\sqrt{\epsilon R_0^\lambda}\frac{1}{WS}\frac{1}{R_0^\lambda + R_0^\mu} \leqslant \frac{2\pi}{\tau_\mu^2}\frac{\epsilon}{WS}\frac{\sqrt{R_0^\mu R_0^\lambda}}{R_0^\mu + R_0^\lambda}. \tag{7.17}$$

This yields the following condition on WS:

$$WS \geqslant \frac{2\pi}{\tau_\mu^2}\frac{\epsilon}{\theta}\frac{\sqrt{R_0^\mu R_0^\lambda}}{R_0^\mu + R_0^\lambda}. \tag{7.18}$$

Therefore, provided that WS verifies Eqs. (7.13) and (7.18), one can safely neglect all $(\mu\overline{\mu}|\underline{\lambda}\overline{\lambda})(\mu\mu|\lambda\lambda)$ that verify

$$R_{\mu\lambda} - R_0^\mu - R_0^\lambda \geqslant WS. \tag{7.19}$$

Conversely, the only $\gamma\delta$ (charge distribution p) and $\kappa\epsilon$ (charge distribution q) that are needed to generate $(\mu\overline{\mu}|\underline{\lambda}\overline{\lambda})$ are the ones that verify

$$R_{pq} \leqslant R_{\mu\lambda} + R_0^\mu + R_0^\lambda, \tag{7.20}$$

$$R_{\mu p} + R_{\lambda q} \leqslant R_0^\mu + R_0^\lambda + WS. \tag{7.21}$$

For the polyglycines and two-dimensional water clusters, the requirements on WS are met with $WS \geqslant 1$ Bohr for $\epsilon = 10^{-10}$ a.u. and $\theta = 10^{-7}$ a.u. Water clusters feature smaller domains than polyglycines do, as it would have been expected for a largely nonbonded system. One of the main concerns for the local-MP2 method is making sure that the assigned domains span enough atomic centers to properly describe all types of bonding, irrespective of the basis set used.[17,21] Because we define domains according to energy contributions, this concern is clearly absent from our thresholding method.

TABLE III. Energy errors (in μHartrees) and percentage of μ,λ pairs ignored due to long-range correlation thresholding as a function of the well-separated distance threshold, WS. All calculations performed using the 3-21G basis-set, $\theta = 10^{-7}$ a.u., $\epsilon = 10^{-10}$ a.u.

	ΔE		Pairs skipped (%)	
WS (Bohrs)	$(H_2O)_{50}$	12-glycine	$(H_2O)_{50}$	12-glycine
2	0.21	0.05	65	28
3	0.39	0.04	62	26
5	0.47	0.08	56	23
7	0.39	0.01	51	22
15	0.12	0.00	30	15
∞	0.0	0.0	0	0

Table III shows the error induced by deliberately screening out the long-range contributions. One can see that this added screening contributes very little to the error. The oscillations with respect to WS are only the reflection of the oscillating nature of $(\mu\overline{\mu}|\underline{\lambda}\overline{\lambda})(\mu\mu|\lambda\lambda)$. Nevertheless, the error induced is remarkably stable with respect to varying WS. Less than one micro-Hartree for $WS = 2$ Bohrs in the case of a two-dimensional $(H_2O)_{50}$ water cluster. In fact, the error is insignificant when compared to the error induced by the Schwarz screening (4.3 micro-Hartrees for $\theta = 10^{-7}$ Hartree). For this moderately big water cluster, 65% of the μ,λ pair interactions are screened out for $WS = 2$ Bohrs. Also, the number of two-electron integrals that need to be generated to construct $(\mu\overline{\nu}|\underline{\lambda}\overline{\sigma})$ drops by a third, down to 14%. For 12-glycine the energy deviation due to our long-range thresholding is even smaller. The number of μ, λ pair contributions neglected is, however, smaller as well. Admittedly, fairly large systems need to be considered before a significant amount of long-range contributions can be neglected in our approach. However, it is in accordance with the $1/R^4$ decay of the individual pair contributions to the interaction energy we discussed in Sec. VI.[39]

One could determine the decay parameters A_μ and τ_μ by data-fitting for each transformed occupied AOs, and automatically choose appropriate settings for ϵ and WS according to Eqs. (7.13) and (7.18). For the purpose of the present paper a more simple approach was taken. Parameter settings $\epsilon = 10^{-10}$ a.u. and $WS = 3$ Bohrs were found appropriate for each system and were used throughout our benchmark calculations.

