Theor Chim Acta (1993) 87: 147–173

Theoretica
Chimica Acta

Møller–Plesset (MP2) perturbation theory for large molecules*

Marco Häser
Institut für Physikalische Chemie, Lehrstuhl für Theoretische Chemie, Universität Karlsruhe,
Kaiserstrasse 12, D-76128 Karlsruhe, Germany

Received November 10, 1992/Accepted April 14, 1993

Summary. A novel formulation of MP2 theory is presented which starts from the Laplace transform MP2 ansatz, and subsequently moves from a molecular orbital (MO) representation to an atomic orbital (AO) representation. Consequently, the new formulation is denoted AO-MP2. As in traditional MP2 approaches electron repulsion integrals still need to be transformed. Strict bounds on the individual MP2 energy contribution of each intermediate four-index quantity allow to screen off numerically insignificant integrals with a single threshold parameter. Implicit in our formulation is a bound to two-particle density matrix elements. For small molecules the computational cost for AO-MP2 calculations is about a factor of 100 higher than for traditional MO-based approaches, but due to screening the computational effort in larger systems will only grow with the fourth power of the size of the system (or less) as is demonstrated both in theory and in application. MP2 calculations on (non-metallic) crystalline systems seem to be a feasible extension of the Laplace transform approach. In large molecules the AO-MP2 ansatz allows massively parallel MP2 calculations without input/output of four-index quantities provided that each processor has in-core memory for a limited number of two-index quantities. Energy gradient formulas for the AO-MP2 approach are derived.

Key words: Correlation – MP2 – AO – Laplace transform – Bounds – Parallel computer – Gradient – Crystal – Solid

1 Introduction

Møller–Plesset second-order perturbation theory (MP2) [1] provides the most economical way of including dynamical electron correlation in *ab initio* electronic structure calculations of molecules. It has been applied to a large variety of problems, and has proven to successfully correct minor deficiencies of the Hartree–Fock self-consistent field (SCF) method [2]. A recent success of MP2 theory is the improvement in calculated magnetic shieldings of ^{17}O nuclei by almost an order of magnitude in accuracy as compared to SCF results [3].

---

* Dedicated to Prof. W. Kutzelnigg whose books on theoretical chemistry aroused my interest in this field

Standard MP2 theory, however, if applied to large molecular systems, is computationally considerably more demanding than SCF techniques. The current computational limit for the applicability of MP2 theory is at about 100 valence electrons; meaningful applications beyond are possible only if aided by high molecular point group symmetry [4].

For large molecules the dominant computational step in MP2 calculations is the transformation of electron repulsion integrals (ERIs) from a given basis set, usually contracted cartesian Gauss type functions (CGTF), to a basis of occupied and virtual canonical molecular orbitals (MO) as obtained by the SCF procedure. The computation time for this transformation scales as nN^4 where n is the number of occupied MOs, and N is the number of CGTFs (we assume little difference between the number of CGTFs and the number of virtual MOs). The computational effort can be reduced if near-zero integrals are neglected in the course of the first three quarter transformation steps [5]. Similar integral screening procedures fail when applied to the last quarter transformation step (with an operation count proportional to n^2N^3) since canonical SCF MOs have non-zero amplitudes practically everywhere in a molecule. Therefore the asymptotic operation count for traditional MP2 algorithms scales with the fifth power of the size of the system.

This situation is in marked contrast to SCF algorithms which completely avoid a transformation of ERIs from the CGTF basis to an MO basis: In the asymptotic limit pre-screening of ERIs (i.e. use of integral bounds to identify near-zero integrals not to be evaluated) reduces the operation count for the construction of the Fock matrix from N^4 towards N^2 (though with a considerably larger prefactor) [5–7]. The computational disparity between MP2 and SCF calculations is thus bound to grow. This is the more true since MP2 calculations on large molecules also involve input/output operations of partially transformed integrals (n^2N^2 or at best $\sim nN^2$ if CGTF ERIs are calculated repeatedly [8]) to some mass storage device while direct SCF techniques [7] can proceed in-core with no more than $\sim N^2$ storage requirements.

To reduce the computational labour involved in the calculation of correlated wave functions many workers have suggested localized occupied MOs together with local correlation spaces [9–11]. The latter were represented, for example, by a truncated expansion of approximate pair natural orbitals (PNO). This approach has been helpful in variational and infinite order calculations of the correlation energy, but is too costly for MP2 calculations.

Unoptimized local correlation spaces constructed from a subset of CGTFs are more economical [12, 13], and have been successfully applied to MP2 theory [14]. The problem here is the deliberate truncation of the local correlation spaces on the grounds of empirical rules. Typical errors are of the order of one per cent in the correlation energy [13], and depend on the choice of the localization procedure. This can certainly be improved by enlarging the local correlation spaces, but a systematic and efficient approach has not been devised.

Much the same can be said about the Local Ansatz [15], to which similar or somewhat larger error bars apply. The Local Ansatz does not use localized molecular orbitals, but relies on a particular choice of two-electron excitation operators which are defined in terms of empirically constructed local regions [16].

This situation can be summarized as follows: there exist methods for the approximate calculation of (MP2) correlation energies (a more complete account can be found in [13]) which guarantee computational savings over standard MP2 theory, but they rely on empirical rules; their errors – typically of the order of 1% – are difficult to assess, and efficient strategies for their systematic reduction are unknown. Similar to semi-empirical approaches, these methods are useful in their own right, and some of them may become part of the main stream of computational quantum chemistry.

However, we desire a rigorous formulation of MP2 theory which in applications to large molecules is computationally feasible on workstation clusters or massively parallel computers with an asymptotic operation count $\sim N^4$ or lower, and storage requirements no higher than $\sim N^2$. It should allow geometry optimizations, that is, energies should be accurate to within 10^{-5} Hartree, and analytical gradients have to be available. Finally the ansatz should have some natural extension towards solid state applications.

In this paper we present a formulation of MP2 theory in the CGTF basis set (actually there is no restriction to any particular choice of basis set). Our new approach is meant to solve *all* of the above-mentioned problems. Some preliminary applications will be discussed.

2 Theory

The starting point of our considerations is the expression for the MP2 energy of a closed shell system in terms of canonical SCF orbitals:

$$E = -\sum_{ijab} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}. \tag{1}$$

As usual, i, j designate doubly occupied MOs while a, b represent virtual MOs. ε_i, $\varepsilon_j, \ldots$ are corresponding orbital energies.

The electron repulsion integrals (ERI) in the MO basis:

$$(ia|jb) = \langle i(1)j(2)|r_{12}^{-1}|a(1)b(2)\rangle \tag{2}$$

are obtained by an integral transformation from ERIs evaluated for quadruples of CGTFs $\nu\mu\chi\lambda$:

$$(ia|jb) = \sum_{\nu\mu\chi\lambda} (\nu\mu|\chi\lambda) C_{\nu i} C_{\mu a} C_{\chi j} C_{\lambda b}. \tag{3}$$

All algorithms (whether direct, semi-direct or completely disc-oriented) which are based on Eqs. (1–3) will be termed conventional in this paper.

2.1 The Laplace transform MP2 ansatz

Asymptotically the most expensive step in conventional algorithms is the transformation, Eq. (3), with a fifth-power dependence on the size of the system. It is the orbital energy denominator in Eq. (1) which necessitates ERIs formed over canonical SCF MOs. Recently Almlöf [17] suggested a novel way to remove this disturbing denominator by a Laplace transform ansatz:

$$E = -\int_0^\infty dt\, e(t) \tag{4}$$

where

$$e(t) = \sum_{ijab} (i(t)a(t)|j(t)b(t))[2(i(t)a(t)|j(t)b(t)) - (i(t)b(t)|j(t)a(t))] \tag{5}$$

and

$$|i(t)\rangle = |i\rangle \exp((\varepsilon_i - \varepsilon_{\mathrm{F}})t/2), \quad |a(t)\rangle = |a\rangle \exp(-(\varepsilon_a - \varepsilon_{\mathrm{F}})t/2) \tag{6}$$

are scaled canonical orbitals. ε_F is a free parameter which for reasons of numerical stability is best chosen between the energies of the highest occupied orbital (HOMO) and the lowest unoccupied orbital (LUMO), e.g. $\varepsilon_{\mathrm{F}} = (\varepsilon_{\mathrm{HOMO}} + \varepsilon_{\mathrm{LUMO}})/2$.

As Almlöf [17] pointed out, Eq. (5) still holds if the t-dependent orbitals, Eq. (6), are localized by some orthogonal transformation, but this approach will not be investigated in this work.

The first obstacle to any application of the Laplace transform MP2 ansatz is the t-integration, Eq. (4). This problem has been solved in [18]. A functional approximation scheme has been introduced which allows to substitute the t-integration by a finite summation:

$$\frac{1}{x} = \int_0^\infty dt\, \mathrm{e}^{-xt} \approx \sum_{\alpha=1}^{\tau} w_\alpha \mathrm{e}^{-xt_\alpha}. \tag{7}$$

This is a useful approximation if the orbital energy differences $x = \varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j > 0$ are confined to some finite interval $[x_{\min}, x_{\max}]$, and if the parameters w_α and t_α are determined by a least-squares condition [18]:

$$\int_{x_{\min}}^{x_{\max}} dx\, f(x) \left[1/x - \sum_{\alpha=1}^{\tau} w_\alpha \exp(-xt_\alpha) \right]^2 = \min! \tag{8}$$

Here $f(x)$ is the distribution density of $x = \varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j$ (which can be calculated from the distribution of $\varepsilon_a - \varepsilon_i$).

The relative error in the MP2 energy, if calculated with the functional approximation, Eqs. (7, 8), can be made as small as desired. In typical applications $\tau = 5 \ldots 8$ exponentials in Eq. (8) suffice for μHartree accuracy [18]. The number τ of exponentials necessary to meet some pre-set accuracy in the MP2 energy E can be determined a priori from the following estimates of the *relative* discretization error [18]:

$$L1 = \int dx\, f(x) \left| \frac{1}{x} - \sum_\alpha w_\alpha \exp(-xt_\alpha) \right| \Big/ \left(\int dx\, f(x)/x \right), \tag{9}$$

and

$$L2 = \left(\int dx\, f(x) \left| \frac{1}{x} - \sum_\alpha w_\alpha \exp(-xt_\alpha) \right|^2 \right)^{1/2} \left(\int dx\, f(x) \right)^{1/2} \Big/ \left(\int dx\, f(x)/x \right). \tag{10}$$

The corresponding formulas in [18] contain some misprints. The number of exponentials τ determined this way depends primarily on the width of the interval $[x_{\min}, x_{\max}]$, but does not depend on the size of the molecular system.