B. Transformation matrices

The Fock matrix eigenvalues and the canonical MOs are not essential prerequisites for constructing the Laplace density matrices. As shown below, one may consider the Chebychev polynomial expansion[43] of $\exp(\boldsymbol{F}_{oo}t_\alpha)$ and $\exp(-\boldsymbol{F}_{vv}t_\alpha)$, where $\boldsymbol{F}_{oo}$ and $\boldsymbol{F}_{vv}$ are the occupied and virtual blocks of the Fock matrix in any convenient

basis.[7,40,41,44] When the Hartree–Fock calculation is carried out using a linear-scaling density matrix search technique such as conjugate gradient density matrix search (CGDMS),[7] the Laplace density matrices may be found by the Chebychev expansion of

$$\begin{aligned} \boldsymbol{H} &= \boldsymbol{UPFPU}^{\mathbf{t}} - \boldsymbol{UQFQU}^{\mathbf{t}}, \\ &= 2\boldsymbol{UPFPU}^{\mathbf{t}} - \boldsymbol{U}^{-\mathbf{1}}\boldsymbol{FU}^{-\mathbf{t}}, \end{aligned} \tag{7.22}$$

where the overlap matrix is decomposed in the Cholesky fashion as $\boldsymbol{S}=\boldsymbol{UU}^{\mathbf{t}}$. The sparsity of $\boldsymbol{UPFPU}^{\mathbf{t}}$ and $\boldsymbol{U}^{-\mathbf{1}}\boldsymbol{FU}^{-\mathbf{t}}$ has already been studied.[7] The occupied and virtual Laplace density matrices are then given by

$$\begin{aligned} \boldsymbol{X}^{\alpha} &= \boldsymbol{PU}\exp[\boldsymbol{H}t_{\alpha}]\boldsymbol{U}^{\mathbf{t}}\boldsymbol{P}, \\ \boldsymbol{Y}^{\alpha} &= (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{U}\exp[\boldsymbol{H}t_{\alpha}]\boldsymbol{U}^{\mathbf{t}}(\boldsymbol{I}-\boldsymbol{P}). \end{aligned} \tag{7.23}$$

We have found that the accuracy of the Chebychev expansion can be reliably estimated by the root-mean-square (rms) deviation with respect to $\exp(-|x|t)$ over the Fock eigenvalue scale. As usual, the eigenvalue scale is determined by the linear-scaling Lanczos algorithm.[45] Because $\boldsymbol{UPFPU}^{\mathbf{t}}$ and $\boldsymbol{UQFQU}^{\mathbf{t}}$ are singular, the HOMO and LUMO energies are determined using instead $a\boldsymbol{UPFPU}^{\mathbf{t}}-b\boldsymbol{UQFQU}^{\mathbf{t}}$, where the constants a and b are made, in sequence, suitably large.

It has been recently shown that the density matrix can be obtained in $\mathcal{O}(N)$ effort by Chebychev expansion of the Fermi-Dirac distribution function.[41,42] In this density matrix search method, the Chebychev polynomial order for an accurate energy calculation typically exceeds 80. By comparison, the Chebychev polynomial order needed to construct $\exp[\boldsymbol{H}t_{\alpha}]$ is relatively low. For $t_{\alpha}=0.02$ a.u., a polynomial expansion order of 7–10 is needed to obtain in practice a rms deviation of 10^{-9} a.u. from the Laplace density matrix constructed by diagonalization, whereas a polynomial order of ~30 is needed for $t_{\alpha}=3.0$ a.u.

VIII. BENCHMARK CALCULATIONS

Even though polyglycines feature fairly large domains — between 17 and and 22 Bohrs depending on whether the orbital is centered on an oxygen, nitrogen or carbon atom — the onset of the linear scaling behavior of the calculation occurs for 517 basis functions (i.e., 12-glycine). As shown in Fig. 5, both Step 1 and Step 2 are clearly scaling linearly with system size beyond this point. The three rightmost data points fit nicely to a straight line. We remind the reader that Step 1 refers to the generation and first quarter transformation of AO integrals and that the CPU times reported here are for the Laplace quadrature point with the smallest t_{α}. CPU times and the number of first-quarter transformed integrals are listed in Table IV. The break seen in Fig. 5 is due to the overhead of the multipass algorithm. The calculation on 22-glycine was carried out twice; first, in 14 passes and second, in 19 passes.

It is interesting to quantify the scaling exponent γ in the asymptotic region. Using the slope of a log–log plot—that is, a least-square fit of the form $y=ax^{\gamma}$ — can be misleading in cases where the straight line extrapolation does not go through the origin. The scaling exponent measured this way

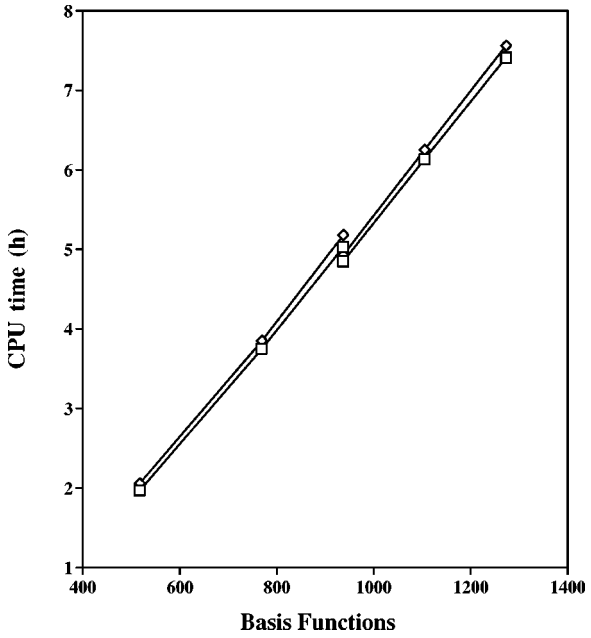

FIG. 5. CPU times (Ref. 38) for Step 1 and 2 of AO-Laplace MP2 for polyglycines using the 3-21G basis set. Step 1: Squares; step 2: Diamonds. The lines are fit to the last three data points.

can be very different for two sets of data taken from two parallel lines. To better measure the distance to the linear-scaling asymptote, we have instead fitted our data to the more general form $y=ax^{\gamma}+b$. The scaling exponents for the CPU times of Step 1 and Step 2 are $\gamma=1.14$ and $\gamma=1.02$, respectively. Because the first quarter transformation is the time dominant step, the scaling exponent for the integral count can also be used to measure the scaling of the AO-Laplace MP2 method. The scaling exponent for the first-quarter transformed integrals is $\gamma=1.1$.

It is perhaps worthwhile noting that linear scaling Hartree–Fock exchange occurs for a similar size linear molecule.[9–12] Carrying the analogy further, linear-scaling Hartree–Fock exchange occurs for water clusters containing between 100 and 150 molecules, therefore, the onset of linear-scaling for our AO-MP2 algorithm should not be expected to occur for a smaller size three-dimensional molecules. We have performed, instead, calculations on a slab of

TABLE IV. CPU times (Ref. 38) (in minutes) for steps 1 and 2 (see Sec. III) and first-quarter-transformed integral count (in billions) for polyglycine chains at the 3-21G level of theory, $\theta=10^{-7}/w$ a.u.

		CPU time		$(\mu\nu\|\lambda\sigma)$
Molecule	Basis functions	Step 1	Step 2	Integral count
8-glycine[a]	349	50	66	0.3251
12-glycine [a]	517	118	161	0.7231
12-glycine [b]	517	118	124	0.7009
18-glycine [b]	769	225	231	1.3064
22-glycine [b]	937	302	311	1.7099
22-glycine [c]	937	291	296	1.7099
26-glycine [c]	1105	368	375	2.1135
30-glycine[c]	1273	445	454	2.5170

[a]6 integral transformation passes, no long-range correlation thresholding.
[b]14 integral transformation passes, $WS=3$ Bohrs.
[c]19 integral transformation passes, $WS=3$ Bohrs.