With the Laplace integration replaced by the finite summation, Eq. (7), the expressions for the MP2 energy, Eqs. (4, 5, 6), now read [18]:

$$E \approx - \sum_{\alpha=1}^{\tau} \operatorname{sgn}(w_\alpha) e_\alpha \tag{11}$$

$$e_\alpha = \sum_{ijab} (i_\alpha a_\alpha | j_\alpha b_\alpha) \left[2(i_\alpha a_\alpha | j_\alpha b_\alpha) - (i_\alpha b_\alpha | j_\alpha a_\alpha) \right] \tag{12}$$

$$|i_\alpha\rangle = |i\rangle\, |w_\alpha|^{1/8} \exp((\varepsilon_i - \varepsilon_\mathrm{F}) t_\alpha / 2),$$

$$|a_\alpha\rangle = |a\rangle\, |w_\alpha|^{1/8} \exp(-(\varepsilon_a - \varepsilon_\mathrm{F}) t_\alpha / 2). \tag{13}$$

The expansion coefficients w_α have always been found to be positive, that is, $\operatorname{sgn}(w_\alpha) = +1$ in Eq. (11).

So far we have recapitulated some of the Laplace transform MP2 theory outlined in [17, 18]. It has been claimed [18] that the new formalism allows to calculate the MP2 two-particle density in the CGTF basis without prior calculation of electron repulsion integrals in the MO basis. This will become fairly obvious after we re-formulate the Laplace MP2 ansatz in the CGTF basis. We term our new formulation AO-MP2 with the familiar acronym AO for atomic orbital (while the subsequent considerations are for pure CGTFs, they also apply to more general atomic basis functions as will be discussed in Sect. 6).

2.2 The AO-MP2 ansatz

To relate the conventional MO-based formulations of MP2 theory to the new approach we need the expansion of the MOs in terms of CGTFs:

$$|i\rangle = \sum_\nu |\nu\rangle\, C_{\nu i}, \qquad |a\rangle = \sum_\nu |\nu\rangle\, C_{\nu a}. \tag{14}$$

Let us define then two symmetric matrices $\bar{D}^{(\alpha)}$ and $\underline{D}^{(\alpha)}$ as:

$$\bar{D}^{(\alpha)}_{\nu\mu} = |w_\alpha|^{1/4} \sum_i C_{\nu i} \exp((\varepsilon_i - \varepsilon_\mathrm{F}) t_\alpha) C_{\mu i}$$

$$\underline{D}^{(\alpha)}_{\nu\mu} = |w_\alpha|^{1/4} \sum_a C_{\nu a} \exp(-(\varepsilon_a - \varepsilon_\mathrm{F}) t_\alpha) C_{\mu a}. \tag{15}$$

$\bar{D}^{(\alpha)}$ and $\underline{D}^{(\alpha)}$ relate to occupied–occupied and virtual–virtual blocks of the exponential of the Fock matrix, respectively. Note that some MOs may deliberately be excluded from the summation in Eq. (15), and thus from the correlation treatment ('frozen orbitals') as is often done for core orbitals ('frozen core') in conventional MP2 theory.

$\bar{D}_{\nu\mu}(w_\alpha^{1/4} = 2, t_\alpha = 0)$ is the closed-shell SCF one-particle density matrix in the CGTF basis (unless occupied orbitals have been frozen).

If we insert Eqs. (14, 15) into Eqs. (12, 13) we obtain:

$$e_\alpha = \sum_{\nu'\mu'\chi'\lambda'} \sum_{\nu\mu\chi\lambda} (\nu'\mu'|\chi'\lambda') \bar{D}^{(\alpha)}_{\nu'\nu}\, \underline{D}^{(\alpha)}_{\mu'\mu}\, \bar{D}^{(\alpha)}_{\chi'\chi}\, \underline{D}^{(\alpha)}_{\lambda'\lambda}\; [2(\nu\mu|\chi\lambda) - (\nu\lambda|\chi\mu)]. \tag{16}$$

Together with Eq. (11) this forms a simple expression for the MP2 energy with matrices and ERIs in the CGTF basis. The impractical eightfold summation over basis function indices can be removed if we define:

$$|\bar{\nu}\rangle = |\bar{\nu}(\alpha)\rangle = \sum_{\mu} |\mu\rangle\, \bar{D}^{(\alpha)}_{\mu\nu}, \quad |\underline{\nu}\rangle = |\underline{\nu}(\alpha)\rangle = \sum_{\mu} |\mu\rangle\, \underline{D}^{(\alpha)}_{\mu\nu}, \tag{17}$$

and (for simplicity we subsequently drop the superscript (α)):

$$(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda}) = \sum_{\lambda'}\left(\sum_{\chi'}\left(\sum_{\mu'}\left(\sum_{\nu'}(\nu'\mu'|\chi'\lambda')\,\bar{D}_{\nu'\nu}\right)\underline{D}_{\mu'\mu}\right)\bar{D}_{\chi'\chi}\right)\underline{D}_{\lambda'\lambda}. \tag{18}$$

Eq. (16) now reads:

$$e_{\alpha} = \sum_{\nu\mu\chi\lambda} |\,(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})\,[2(\nu\mu|\chi\lambda) - (\nu\lambda|\chi\mu)]. \tag{19}$$

To make our notation transparent we recall that the bars and underbars in $\bar{\nu}$ or $\underline{\mu}$ represent linear mappings, Eq. (17), of the unbarred basis functions. A primed basis function on the other hand, e.g. ν', is a basis function and nothing else. Accordingly, $\bar{\nu}'$ is meant to be the mapped basis function ν'.

For each $\alpha = 1 \ldots \tau$ four N^5 transformation steps are necessary to establish $(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})$ according to Eq. (18) (it is possible to evaluate e_{α} in Eq. (17) by only three N^5 transformation steps, but this we will not be used here). Computationally this is much worse than the four steps of a conventional integral transformation for MP2 energies, Eq. (3), which formally scale with nN^4 and n^2N^3: In typical applications the conventional MP2 algorithm will be 100 times more efficient than the AO-MP2 ansatz. In the following Sect. 2.3 we discuss how the operation count for the AO-MP2 ansatz can be significantly reduced in applications to large molecules.

2.3 Integral screening in the AO-MP2 ansatz

The AO-MP2 ansatz can become competitive if an efficient strategy is devised to reduce the computational labour involved in the evaluation of e_{α}, Eqs. (18, 19), by a factor of 100 or more. For large molecules this may be achieved by screening integrals with near-zero contributions to the MP2 energy in each quarter transformation step, Eq. (18), and in the final assembly of e_{α}, Eq. (19). Here we outline our screening procedure while the meaning of 'large' and 'near-zero' will be addressed in Sects. 4 and 5.

For all of the intermediate quantities in Eq. (18), that is, for $(\nu'\mu'|\chi'\lambda')$, $(\bar{\nu}\mu'|\chi'\lambda')$, $(\bar{\nu}\underline{\mu}|\chi'\lambda')$, $(\bar{\nu}\underline{\mu}|\bar{\chi}\lambda')$, and $(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})$, the absolute values of their individual contributions to the MP2 energy are given by:

$$\begin{aligned} |(\nu'\mu'|\chi'\lambda')\,[2(\bar{\nu}'\underline{\mu}'|\bar{\chi}'\underline{\lambda}') - (\bar{\nu}'\underline{\lambda}'|\bar{\chi}'\underline{\mu}')]| &\leqslant |(\nu'\mu'|\chi'\lambda')|[2Z_{\nu'\mu'}\,Z_{\chi'\lambda'} + Z_{\nu'\lambda'}\,Z_{\chi'\mu'}] \\ &\leqslant Q_{\nu'\mu'}\,Q_{\chi'\lambda'}\,[2Z_{\nu'\mu'}\,Z_{\chi'\lambda'} + Z_{\nu'\lambda'}\,Z_{\chi'\mu'}], \quad (20) \\ |(\bar{\nu}\mu'|\chi'\lambda')\,[2(\nu\underline{\mu}'|\bar{\chi}'\underline{\lambda}') - (\nu\underline{\lambda}'|\bar{\chi}'\underline{\mu}')]| &\leqslant |(\bar{\nu}\mu'|\chi'\lambda')|[2Y_{\mu'\nu}\,Z_{\chi'\lambda'} + Y_{\lambda'\nu}\,Z_{\chi'\mu'}] \\ &\leqslant X_{\nu\mu'}\,Q_{\chi'\lambda'}\,[2Y_{\mu'\nu}\,Z_{\chi'\lambda'} + Y_{\lambda'\nu}\,Z_{\chi'\mu'}], \quad (21) \end{aligned}$$

$$\begin{aligned}|(\bar{\nu}\underline{\mu}|\chi'\lambda')\,[2(\nu\mu|\bar{\chi}'\underline{\lambda}') - (\nu\underline{\lambda}'|\bar{\chi}'\mu)]| &\leqslant |(\bar{\nu}\underline{\mu}|\chi'\lambda')|[2Q_{\nu\mu}\,Z_{\chi'\lambda'} + Y_{\lambda'\nu}\,X_{\chi'\mu}]\\ &\leqslant Z_{\nu\mu}\,Q_{\chi'\lambda'}\,[2Q_{\nu\mu}\,Z_{\chi'\lambda'} + Y_{\lambda'\nu}\,X_{\chi'\mu}],\end{aligned} \tag{22}$$

$$\begin{aligned}|(\bar{\nu}\underline{\mu}|\bar{\chi}\lambda')\,[2(\nu\mu|\chi\underline{\lambda}') - (\nu\underline{\lambda}'|\chi\mu)]| &\leqslant |(\bar{\nu}\underline{\mu}|\bar{\chi}\lambda')|[2Q_{\nu\mu}\,Y_{\lambda'\chi} + Y_{\lambda'\nu}\,Q_{\chi\mu}]\\ &\leqslant Z_{\nu\mu}\,X_{\chi\lambda'}\,[2Q_{\nu\mu}\,Y_{\lambda'\chi} + Y_{\lambda'\nu}\,Q_{\chi\mu}],\end{aligned} \tag{23}$$

and

$$\begin{aligned}|(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})\,[2(\nu\mu|\chi\lambda) - (\nu\lambda|\chi\mu)]| &\leqslant |(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})|[2Q_{\nu\mu}\,Q_{\chi\lambda} + Q_{\nu\lambda}\,Q_{\chi\mu}]\\ &\leqslant Z_{\nu\mu}\,Z_{\chi\lambda}\,[2Q_{\nu\mu}\,Q_{\chi\lambda} + Q_{\nu\lambda}\,Q_{\chi\mu}],\end{aligned} \tag{24}$$

respectively. The right-hand sides are strict upper bounds to the magnitudes of these correlation energy contributions. The quantities in the bounds are:

$$Q_{\nu\mu} = (\nu\mu|\nu\mu)^{1/2} \tag{25}$$

$$X_{\nu\mu} = (\bar{\nu}\mu|\bar{\nu}\mu)^{1/2} \tag{26}$$

$$Y_{\nu\mu} = (\underline{\nu}\mu|\underline{\nu}\mu)^{1/2} \tag{27}$$

$$Z_{\nu\mu} = (\bar{\nu}\underline{\mu}|\bar{\nu}\underline{\mu})^{1/2}. \tag{28}$$

Later we will change our definition of Z for computational convenience (Eq. (33)).

The proofs of inequalities (20–24) follow directly from Schwarz' inequality for interacting charge distributions [19]:

$$|(\nu\mu|\chi\lambda)| \leqslant (\nu\mu|\nu\mu)^{1/2}(\chi\lambda|\chi\lambda)^{1/2} \tag{29}$$

which holds for any type of orbitals provided the r.h.s. exists. The efficiency of Schwarz' inequality, Eq. (29), in electronic structure calculations has first been demonstrated in [20].

For the inequalities (20–24) to be useful in a screening procedure, the matrices Q, X, Y, and Z, Eqs. (25–28), have to be calculated and stored beforehand. This is certainly not a problem for the symmetric matrix Q, which is also used in direct SCF programs [20], and can be obtained by N^2 integral evaluations, Eq. (25). X and Y, Eqs. (26, 27), require N^3 integral evaluations and a subsequent transformation step (with N^4 floating point operations) which can be carried out in a fully direct mode, that is, without storing intermediate matrix quantities:

$$X^2_{\nu\mu} = \sum_{\chi\lambda} (\chi\mu|\lambda\mu)\bar{D}_{\chi\nu}\,\bar{D}_{\lambda\nu}$$

$$Y^2_{\nu\mu} = \sum_{\chi\lambda} (\chi\mu|\lambda\mu)\underline{D}_{\chi\mu}\,\underline{D}_{\lambda\nu}. \tag{30}$$

Integral screening can be used at this stage, but it has to be kept in mind that matrix elements of Q, X, and Y, which are always positive, may not be assigned values smaller than the magnitudes of their possible errors – otherwise their use in integral bounds will lead to gross errors.

We now turn to the evaluation of matrix Z, Eq. (28). If we proceed as in the calculation of matrices X and Y, Eq. (30), we will end up with (formally) N^4 integral evaluations plus a subsequent N^5 (discless) transformation step: this approach is almost as costly as the operations we want to avoid. Fortunately, however, we can construct upper bounds to $Z_{\nu\mu} = (\bar{\nu}\underline{\mu}|\bar{\nu}\underline{\mu})^{1/2}$ which will meet our demands:

$$(\bar{\nu}\underline{\mu}|\bar{\nu}\underline{\mu}) = \sum_{\chi\lambda} (\bar{\nu}\chi|\bar{\nu}\lambda)\underline{D}_{\chi\mu}\underline{D}_{\lambda\mu} \leqslant \left(\sum_{\chi} X_{\nu\chi}|\underline{D}_{\chi\mu}|\right)^2 = x_{\nu\mu}^2, \tag{31}$$

and

$$(\bar{\nu}\underline{\mu}|\bar{\nu}\underline{\mu}) = \sum_{\chi\lambda} (\chi\underline{\mu}|\lambda\underline{\mu})\bar{D}_{\chi\nu}\bar{D}_{\lambda\nu} \leqslant \left(\sum_{\chi} Y_{\mu\chi}|\bar{D}_{\chi\nu}|\right)^2 = y_{\nu\mu}^2. \tag{32}$$

These inequalities follow from the triangular inequality and from Schwarz inequality, the latter being applied to $(\bar{\nu}\chi|\bar{\nu}\lambda)$ and to $(\chi\underline{\mu}|\lambda\underline{\mu})$, respectively. The matrices x and y, which are defined by Eqs. (31, 32), are thus available by an N^3 matrix multiplication step, and may be used for a re-definition of $Z_{\nu\mu}$:

$$Z_{\nu\mu} = \min(x_{\nu\mu}, y_{\nu\mu}). \tag{33}$$

This new definition of Z can be inserted into inequalities (20–24) since $(\bar{\nu}\underline{\mu}|\bar{\nu}\underline{\mu})^{1/2} \leqslant \min(x_{\nu\mu}, y_{\nu\mu})$. The preliminary definition according to Eq. (28) will no longer be used.

We sum up the contents of this section: For each individual partially transformed integral in Eq. (18) we have devised a bound to its contribution to the MP2 energy. These bounds, Eqs. (20–24), rely on four N^2 matrices: Q, X, Y, and Z as defined by Eqs. (25, 30–33). These four matrices can be obtained by at most N^3 integral evaluations and a direct N^4 integral transformation step. The bounds are derived from Schwarz' inequality, Eq. (29), and can thus be expected to be very efficient [20]. Their performance in the present context, however, will have to pass some practical tests (Sects. 3.2 and 5).

3 Algorithms

It is appropriate to clarify the application of our ansatz in algorithmic form. We will present two different algorithms. So far only the first algorithm, a two-step out-of-core transformation algorithm, has been implemented, and all numerical tests in Sect. 5 relate to this preliminary implementation. Later we discovered a superior algorithm, Sect. 3.3, which solves the memory bottleneck problem in MP2 calculations.

3.1 Two-step out-of-core algorithm

We outline our preliminary implementation of an AO-MP2 program, which uses semi-direct two-step out-of-core transformations [21].

1. Get basic data (molecular geometry, atomic basis sets, SCF MOs, required accuracy).
2. Determine the distribution function $f(x)$ of orbital energy differences $\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j$ [18].
3. From the required accuracy and from $f(x)$ calculate the parameters τ, w_α, $t_\alpha(\alpha = 1 \,..\, \tau)$ of the exponential approximation to $1/x$, Eqs. (7–10).

4. Choose a threshold parameter ϑ for screening of partially transformed integrals in accordance with the required accuracy.
5. Calculate the integral bounds Q, Eq. (25).
6. For each $\alpha = 1 \,..\, \tau$ evaluate a contribution e_α to the MP2 energy E, Eqs. (11, 16).
7. Output the MP2 energy E.

The evaluation of each e_α in step 6 proceeds as follows:

6.1 Determine the matrices $\bar{D}$ and $\underline{D}$ according to Eq. (15).
6.2 Evaluate the integral bounds X and Y, Eq. (30), and the integral bound Z, Eqs. (31–33).
6.3. Calculate and transform CGTF ERIs $(\nu'\mu'|\chi'\lambda')$. Half-transformed integrals $(\bar{\nu}\underline{\mu}|\chi'\lambda')$ are written to direct access file.
6.4 Read half-transformed integrals in reverse order as $(\chi'\lambda'|\bar{\nu}\underline{\mu})$ and transform them to $(\bar{\chi}\underline{\lambda}|\bar{\nu}\underline{\mu})$. The fully transformed integrals $(\bar{\chi}\underline{\lambda}|\bar{\nu}\underline{\mu})$ can then be processed directly with CGTF integrals $(\chi\lambda|\nu\mu)$ and $(\chi\mu|\nu\lambda)$, which are evaluated 'on-the-fly', to yield an energy contribution to e_α, Eq. (19).

Integral screening is used in all critical stages of the calculation:

In step 6.3 integrals $(\nu'\mu'|\chi'\lambda')$ will be evaluated only if:

$$Q_{\nu'\mu'}\, Q_{\chi'\lambda'}[2Z_{\nu'\mu'}\, Z_{\chi'\lambda'} + Z_{\nu'\lambda'}\, Z_{\chi'\mu'}] > \vartheta. \tag{34}$$

After evaluation they will be transformed to $(\bar{\nu}\mu'|\chi'\lambda')$, provided they meet the condition:

$$|(\nu'\mu'|\chi'\lambda')|[2Z_{\nu'\mu'}\, Z_{\chi'\lambda'} + Z_{\nu'\lambda'}\, Z_{\chi'\mu'}] > \vartheta. \tag{35}$$

Quarter-transformed integrals $(\bar{\nu}\mu'|\chi'\lambda')$ will be transformed to half-transformed integrals $(\bar{\nu}\underline{\mu}|\chi'\lambda')$ only if:

$$|(\bar{\nu}\mu'|\chi'\lambda')|[2Y_{\mu'\nu}\, Z_{\chi'\lambda'} + Y_{\lambda'\nu}\, Z_{\chi'\mu'}] > \vartheta. \tag{36}$$

Half-transformed integrals (values and indices) are written to direct access file if:

$$|(\bar{\nu}\underline{\mu}|\chi'\lambda')|[2Q_{\nu\mu}\, Z_{\chi'\lambda'} + Y_{\lambda'\nu}\, X_{\chi'\mu}] > \vartheta. \tag{37}$$

The re-ordered half-transformed integrals $(\chi'\lambda'|\bar{\nu}\underline{\mu})$, which are read from file in step 6.4, may then be transformed to $(\bar{\chi}\lambda'|\bar{\nu}\underline{\mu})$ without further screening.