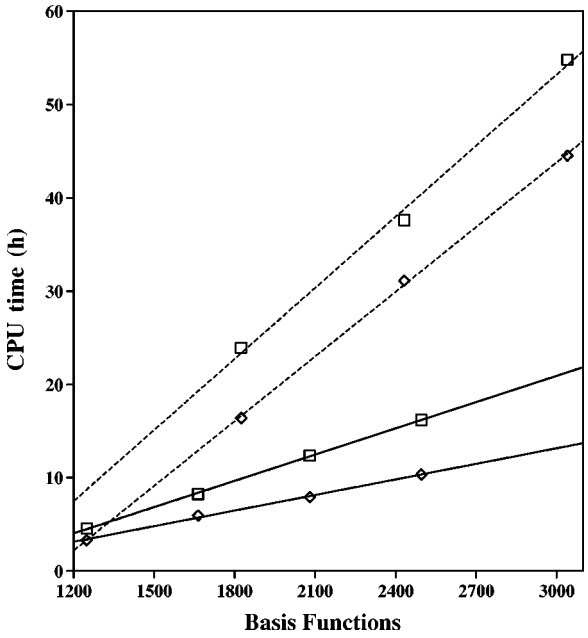

FIG. 6. CPU times (Ref. 38) for Steps 1 and 2 of AO-Laplace MP2 for water clusters using the 6-31G* basis set (dashed lines) and the 3-21G basis set (solid lines). Step 1: Squares; step 2: Diamonds. The lines are fit to all data points.

water clusters assembled by reproducing in two dimensions a three-dimensional unit cell of eight water molecules. It is also useful to compare the computational cost of our AO-Laplace MP2 algorithm with conventional Hartree–Fock algorithms. Gaussian's[33] quadratic scaling direct Hartree–Fock algorithm was used for all calculations presented here. The crossover between the full Hartree–Fock calculation and one quadrature point for the AO-Laplace MP2 calculation occurs for water clusters containing between 128 and 160 water molecules for both 3-21G and 6-31G* basis sets. The largest clusters studied here are $(H_2O)_{192}$, or 2496 basis functions using the 3-21G basis set, and $(H_2O)_{160}$, or 3040 basis functions using the 6-31G* basis set. Figure 6 shows the growth of the computational effort with system size. The CPU times for these very large systems as well as the integral counts are listed in Table V. As shown in Fig. 6, the computational effort grows linearly. Once again, linear scaling for our algorithm is substantiated by the line fit of the CPU times as well as least-square fit scaling exponents. Using the 3-21G basis-set, the scaling exponents for the CPU times of Step 1 and Step 2 are $\gamma=1.22$ and $\gamma=1.12$, respectively. The scaling exponent for the number of quarter transformed integrals evaluated and stored is $\gamma=1.15$. As expected, when using the 6-31G* basis set, the growth of the computational effort exhibits a steeper slope but remains linear. The scaling exponents for the two transformation steps are $\gamma=1.48$ and $\gamma=1.20$. To the best of our knowledge, the calculations presented here are the largest MP2 calculations to date.

TABLE V. CPU times (Ref. 38) (in min.) for Steps 1 and 2 (see Sec. III) and first-quarter-transformed integral count (in billions) for two-dimensional water clusters at the 3-21G and 6-31G* levels of theory. $\theta=10^{-7}/w$ a.u. and $WS=3$ Bohrs.

			CPU time		$(\mu\nu\|\lambda\sigma)$
Molecule	Basis set	Basis functions	Step 1	Step 2	Integral count
$(H_2O)_{96}$	3-21G	1248	274	198	1.6456
	6-31G*	1824	1436	983	6.0637
$(H_2O)_{128}$	3-21G	1664	495	357	2.6127
	6-31G*	2432	2258	1866	10.0056
$(H_2O)_{160}$	3-21G	2080	742	476	3.6562
	6-31G*	3040	3188	2781	14.0248
$(H_2O)_{192}$	3-21G	2496	972	621	4.7013

IX. CONCLUSION

In this paper we have shown that by expressing the MP2 correlation energy in the atomic orbital basis via the Laplace transform ansatz, the power-law decay of the Coulomb correlation energy and the exponential decay of the exchange energy of molecules with large HOMO–LUMO gaps are made apparent. These decay properties make it possible to obtain the canonical MP2 energy with a computational effort scaling quadratically with system size. Moreover, we have shown that the long-range contributions can be thresholded in a reliable and consistent fashion, resulting in a linear-scaling MP2 algorithm. We have also demonstrated that using present computer technology, studies of very large molecular systems, systems containing several hundreds of atoms, are now within reach at the MP2 level of theory.

ACKNOWLEDGMENTS

We are thankful to Andreas Savin for many stimulating and fruitful discussions on the topic of long-range correlation (Sec. VI). We also thank Nicholas Handy for calling our attention to the Euler–McLaurin quadrature scheme. This work was supported by the National Science Foundation (CHE-9618323). We dedicate this paper to the memory of Jan Almlöf and Marco Häser.