The last quarter transformation is carried out for integrals $(\bar{\chi}\lambda'|\bar{\nu}\underline{\mu})$ which fulfill:

$$|(\bar{\chi}\lambda'|\bar{\nu}\underline{\mu})|[2Q_{\nu\mu}\, Y_{\lambda'\chi} + Y_{\lambda'\nu}\, Q_{\chi\mu}] > \vartheta. \tag{38}$$

Finally, for each fully transformed integral $(\bar{\chi}\underline{\lambda}|\bar{\nu}\underline{\mu})$ the ERIs $(\chi\lambda|\nu\mu)$ and $(\chi\mu|\nu\lambda)$ need to be evaluated only if:

$$|(\bar{\chi}\underline{\lambda}|\bar{\nu}\underline{\mu})|[2Q_{\nu\mu}\, Q_{\chi\lambda} + Q_{\nu\lambda}\, Q_{\chi\mu}] > \vartheta. \tag{39}$$

All integral evaluations are of course carried out in batches [22], and corresponding quantities are arranged in appropriate order. Canonical symmetries of the type $(\nu\mu|\chi\lambda) = (\mu\nu|\chi\lambda) = (\nu\mu|\lambda\chi) = (\mu\nu|\lambda\chi)$ have been utilized in all integral evaluations, that is, $N^4/4$ ERIs and $3N^4/4$ ERIs would be calculated in steps 6.3 and 6.4, respectively, if no integral screening were used. The screening criteria, Eqs. (34–39), have been adapted accordingly. Half-transformed integrals $(\bar{\nu}\underline{\mu}|\chi'\lambda')$ are calculated and stored only for $\chi' \geqslant \lambda'$.

In our implementation an exploitation of $(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda}) = (\bar{\chi}\underline{\lambda}|\bar{\nu}\underline{\mu})$ as a means to cut computation cost in step 6.4 (by running the loop over χ with the restriction $\chi \leqslant \nu$) was incompatible with integral screening by upper bounds of energy contributions, Eqs. (34–39). As a possible rationalization we point out that integral neglect in one quarter-transformation step is certainly not independent from integral neglect at other stages of an AO-MP2 calculation. Usually such mutual effects contribute only in second order to the total error and become insignificant as ϑ becomes reasonably small ($\leqslant 10^{-7}$).

In this section we refrain from a discussion of parallelization or memory requirements as it would concern only our preliminary implementation (see Sect. 3.3 instead). It is worth mentioning that we obtain an a priori estimate of the number of non-negligible half-transformed integrals by counting all quadruples $\nu\mu\chi'\lambda'$ with:

$$Z_{\nu\mu}\, Q_{\chi'\lambda'}[2Q_{\nu\mu}\, Z_{\chi'\lambda'} + Y_{\lambda'\nu}\, X_{\chi'\mu}] > \vartheta. \tag{40}$$

If the required disc space for half-transformed integrals is predicted to be larger than the available space, we switch to a multiple pass algorithm in close analogy to [8].

3.2 *A note on multiple pass algorithms*

Multiple pass algorithms have been devised in conventional MP2 schemes with (semi-)direct ERI transformations [8] in order to cut disc space (or in-core memory) demands for (partially) transformed integrals, e.g. from n^2N^2 down to a minimum of nN^2 (we recall that n is the number of doubly occupied MOs). This is accomplished by partitioning the occupied space into $m \leqslant n$ sub-spaces M, and performing a partial MP2 calculation m times, each time evaluating quarter-transformed integrals $(i\mu|\chi\lambda)$ (and subsequent intermediates) only for $i \in M$. During the complete MP2 calculation the CGTF integrals thus have to be re-calculated for each occupied sub-space M, that is, integral evaluation formally becomes an mN^4 operation. Integral screening could reduce this asymptotically to $\sim mN^2$ with a large pre-factor, but with existing integral pre-screening algorithms for conventional MP2 methods, which do little better than to ignore CGTFs $(\nu\mu|\chi\lambda)$ if $Q_{\nu\mu}\, Q_{\chi\lambda} < \vartheta$, this asymptotic limit has not yet been attained.

As already mentioned in Sect. 3.1, we also admit multiple passes in the first quarter-transformation of our out-of-core AO-MP2 program. M then becomes a sub-set of CGTFs. Integral pre-screening is based on Eq. (34) rather than on $Q_{\nu\mu}\, Q_{\chi\lambda} > \vartheta$, and consequently we can expect to reach the $\sim mN^2$ limit for multiple ERI evaluation costs earlier than conventional direct multiple pass MP2 schemes. This claim may be backed by a comparison of these two pre-screening conditions carried out for the porphyrin molecule $C_{20}H_{14}N_4$ (split valence plus polarization (SVP) basis set = 430 CGTFs, $\vartheta = 10^{-9}$, pre-screening of whole integral batches [22]): For $Q_{\nu\mu}Q_{\chi\lambda} > \vartheta$ as pre-screening condition 29% of the integrals have to be evaluated while this number is reduced to 19% if Eq. (34) is used. For the distributions of magnitudes of matrix elements Q and Z cf. Table 3.

Pre-screening formulas like Eq. (34) may also be incorporated into conventional multiple pass MP2 algorithms (with appropriate adjustments like a discretized t-integration over $Z_{\nu'\mu'}(t)Z_{\chi'\lambda'}(t)$: this provides a bound to the relevant part of the two-particle density in the CGTF basis).

Finally let us point out that m passes through the CGTF evaluation in step 6.3 do not automatically imply an increase of integral evaluation costs by a factor of m. Using the same derivation as in Eq. (20) it may be shown that $Z_{\nu'\mu'}$ (and similarly also $Z_{\nu'\lambda'}$) may be replaced in screening formula (34) by modified matrix elements:

$$Z_{\nu'\mu'}(M) = \min\left(Z_{\nu'\mu'}, m \sum_{\chi\in M} Y_{\mu'\chi} |\bar{D}_{\chi\nu'}| \right). \tag{41}$$

This is a direct consequence of the fact that in pass M an integral $(\nu'\mu'|\chi'\lambda')$ will no longer contribute to the total MP2 energy E, but only to some well-defined share $E(M)$. Such a modified screening formula has not yet been implemented or tested.

3.3 In-core multiple pass algorithm

Other than conventional MP2 algorithms the previously outlined AO-MP2 algorithm requires an evaluation of CGTF integrals $(\nu\mu|\chi\lambda)$ and $(\nu\lambda|\chi\mu)$ for each fully transformed integral $(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})$ in the course of the second half-transformation, step 6.4 in Sect. 3.1. This part is not much affected by possible multiple passes in the first half-transformation.

Another type of AO-MP2 algorithm is obtained if we admit multiple pass type integral re-evaluation both at the beginning *and* at the end of the AO-MP2 transformation procedure. Only steps 6.3 and 6.4 of our previous algorithm, Sect. 3.1, need to be replaced by the following in-core multiple pass scheme:

6.A Partition the CGTF basis into m sub-spaces M of (maximum) dimension $\dim(M)$; for ERI evaluation in batches each sub-space M should contain only complete shells of CGTFs.

6.B For each pair of sub-spaces M, M' calculate a contribution $e_\alpha(M, M')$ to e_α, and thus to the MP2 energy E.

Step 6.B needs to be detailed:

B.1 Calculate all quarter-transformed integrals $(\bar{\nu}\mu'|\chi'\lambda')$ for $\nu \in M$, $\lambda' \in M'$, and all μ', χ' (integral screening, while applicable, will not be given attention in this section). The memory requirement for the quarter-transformed integrals is $\dim(M)^2 N^2$. The number of CGTF integrals $(\nu'\mu'|\chi'\lambda')$, which are evaluated and transformed 'on-the-fly', is $\dim(M)N^3$. The number of floating point operations for the transformation is $\dim(M)^2 N^3$.

B.2 Continue the in-core transformation to obtain all three-quarter transformed integrals $(\bar{\nu}\underline{\mu}|\bar{\chi}\lambda')$ for $\nu \in M$, $\lambda' \in M'$, and all μ, χ. Again the memory requirement is $\dim(M)^2 N^2$ while the number of floating point operations is $\dim(M)^2 N^3$.

B.3 Calculate and process immediately CGTF integrals $(\nu\mu|\chi\lambda)$ and $(\nu\lambda|\chi\mu)$ for $\nu \in M$ and all μ, χ, λ, i.e. evaluate:

$$\bar{R}_{\lambda\lambda'}(M, M') = \sum_{\nu\in M} \sum_{\mu,\chi} (\bar{\nu}\underline{\mu}|\bar{\chi}\lambda') \left[2(\nu\mu|\chi\lambda) - (\nu\lambda|\chi\mu)\right]. \tag{42}$$

This step takes $\dim(M)N^3$ integral evaluations, and $\dim(M)^2 N^3$ floating point operations for the construction of matrix $\bar{R}(M, M')$.

B.4 Evaluate:

$$e_\alpha(M, M') = \sum_{\lambda' \in M'} \sum_{\lambda} \bar{R}_{\lambda\lambda'}(M, M') \underline{D}_{\lambda'\lambda} . \tag{43}$$

Without integral screening taken into account, this multiple pass algorithm allows to calculate e_α (and thus the MP2 energy) by $N^5/\dim(M) = mN^4$ integral evaluations plus an N^5 operation count for the transformation steps. The largest matrices to be kept in memory have dimension $\dim(M)^2 N^2$, and since $\dim(M)$ can be any small positive number, e.g. 10, *the evaluation of e_α can be performed in-core*, that is, *without mass storage devices.* Furthermore, each contribution $e_\alpha(M, M')$ to the MP2 energy can be evaluated independently, e.g. *on different nodes of a massively parallel computer.*

4 Asymptotic operation count in very large systems

Before we monitor the performance of integral screening in the AO-MP2 algorithm in applications to medium sized molecules (Sect. 5) we want to promote some understanding by means of model systems.

The efficiency of integral screening is tied to the sparsity of matrices Q, X, Y, and Z, because the correlation energy contributions of individual partially transformed ERIs are bound by a product of four of these matrix elements, Eqs. (20–24). Of the four matrices Q, X, Y, and Z, asymptotic characteristics typical of very large molecules will first be exhibited by Q as it relates to the original (and well localized) basis set ($Q_{\nu\mu} = (\nu\mu|\nu\mu)^{1/2}$). As matrix elements of Q appear at least once in all bounds, Eqs. (20–24), their size distribution is most important with respect to the efficiency of integral screening in medium sized molecules ($\approx$ 20–50 atoms).

4.1 Sparsity in Q and implications

Let us introduce a simple mathematical model for the sparsity in Q. We consider a homogeneous distribution of s-type Gaussians in d dimensions ($d = 1, 2$, or 3) with ρ as the number of Gaussians per unit volume. For simplicity there is only one orbital exponent ζ. We then obtain for two Gaussians ν, μ separated by a distance $r = r_{\nu\mu}$:

$$Q_{\nu\mu} = (4\zeta/\pi)^{1/4} \exp(-\zeta r^2/2). \tag{44}$$

It is convenient to define the maximum matrix element:

$$Q_{\max} = \max Q_{\nu\mu} = (4\zeta/\pi)^{1/4}. \tag{45}$$

The number density n_r of basis function pairs ν, μ separated by a distance r is:

$$n_r\, dr = f_d\, \rho^2\, r^{d-1}\, dr, \tag{46}$$

where $f_1 = 2$, $f_2 = 2\pi$, and $f_3 = 4\pi$. From Eqs. (44–46) the number density n_q of matrix elements $Q_{\nu\mu}$ with:

$$q_{\nu\mu} = \ln(Q_{\max}/Q_{\nu\mu}) \tag{47}$$

in the infinitesimal interval between q and $q + dq$ $(q > 0)$ can be inferred as:

$$n_q\, dq = (f_d\, \rho^2/\zeta^{d/2})\,(2q)^{(d-2)/2}\, dq. \tag{48}$$

In planar systems $(d = 2)$ $n_q(q)$ is a step function, that is, the number of matrix elements $Q_{\nu\mu}$ with magnitudes between 10^{-4} and 10^{-5} is the same as between 10^{-9} and 10^{-10} (provided $Q_{\max} > 10^{-4}$). The latter result remains valid for any realistic distribution of orbital exponents (ζ in Eq. (48) then adopts the value of the harmonic mean of all occurring orbital exponents).

From Eq. (48) one can derive an asymptotic formula for the number $I(\vartheta)$ of basis function quadruples ν, μ, χ, λ which obey the simplified pre-screening criterion:

$$Q_{\nu\mu}\, Q_{\chi\lambda} > \vartheta \tag{49}$$

or equivalently:

$$q_{\mu\nu} + q_{\chi\lambda} < \Theta = \ln(Q^2_{\max}/\vartheta). \tag{50}$$

Using the distribution function n_q, Eq. (48), we obtain after two integrations:

$$I(\vartheta) = \frac{\rho^4}{d!}\,(2\pi\Theta/\zeta)^d\, V^2, \tag{51}$$

where V is the actual spatial extent of the system. Introducing the total number of basis functions $N = \rho V$ we may rewrite Eq. (51):

$$I(\vartheta) = \rho^2\, N^2 \left(\frac{\pi}{\zeta} \ln\,(4\zeta/(\pi\vartheta^2))\right)^d \Big/ d!\,. \tag{52}$$

In a three-dimensional system and for realistic choices of parameters the asymptotic number of ERIs obeying Eq. (49) becomes $I(10^{-10}) \approx 10^5\, N^2$. The number of (untransformed) ERIs to be processed in the AO-MP2 algorithm will be below the estimate given in Eq. (52) as the more stringent screening criteria Eqs. (34–35) take into account bounds to the relevant two-particle density.

The asymptotic number of partially transformed integrals, which are screened by Eqs. (36–38), remains to be discussed. In all these inequalities there is at least one charge distribution which originates from a pair of CGTFs, either in a partially transformed ERI, or in a Q matrix element. Thus we expect at worst $\sim N^3$ partially transformed integrals in the asymptotic limit. This could be reduced to $\sim N^2$ partially transformed integrals, if the transformed CGTFs $\bar{\nu}$ (and/or $\underline{\mu}$) are well localized.

4.2 Localization of the transformed basis functions $\bar{\nu}$, $\underline{\mu}$ in non-metallic crystals

The localization of transformed CGTFs $\bar{\nu}$ will be guaranteed if the absolute values of the matrix elements $\bar{D}_{\nu\mu}$, Eq. (15), decrease exponentially with the distance between basis functions ν and μ. This would indeed be true if $\bar{D}$ were the density matrix of a non-metallic crystal [23]. Since $\bar{D}$ becomes proportional to the density matrix for $t \to 0$ one may expect for extended (non-metallic) systems that the transformed basis functions $\bar{\nu}$ remain localized for $t > 0$. This will subsequently be demonstrated for a crystal with non-degenerate band structure [24].

Our assumptions are:

(1) The Hartree–Fock Bloch functions can be expanded in terms of a limited number (per unit cell) of atomic basis functions $v_n(r)$:

$$\varphi_h(K, r) = \sum_{v,n} v_n(r)\, C_{v_n h}(K), \tag{53}$$

where n is a direct lattice vector, K is a wave vector in the (first) Brilloin zone BZ, h is a completely occupied band, and the basis set expansion coefficients $C_{v_n h}(K)$ are solutions of the SCF equations [25]:

$$\sum_{\mu,m} F_{v_n \mu_m} C_{\mu_m h}(K) = \left(\sum_{\mu,m} S_{v_n \mu_m} C_{\mu_m h}(K) \right) \varepsilon_h(K). \tag{54}$$

(2) The coefficients $C_{v_n h}(K)$ exhibit the translational (symmetry) property:

$$C_{v_{n+m} h}(K) = C_{v_n h}(K)\, e^{iKm}. \tag{55}$$

(3) The coefficients $C_{v_n h}(K)$ are periodic and analytic functions of the wave vector K, that is, they can be represented by a locally convergent Taylor series in terms of the components of K at every point of the non-degenerate band h [26].

We then observe that the left hand side of Eq. (54) and the expression in brackets define two sets of analytic functions $f_{v_n h}(K)$ and $s_{v_n h}(K)$, provided the infinite summations converge uniformly in a complex (!) neighborhood of BZ. We briefly prove uniform convergence: $C_{v_n h}(K)$, $|n| \to \infty$, may only increase as $(1 + \delta)^{|n|}$, where $\delta > 0$ relates to the imaginary part of K, Eq. (55), and can be chosen as small as desired. Uniform convergence will then be guaranteed for a small enough imaginary part of K, provided $S_{v_n \mu_m}$ and $F_{v_n \mu_m}$ decrease exponentially with $|n - m| \to \infty$. The latter is certainly true for the overlap matrix and for the local part of the Fock matrix. In the absence of partially filled bands the non-local exchange part of the Fock matrix also decreases exponentially since the density matrix exhibits exponential decay in this case [23].

Analyticity of $f_{v_n h}(K)$ and $s_{v_n h}(K)$ thus proven we note (the asterisk indicates complex conjugation):

$$\sigma_h(K) = \sum_{v} C^*_{v_0 h}(K^*)\, s_{v_0 h}(K) \neq 0 \tag{56}$$

for all K in a neighborhood of BZ. The opposite would imply that either the basis set is linearly dependent, or $\varphi_h(K) \equiv 0$. We may therefore conclude that the orbital energies:

$$\varepsilon_h(K) = \left(\sum_{v} C^*_{v_0 h}(K^*) f_{v_0 h}(K) \right) \Big/ \sigma_h(K) \tag{57}$$

are analytic.

This result does not hold for metallic systems as is well known [27]: the exchange contribution to $F_{v_n \mu_m}$ does not decay exponentially for $|n - m| \to \infty$.

As the functions $C_{\nu_n h}(K)$ and $\varepsilon_h(K)$ are analytic (and periodic) this also applies to:

$$\bar{C}^{(\alpha)}_{\nu_n h}(K) = |w_\alpha|^{1/8}\, C_{\nu_n h}(K)\exp((\varepsilon_h(K) - \varepsilon_F)t_\alpha/2), \tag{58}$$

and to:

$$\bar{d}^{(\alpha)}_{\nu_n \mu_m}(K) = \sum_h \bar{C}^{(\alpha)}_{\nu_n h}(K)\,\bar{C}^{(\alpha)*}_{\mu_m h}(K^*). \tag{59}$$

The constants w_α, t_α, ε_F have been defined in Sect. 2.1.

Integration over the first Brilloin zone yields the crystal analogue of $\bar{D}^{(\alpha)}_{\nu\mu}$, Eq. (15):

$$\bar{D}^{(\alpha)}_{\nu_n \mu_m} = \frac{1}{V_{BZ}} \int d^3K\, \bar{d}^{(\alpha)}_{\nu_n \mu_m}(K). \tag{60}$$

Using the translational properties of Bloch functions, Eq. (55), in conjunction with Eqs. (58, 59) we observe:

$$\bar{D}^{(\alpha)}_{\nu_n \mu_{n+m}} = \frac{1}{V_{BZ}} \int d^3K\, \bar{d}^{(\alpha)}_{\nu_n \mu_n}(K)\, e^{-iKm}. \tag{61}$$

Obviously this integral defines $\bar{D}^{(\alpha)}_{\nu_n \mu_{n+m}}$ to be a Fourier coefficient of function $\bar{d}^{(\alpha)}_{\nu_n \mu_n}(K)$.

By virtue of a theorem on Fourier coefficients of periodic and analytic functions [28] we finally conclude that there is a positive constant η with:

$$\lim_{|m| \to \infty} |\bar{D}^{(\alpha)}_{\nu_n \mu_{n+m}}| \exp(\eta |m|) = 0, \tag{62}$$

that is, $\bar{D}^{(\alpha)}_{\nu_n \mu_m}$ decreases exponentially as $|n - m| \to \infty$, and the transformed basis functions $\bar{\nu}$, Eq. (17), are localized.

For the second type of transformed basis functions $\underline{\nu}$, Eq. (17), similar asymptotic characteristics may be inferred within the finite basis set (per unit cell) approximation.

It has thus been demonstrated that there are only $\sim N^2$ partially transformed integrals to be processed – at least in very large systems. Moreover, as matrix elements $\bar{D}_{\nu\mu}$ and $\underline{D}_{\nu\mu}$ decrease exponentially with distance between ν and μ, only a limited number of them need be considered in a transformation step: the $\sim N^5$ transformation steps of conventional MP2 theory ultimately become $\sim N^2$ transformation steps in the AO-MP2 approach as we consider very large molecular systems with localizable electronic structure.

With respect to solid state MP2 calculations we note that MP2 energy contributions of distant charge distributions, e.g. $(\bar{\nu}\mu|\bar{\chi}\lambda)\,(\nu\underline{\mu}|\chi\underline{\lambda})$ with basis function pair $(\nu\mu)$ well separated from $(\chi\lambda)$, may be evaluated by multipole expansion and direct lattice sums (additional considerations are necessary to enforce convergence). Alternatively one may exploit the Fourier integral over $\bar{C}^{(\alpha)}_{\nu_n h}(K)$, Eq. (58), which defines a localized (exponentially decaying) Wannier type function. The latter approach is more in the spirit of Almlöf's original formulation of Laplace transform MP2 theory [17]. The leading term in a multipole expansion involving products of localized Wannier-type orbitals will be $\sim 1/R^6$. Probably there is a relation to molecular van der Waals coefficients which allows further reductions in the operation count.