[1] M. Häser, J. Almlöf, and G. E. Scuseria, Chem. Phys. Lett. **181**, 497 (1991).
[2] M. Schutz and R. Lindh, Theor. Chim. Acta **95**, 13 (1997).
[3] M. Bühl, S. Patchkovskii, and W. Thiel, Chem. Phys. Lett. **275**, 14 (1998).
[4] A. D. Boese and G. E. Scuseria, Chem. Phys. Lett. (in press).
[5] M. C. Strain, G. E. Scuseria, and M. J. Frisch, Science **271**, 51 (1996).
[6] C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon, Chem. Phys. Lett. **253**, 268 (1996).
[7] J. M. Millam and G. E. Scuseria, J. Chem. Phys. **106**, 5569 (1997).
[8] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, Chem. Phys. Lett. **257**, 213 (1996).
[9] J. C. Burant, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **105**, 8969 (1996).
[10] E. Schwegler and M. Challacombe, J. Chem. Phys. **105**, 2726 (1996).
[11] E. Schwegler, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[12] C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[13] S. Goedecker, Rev. Mod. Phys. (in print).
[14] P. R. Taylor, Int. J. Quantum Chem. **31**, 521 (1987).
[15] M. Head-Gordon, J. A. Pople, and M. J. Frisch, Chem. Phys. Lett. **153**, 503 (1988).
[16] S. Saebo and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[17] J. W. Boughton and P. Pulay, J. Comput. Chem. **14**, 736 (1993).
[18] G. Rauhut, P. Pulay, and H. J. Werner, J. Comput. Chem. **19**, 1241 (1998).
[19] G. Hetzer, P. Pulay, and H. J. Werner, Chem. Phys. Lett. **290**, 143 (1998).
[20] M. Schutz, G. Rauhut, and H. J. Werner, J. Phys. Chem. **102**, 5997 (1998).

[21] R. B. Murphy, M. D. Beachy, R. A. Friesner, and M. N. Ringnalda, J. Phys. Chem. **103**, 1481 (1995).
[22] R. B. Murphy, W. T. Pollard, and R. A. Friesner, J. Chem. Phys. **106**, 5073 (1997).
[23] F. Weigand, M. Häser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[24] D. E. Bernholdt and R. J. Harrison, J. Chem. Phys. **109**, 1593 (1998).
[25] F. Weigand and M. Häser, Theor. Chem. Acc. **97**, 331 (1997).
[26] M. W. Feyereisen, D. Feller, and D. A. Dixon, J. Chem. Phys. **100**, 2993 (1996).
[27] D. Feller, E. D. Glendening, D. E. Woon, and M. W. Feyereisen, J. Chem. Phys. **103**, 3526 (1995).
[28] J. Almlöf, Chem. Phys. Lett. **181**, 319 (1991).
[29] M. Häser and J. Almlöf, J. Chem. Phys. **96**, 489 (1992).
[30] M. Häser, Theor. Chim. Acta **87**, 147 (1993).
[31] A. K. Wilson and J. Almlöf, Theor. Chim. Acta **95**, 49 (1997).
[32] C. W. Murray, N. C. Handy, and G. J. Laming, Mol. Phys. **78**, 997 (1993).
[33] GAUSSIAN 99, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, and J. A. Pople (Gaussian, Inc., Pittsburgh, PA, 1998).
[34] J. Almlöf, K. Faegri, and K. Korsell, J. Comput. Chem. **3**, 385 (1982).
[35] M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[36] D. L. Strout and G. E. Scuseria, J. Chem. Phys. **102**, 8448 (1995).
[37] W. Kohn, Int. J. Quantum Chem. **56**, 229 (1995).
[38] All timings are for a single MIPS R10K/195 MHz processor of a SGI Origin 2000 computer. In all calculations, molecular point group symmetry was ignored and all electrons are correlated.
[39] For reference, we point out that 10^{-6} Hartree corresponds to $1/R_0^4$ where $R_0 = 32$ Bohrs.
[40] J. P. Stewart, P. Császár, and P. Pulay, J. Comput. Chem. **3**, 227 (1982).
[41] K. R. Bates, A. D. Daniels, and G. E. Scuseria, J. Chem. Phys. **109**, 3008 (1998).
[42] R. Baer and M. Head–Gordon, J. Chem. Phys. **107**, 10003 (1997).
[43] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and W. P. Flannery, *Numerical Recipes in Fortran*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[44] J. Rubio, A. Povill, J. P. Malrieu, and P. Reinhardt, J. Chem. Phys. **107**, 10044 (1997).
[45] B. N. Bartlett, *The Symmetric Eigenvalue Problem* (Prentice-Hall, Englewood Cliffs, New Jersey, 1980).