5 Applications: efficiency and over-all accuracy of the AO-MP2 ansatz

Two parameters in the AO-MP2 ansatz determine the accuracy and the computational cost of a calculation: the number τ of exponentials used in the discretized Laplace integration, Eq. (7), and the screening threshold ϑ in Eqs. (34–39). In this section we will monitor for representative applications:

(1) the over-all accuracy of the MP2 correlation energy as a function of τ and ϑ;
(2) the number of (partially transformed) integrals admitted by the screening formulas, Eqs. (34–39);
(3) the distribution of magnitudes of the matrix elements of Q, X, Y, and Z, Eqs. (25–27, 30–33), which are crucial for efficient screening.

We will demonstrate that any pre-determined accuracy can be accomplished by proper choices of τ and ϑ, and that in systems the size of porphyrin ($C_{20}H_{14}N_4$) the number of (partially transformed) integrals passing the bounds, Eqs. (34–39), grows at most with the third power of the size of the system when compared to similar, but smaller molecules (same screening threshold ϑ).

In a previous paper [18] we studied the performance of the Laplace transform MP2 ansatz, Sect. 2.1, for the medium sized molecule *p*-chloro-phosphabenzene C_5H_4ClP using a (8*s*4*p*1*d*; 4*s*1*p*; 11*s*6*p*1*d*; 11*s*6*p*1*d*)/[4*s*2*p*1*d*; 2*s*1*p*; 5*s*4*p*1*d*; 5*s*4*p*1*d*] basis set (abbreviated DZP) with a total of 148 cartesian CGTFs. The core orbitals were frozen. Here we use the same molecule to test the AO-MP2 ansatz.

Due to a wide spread of orbital energies, $x_{\min} = 0.77$, $x_{\max} = 358.1$ in Eq. (8), a comparatively large number of terms, $\tau = 8$, is needed in the functional approximation, Eqs. (7, 8), to attain micro-hartree accuracy in the MP2 energy [18]. The parameters w_α, t_α are given in Table 3 of [18].

Although our experimental AO-MP2 program does not explicitly exploit point group symmetry, the planarity of C_5H_4ClP means the CGTFs accidentally form a symmetry-adapted basis with respect to the molecular plane. Eqs. (35–38) thus screen off about 50% of the ERIs just on symmetry grounds. This will be taken into consideration in subsequent comparisons.

In Table 1 we show how variation of the parameters τ (number of exponentials) and ϑ (neglect threshold, Eqs. (34–39)) affects the approximate MP2 energy, Eqs. (11, 19), and how it affects the number of half-transformed integrals written to direct access file according to Eq. (37).

While in all calculations on C_5H_4ClP errors in the MP2 energy E are below 1% (at the same time 94% of the half-transformed integrals are screened off) the computation costs grow by a factor of ten as the over-all accuracy is increased to almost μHartree accuracy ($\vartheta = 10^{-10}$). At the highest accuracy ($\vartheta = 10^{-10}$, $\tau = 8$) the computation time was about 44 h on an IBM RISC/6000-320 workstation computer: this is about a factor of 50–100 more than what would be expected for an efficient conventional semi-direct MP2 program.

It is important to note that the error which arises from integral screening, termed '*SE*' in Table 1, changes sign as ϑ is varied. Its magnitude agrees with a statistical superposition of small errors due to neglect of individual ERIs. This can be concluded, e.g. from the difference in the numbers of half-transformed integrals for $\vartheta = 10^{-9}$ and $\vartheta = 10^{-10}$, $21 \cdot 10^6$, and from the difference in *SE*, $2 \cdot 10^{-5}$ (the total screening errors in each e_α, $\alpha = 1 \ldots \tau$, probably add up systematically in *SE*).

Table 1. MP2 energies E calcualted by the AO-MP2 ansatz for p-chloro-phospha-benzene C_4H_5ClP (DZP basis sets) for different integral screening parameters ϑ, Eqs. (34–39). The exact result would be $-.911294$ Hartree. τ is the number of exponentials used in the Laplace integration, Eqs. (7, 11). $L1$, Eq. (9), is an a priori measure for the expected relative error in E due to the discretization of the Laplace transform. The observed error [18] from the discretization of the Laplace transform is LE (absolute error). SE is the error originating from integral screening. Also given is the maximum number (in million and in %) of half-transformed integrals $(\bar{\nu}\underline{\mu}|\chi'\lambda')$ written to direct access file according to Eq. (37), and its a priori estimate based on Eq. (40) (the maximum is taken from $\alpha = 1 \mathinner{..} \tau$)

ϑ	τ	$L1$	E	LE	SE	est. # $(\bar{\nu}\underline{\mu}\|\chi'\lambda')$	# $(\bar{\nu}\underline{\mu}\|\chi'\lambda')$
10^{-6}	8	2.7E − 6	− .908822	+ .000001	+ .002471	29 (11%)	13 (5%)
10^{-7}	8	2.7E − 6	− .911316	+ .000001	− .000023	51 (21%)	21 (9%)
10^{-8}	8	2.7E − 6	− .911332	+ .000001	− .000039	78 (31%)	32 (13%)
10^{-9}	8	2.7E − 6	− .911271	+ .000001	+ .000022	104 (42%)	46 (19%)
10^{-10}	8	2.7E − 6	− .911292	+ .000001	+ .000001	125 (51%)	61 (25%)
10^{-9}	7	1.5E − 5	− .911264	+ .000012	+ .000018	117 (48%)	47 (19%)
10^{-8}	6	4.5E − 5	− .911253	+ .000040	+ .000001	79 (32%)	33 (14%)
10^{-7}	5	2.8E − 4	− .911262	+ .000214	− .000182	54 (22%)	23 (9%)
10^{-6}	3	1.5E − 3	− .906521	+ .001377	+ .003396	34 (13%)	14 (6%)

Table 2. Number of ERIs included in an AO-MP2 calculation of p-chloro-phospha-benzene with $\vartheta = 10^{-10}$ and $\tau = 8$ according to Eqs. (34–39). The integral numbers have been termed ERI34–ERI39 respectively and are given in % (their total numbers cannot be compared directly due to different canonical symmetry). The numbers ERI34 and ERI39 are somewhat larger than the rest since in these steps whole integral batches have to be considered at once. Also shown are the energy contributions e_α arising from each exponential term $\alpha = 1 \mathinner{..} \tau$, Eqs. (7, 11, 16)

α	e_α	ERI34	ERI35	ERI36	ERI37	ERI38	ERI39
1	− .026270	57%	19%	23%	20%	16%	43%
2	− .101457	61%	21%	26%	23%	21%	47%
3	− .246319	62%	22%	28%	25%	23%	49%
4	− .277971	61%	22%	27%	25%	24%	49%
5	− .180737	59%	20%	25%	23%	22%	47%
6	− .065095	55%	18%	22%	20%	19%	43%
7	− .012326	50%	14%	16%	15%	13%	35%
8	− .001118	37%	8%	7%	6%	4%	21%

For any required accuracy in the MP2 correlation energy we are thus able to choose optimal values of

(1) τ by means of the a priori error estimates $L1$ (also shown in Table 1) and $L2$, Eqs. (9, 10) [18], and of

(2) ϑ by assuming a statistical superposition of small screening errors; their number is known a priori from Eq. (40).

In Table 2 we compare the performance of our screening procedure for all types of (partially transformed) ERIs. All bounds screen about the same percentage of ERIs, except where integrals have to be processed and screened in batches. Not all MP2 energy contributions e_α, $\alpha = 1 \dots 8 = \tau$, come at the same price: $\alpha = 7$ and $\alpha = 8$, which correspond to high exponential coefficients of $t_7 = 2.7$ and $t_8 = 5.4$,

Table 3. Size distributions of $|\bar{D}_{\nu\mu}|$, $|\underline{D}_{\nu\mu}|$, $Q_{\nu\mu}$, $X_{\nu\mu}$, $Y_{\nu\mu}$, and $Z_{\nu\mu}$, Eqs. (15, 16, 25, 26, 27, 33), in *p*-chloro-phospha-benzene (core electrons frozen). Given is the percentage of matrix elements in the intervals $[10^{-k}, 10^{-k+1}]$ for $k = 0 \ldots 8$, and the percentage of matrix elements smaller than 10^{-8}. Below the same numbers are given for porphyrin $C_{20}H_{14}N_4$ (core electrons frozen)

k	0	1	2	3	4	5	6	7	8	smaller
$\lvert\bar{D}_{\nu\mu}\rvert$	0.0%	0.7%	7.7%	19.3%	18.8%	9.9%	2.4%	0.2%	0.0%	41.0%
$\lvert\underline{D}_{\nu\mu}\rvert$	5.4%	12.5%	17.6%	15.6%	6.8%	1.2%	0.1%	0.0%	0.0%	41.0%
$Q_{\nu\mu}$	0.7%	21.2%	22.2%	14.9%	9.5%	5.5%	4.2%	4.2%	3.2%	14.3%
$X_{\nu\mu}$	0.0%	2.8%	29.6%	41.8%	21.5%	4.0%	0.2%	0.1%	0.0%	0.0%
$Y_{\nu\mu}$	0.1%	17.3%	39.2%	33.8%	8.9%	0.7%	0.0%	0.0%	0.0%	0.0%
$Z_{\nu\mu}$	0.0%	8.4%	38.2%	36.9%	15.6%	1.0%	0.0%	0.0%	0.0%	0.0%
$\lvert\bar{D}_{\nu\mu}\rvert$	0.0%	0.2%	3.4%	12.2%	16.9%	15.3%	9.2%	3.7%	0.8%	38.3%
$\lvert\underline{D}_{\nu\mu}\rvert$	4.9%	9.4%	15.3%	16.1%	12.1%	3.5%	0.5%	0.0%	0.0%	38.2%
$Q_{\nu\mu}$	0.2%	9.3%	10.2%	8.6%	6.3%	5.9%	6.1%	5.5%	3.9%	44.1%
$X_{\nu\mu}$	0.0%	1.1%	13.0%	31.0%	30.5%	18.4%	5.7%	0.2%	0.0%	0.0%
$Y_{\nu\mu}$	0.2%	9.9%	24.6%	32.1%	28.7%	4.5%	0.1%	0.0%	0.0%	0.0%
$Z_{\nu\mu}$	0.0%	6.5%	25.0%	36.9%	23.6%	7.6%	0.3%	0.0%	0.0%	0.0%

Table 4. MP2 energies E calculated by the AO-MP2 ansatz for P_7H_3 with an SVP basis set for different integral screening parameters ϑ, Eqs. (34–39), and for different numbers τ of exponentials used in the Laplace integration, Eqs. (7, 11). The exact result would be $E = -.808210$ Hartree. $L1$, Eq. (9), is an a priori measure for the expected relative error in E due to the discretization of the Laplace transform. The observed error $LE + SE$ results from the discretization of the Laplace transform and from integral screening. Also given is the maximum number (in million and in %) of half-transformed integrals $(\bar{\nu}\underline{\mu}|\chi'\lambda')$ written to direct access file according to Eq. (37), and its a priori estimate based on Eq. (40) (the maximum is taken from $\alpha = 1 \,..\, \tau$)

ϑ	τ	$L1$	E	$LE + SE$	est. # $(\bar{\nu}\underline{\mu}\mid\chi'\lambda')$	# $(\bar{\nu}\underline{\mu}\mid\chi'\lambda')$
10^{-6}	2	3.0E − 3	−.809774	−.001564	63 (25%)	26 (10%)
10^{-7}	3	1.9E − 4	−.808815	−.000605	92 (37%)	44 (18%)
10^{-8}	4	1.3E − 5	−.808223	−.000013	118 (48%)	68 (28%)
10^{-9}	5	8.6E − 7	−.808199	+.000011	144 (58%)	97 (40%)
10^{-10}	6	5.9E − 8	−.808208	+.000002	164 (67%)	128 (52%)

Eq. (7), are comparatively inexpensive as can best be rationalized on the grounds of Eqs. (15, 30).

Table 3 provides the size distribution of all two-index quantities involved in transformations and screening for *p*-chloro-phospha-benzene. Remember that the correlation energy contribution of each ERI is bound by a product of *four* matrix elements from Q, X, Y, or Z. It is the spread of the distributions which makes integral screening efficient.

For comparison we also studied ERI screening and resulting accuracies for the more globular, compact P_7H_3 molecule. Core orbitals have not been correlated. With a (10*s*7*p*1*d*; 4*s*1*p*)/[4*s*3*p*1*d*; 2*s*1*p*] basis set (split valence plus polarization or SVP), which yields a total of 148 cartesian CGTFs, we observe similar accuracies as for the planar C_5H_4ClP, compare Tables 1 and 4. For the SVP basis set in P_7H_3, however, the spread of orbital energies is narrower than for a DZP basis set in

C_5H_4ClP, and a smaller number τ of terms in the exponential approximation, Eq. (7), suffices to attain similar accuracies in the Laplace transform MP2 energies as can be seen from a comparison of the Laplace integration error estimates $L1$ in Tables 1 and 4. At the same total accuracy the number of half-transformed integrals to be processed in P_7H_3 is twice that in C_5H_4ClP: this is largely a consequence of planarity in the latter system.

To ascertain the effects of growing molecular size we have chosen the planar porphyrin $C_{20}H_{14}N_4$ (SCF equilibrium geometry) to be our last test molecule. The basis set employed is (*7s4p1d*; *4s1p*; *7s4p1d*)/[*3s2p1d*; *2s1p*; *3s2p1d*] (SVP), which totals 430 CGTFs. Porphyrin has an extended π-system and may be among the more difficult large molecules to study as there is electron delocalization and exchange interactions may be long ranged (cf. Sect. 4.2). This would also pose problems to localization based correlation approaches.

In Table 3 the distributions of matrix element magnitudes of Q, X, Y, and Z, Eqs. (25, 30–33), can be compared for the systems *p*-chloro-phospha-benzene and porphyrin. The largest difference appears in Q as the logarithms of its matrix element magnitudes approach a uniform distribution in porphyrin. This is already close to what can be expected for a large planar molecule in the asymptotic limit according to Sect. 4.1. One would therefore expect at worst $\sim N^3$ of the partially transformed integrals to pass the screening formulas, Eqs. (36–38). The other size distributions (for X, Y, and Z) gain up to a decade in width as compared to *p*-chloro-phospha-benzene. This may help to reduce the number of partially transformed integrals, but is insufficient to reach the $\sim N^2$ asymptotic integral count at this molecular size.

The changes in the size distributions of $X_{\nu\mu}$ and of $Y_{\nu\mu}$ are in phase with those of $|\bar{D}_{\nu\mu}|$ and $|\underline{D}_{\nu\mu}|$, respectively, Table 3. However, besides symmetry zeros in $\bar{D}$ and $\underline{D}$, sparsity in these matrices is not sufficient to reward running innermost transformation loops only over significant matrix elements of $\bar{D}$ and $\underline{D}$ (as may be achieved by pre-sorting and indirect addressing).

To test the efficiency of integral screening for porphyrin we performed an all-electron AO-MP2 calculation with parameters $\tau = 1$ and $\vartheta = 10^{-9}$. The number of CGTF ERIs to be evaluated for the first quarter-transformation step was $1.7 \cdot 10^9$ (19%) due to screening by Eq. (34) (adapted to integral batches). After evaluation only 5% of the CGTF ERIs were transformed as a consequence of Eq. (35) (applied to *individual* ERIs). Screening of quarter-, half-, and three-quarter-transformed integrals by Eqs. (36–38) admitted only 7%, 5%, and 4% of the ERIs into the subsequent transformation steps, respectively.

In porphyrin integral screening in the transformation steps thus improves performance by a factor of 20. However, only a factor of 10 is due to the size of the system while planarity accounts for a factor of 2. The total number of half-transformed integrals written to disc was $9 \cdot 10^8$.

A comparison to *p*-chloro-phospha-benzene shows that while we tripled the number of basis functions the performance gain due to integral screening increased by a factor of 4.

We have thus left the N^5 dependence of the operation count on the size of the system, and have entered the N^4 regime. Unfortunately this is not sufficient to win over an optimal conventional MP2 algorithm for which the number of half-transformed integrals $(i\nu|j\mu)$ is $n^2N^2/4 = 3 \cdot 10^8$ (we included a factor of 1/2 for planarity). Considering that we have chosen $\tau = 1$ while $\tau = 5$ would have been adequate (this translates into a factor of ≈ 3.5 in computation time, see Table 2), there is still a factor of about ten in performance we need to gain.

We conclude: We have shown that rigorous MP2 correlation energies can be evaluated by the AO-MP2 ansatz with an operation count $\sim N^4$ in systems like porphyrin $C_{20}H_{14}N_4$, 430 CGTFs. However, the AO-MP2 ansatz is a factor of about 100 more costly in small molecules than the conventional MP2 approach, and in porphyrin this difference is reduced by a factor of 10. Thus our current implementation of the AO-MP2 ansatz is not competitive if μHartree accuracy is required. At this accuracy we would expect the break-even point at about 200 atoms (provided the size distributions of X, Y, and Z gain further width). In the next section we look at possible improvements of the AO-MP2 ansatz which may help to attain the break-even point at a smaller molecular size.

6 Further improvements

The AO-MP2 ansatz could be more efficient, if the matrices $\bar{D}$ and $\underline{D}$, Eq. (15), were sparse even in medium sized molecules. Their appearance definitely hinges on the basis set we use: if we switch from the CGTF representation to the canonical SCF MO basis, we obtain diagonal matrices $\bar{D}$ and $\underline{D}$, and we are back at the original Laplace transform MP2 ansatz, Eqs. (11–13). However, with this drastic transition we loose most of our integral screening potential, and a number of other advantages of the AO-MP2 ansatz as well, cf. Sect. 3.3.

Fortunately there is a smooth transition between those two extreme approaches which probably allows to balance their respective advantages and disadvantages. A well-known example from within this transition region are symmetry-adapted basis functions [29], but this will not be our concern here.

Let us partition our CGTF basis into sub-sets and corresponding sub-spaces M (the identifier M will be used for both) as in Sects. 3.2 and 3.3. Such a sub-set M could, for instance, comprise all CGTFs which are centered at the same atom, or all basis functions which belong to a small group of atoms like a methyl group or a carbonyl group. This will be our choice.

Next we introduce a new basis $B(M)$ in sub-space M which for our computational convenience is orthonormal in M. Basis functions from different subspaces M and M' would usually not be orthogonal. The AO-MP2 ansatz may now be formulated in terms of the new molecular basis set $B = \cup B(M)$, which thus consists of modified atomic orbitals (AO) or group orbitals (GO). Subsequently we will use the acronym AO for both.

We may also use a mixture of basis sets, e.g. the original CGTF basis for those functions which will subsequently be transformed by $\underline{D}$, and the modified AO basis set B otherwise, cf. Eq. (17).

Which particular choice of B offers the highest rewards in such a mixed scheme? If our primary CGTF basis is large in the sense that it is designed to retrieve a high proportion of the dynamical correlation energy (this is a case against the AO-MP2 ansatz), we can expect that $\bar{D}$, which corresponds to the occupied–occupied block of the exponential Fock matrix, can efficiently be represented by a sub-set of B. This will be our aim.

We express the occupied orbitals in our as yet unoptimized basis $\beta, \gamma, \delta \in B$:

$$|i\rangle = \sum_{\beta \in B} |\beta\rangle C^{(B)}_{\beta i}, \tag{63}$$

and obtain for matrix $\bar{D}^{(\alpha)}$, Eq. (15), the new representation:

$$\bar{\Delta}_{\beta\gamma} = |w_\alpha|^{1/4} \sum_i C^{(B)}_{\beta i} \exp((\varepsilon_i - \varepsilon_{\mathrm{F}}) t_\alpha) C^{(B)}_{\gamma i}. \tag{64}$$

Basis B will be optimal if $\bar{\Delta}$ has as many near-zero rows and columns as possible (or has many rows and columns with near-zeros everywhere with the possible exception of the (M, M) diagonal blocks). In other words we choose each sub-basis $B(M)$ so as to diagonalize the positive semi-definite matrix:

$$\bar{d}(M)_{\beta\delta} = \sum_{\gamma \in B'} \bar{\Delta}_{\beta\gamma} \bar{\Delta}_{\gamma\delta}, \tag{65}$$

where β, $\delta \in B(M)$, and $B' = B$ (or $B' = B \backslash B(M)$). The near-zero eigenvalues of $\bar{d}(M)$ would relate, for example, to atomic orbitals (or group orbitals) with a large number of nodes which do not represent the occupied space, but are only involved in the description of dynamical correlation effects. In the limiting case, where the CGTF basis is partitioned into only one sub-space M, the eigenvalues of $\bar{d}(M)$ would be zero (virtual orbitals) or equal to $\exp(2(\varepsilon_i - \varepsilon_{\mathrm{F}})t_\alpha)$ (occupied orbitals).

We have not yet tested this modified AO approach, but we may refer the reader to similar considerations in a very different context: modified atomic orbitals (MAO) have previously been used in population analyses, and it has been shown that a *small* number of carefully optimized MAOs at each atom (corresponding to a single zeta description) can to a good approximation serve as a basis of a SCF one-electron density originally expressed in a much larger basis set (the 'unassigned charge' is usually well below 1%) [30].

The calculation of mixed CGTF/MAO ERIs is certainly not a problem as similar ERIs have efficiently been implemented in context with atomic natural orbital (ANO) generalized contraction basis sets [31, 32].

7 Conclusions

With the AO-MP2 ansatz we have presented a novel formulation of MP2 theory. Starting point of our approach was the Laplace transform MP2 ansatz [17, 18]. Subsequently we moved into an atomic orbital (AO) basis representation and eliminated the need to calculate electron repulsion integrals for molecular orbitals. For the first time we calculated MP2 correlation energies of medium-sized molecular in an AO basis.

Our aim has been to remove all bottlenecks which hinder an application of MP2 theory to large molecular systems:

(1) We demonstrated in theory, Sect. 4, and in applications, Sect. 5, that the operation count of the AO-MP2 ansatz grows at most with the fourth power of the size of the system, and *not* with the fifth power as in conventional MP2 theory. This has been made possible by strict bounds, Sect. 2.3, which eliminate numerically insignificant four-index quantities from all intermediate operations. This elimination rate approached 95% in an application to the porphyrin molecule $C_{20}H_{14}N_4$. Some of these findings also have implications with regard to standard MP2 programs, cf. Sect. 3.2.

(2) In systems with localizable electronic structure the asymptotic operation count may be as low as $\sim N^2$ (Sect. 4.2). An application to (non-metallic) crystals seems to be a natural extension of the AO-MP2 ansatz.

(3) In Sect. 3.3 we devised an AO-MP2 algorithm which allows to perform molecular MP2 calculations in-core with only $\sim N^2$ memory requirements.

(4) The same algorithm is ideally suited to be implemented on massively parallel computers and will run with high efficiency as virtually no input/output or inter-processor communication is necessary, provided that each individual processor can store several N^2 matrices (we recall that N is the number of basis functions).

In the appendix we outline the formulation of MP2 gradients within the framework of the AO-MP2 ansatz, and there is good reason to believe that points (1)–(4) above also apply to AO-MP2 derivatives.

However, we failed one objective: our experimental AO-MP2 program does not outperform a conventional (semi-direct) MP2 program in molecular systems with 100 correlated electron pairs or less. The reason is that in small molecules the AO-MP2 ansatz starts off with about a factor of 100 more computational operations than MO-based MP2 algorithms. If μHartree accuracy is required we estimate that computation times superior to MO-based algorithms will be achieved with the present AO-MP2 algorithm for molecular systems of 200 atoms. For moderate quality basis sets (like 'split valence plus polarization') this is a conservative estimate, and there are probably smaller systems, e.g. molecular clusters interacting through hydrogen bonds, for which the AO-MP2 ansatz is already competitive. High quality basis sets, on the other hand, will decrease the competitivity of the straight AO-MP2 ansatz.

With the introduction of modified atomic orbitals or modified group orbitals in Sect. 6 we hope to eliminate this drawback, and aim to outperform traditional MO-based MP2 algorithms already in applications to molecular systems only slightly larger than porphyrin. Other conceivable means to improve the operation count are auxiliary basis sets for charge distributions [19, 33] or numerical integration techniques [34].

Ten years time have passed since Almlöf et al. suggested the 'direct' SCF method in 1982 [7], designed for an application to large molecules. At first this approach was not competitive, but this changed as the meaning of 'large' changed. As computational hardware seems to become ever more powerful it does not take much prophecy to foretell that 'large' will again change meaning in electronic structure theory in the next ten years. The AO-MP2 ansatz is ideally suited for the computer generations to come, and might become a standard approach for the evaluation of correlation energies and correlated properties within the framework of Møller–Plesset (MP2) perturbation theory in systems larger than porphyrin.

Appendix: AO-MP2 derivatives

The first expressions for conventional MP2 gradients have been given in [35]. Here we derive expressions for the derivatives E^{ξ} of the MP2 energy E with respect to a set Ξ of (real) perturbations ξ in the AO-MP2 formalism.

We adopt the convention that a superscript ξ indicates a total derivative with respect to the perturbation ξ while a superscript (ξ) means a partial derivative which does not apply to SCF molecular orbital coefficients and orbital energies. To keep our notation short we will also imply implicit summation over indices which occur in more than one factor (not counting orbital energies ε_p) of a product expression. These conventions are illustrated by the familiar coupled-perturbed Hartree–Fock (CPHF) equations [36]:

$$F^{(\xi)}_{pq} - S^{(\xi)}_{pq}\,\varepsilon_q + (\varepsilon_p - \varepsilon_q)U^{\xi}_{pq} + A_{pqai}\,U^{\xi}_{ai} - \tfrac{1}{2}A_{pqji}\,S^{(\xi)}_{ji} = \varepsilon^{\xi}_p\,\delta_{pq}. \tag{A1}$$

which together with the orthogonality constraint:

$$U^{\xi}_{pq} + U^{\xi}_{qp} + S^{(\xi)}_{pq} = 0 \tag{A2}$$

determine the first-order response of the SCF MO coefficients with respect to a perturbation ξ:

$$C^{\xi}_{\nu p} = C_{\nu q}\,U^{\xi}_{qp}. \tag{A3}$$

Here and in the following equations p, q, r, s label all SCF MOs (occupied and virtual), and:

$$A_{pqrs} = 4(pq|rs) - (ps|rq) - (pr|sq). \tag{A4}$$

With Eq. (A1) we have chosen a formulation of the CPHF equations which retains the canonicity of the perturbed MOs.

Our aim is to express the MP2 correlation energy gradient as:

$$\begin{aligned} E^{\xi} &= E^{(\xi)} + \Phi_{pq}\,F^{(\xi)}_{pq} - \Sigma_{pq}\,S^{(\xi)}_{pq} + \Omega_{ai}\,U^{\xi}_{ai} \\ &= -\int_0^{\infty} dt(e^{(\xi)}(t) + \varphi_{pq}(t)\,F^{(\xi)}_{pq} - \sigma_{pq}(t)\,S^{(\xi)}_{pq} + \omega_{ai}(t)\,U^{\xi}_{ai}) \end{aligned} \tag{A5}$$

with the unknowns $E^{(\xi)}$, Φ, Σ, and Ω all evaluated within the AO-MP2 ansatz. As in conventional MP2 gradient schemes the solutions U^{ξ}_{ai} of the CPHF equations are not needed in order to evaluate E^{ξ} from Eq. (A5) since they can be eliminated by the Handy–Schaefer trick [37] after the 'Lagrangian' Ω has been determined. This elimination step amounts to the solution of only one CPHF type equation (also called the Z-vector equation) which for large molecules is best carried out in the CGTF basis in the spirit of direct SCF techniques [38].

We start our derivation of Eq. (A5) by taking the derivative of Eq. (4):

$$E^{\xi} = -\int_0^{\infty} dt\; e^{\xi}(t). \tag{A6}$$

We recall that within the AO-MP2 ansatz:

$$e(t) = (\nu'\mu'|\chi'\lambda')\,\bar{D}_{\nu'\nu}\,\underline{D}_{\mu'\mu}\,\bar{D}_{\chi'\chi}\,\underline{D}_{\lambda'\lambda}[2(\nu\mu|\chi\lambda) - (\nu\lambda|\chi\mu)], \tag{A7}$$

with

$$\bar{D}_{\nu\mu} = \bar{D}_{\nu\mu}(t) = C_{\nu i}\exp(\varepsilon_i t)\,C_{\mu i} \tag{A8}$$

and

$$\underline{D}_{\nu\mu} = \underline{D}_{\nu\mu}(t) = C_{\nu a}\exp(-\,\varepsilon_a t)\,C_{\mu a}, \tag{A9}$$

in close analogy to Eqs. (15, 16), but with two differences: (a) We set $\varepsilon_F = 0$ for simplicity. (b) We still use the Laplace transform scheme without discretization by means of Eq. (7).

The total derivative of $e(t)$, Eq. (A7), with respect to ξ is:

$$e^{\xi}(t) = e^{(\xi)}(t) + 2\bar{R}_{\lambda\lambda'}\,\underline{D}^{\xi}_{\lambda\lambda'} + 2\underline{R}_{\chi\chi'}\,\bar{D}^{\xi}_{\chi\chi'}, \tag{A10}$$

where the first term:

$$\begin{aligned} e^{(\xi)}(t) &= 2(\nu'\mu'|\chi'\lambda')\,\bar{D}_{\nu'\nu}\,\underline{D}_{\mu'\mu}\,\bar{D}_{\chi'\chi}\,\underline{D}_{\lambda'\lambda}[2(\nu\mu|\chi\lambda)^{\xi} - (\nu\lambda|\chi\mu)^{\xi}] \\ &= 2(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})\,[2(\nu\mu|\chi\lambda)^{\xi} - (\nu\lambda|\chi\mu)^{\xi}] \end{aligned} \tag{A11}$$

already yields $E^{(\xi)}$, Eq. (A5), after subsequent t-integration. Obviously $E^{(\xi)}$ can be obtained by the same algorithms as E, Sects. 3.1 and 3.3, at a similar computational cost (the only difference is between $(\nu\mu|\chi\lambda)$ and $(\nu\mu|\chi\lambda)^{\xi}$).

The other terms in Eq. (A10) are:

$$\bar{R}_{\lambda\lambda'} = (\bar{\nu}\underline{\mu}|\bar{\chi}\lambda')\,[2(\nu\mu|\chi\lambda) - (\nu\lambda|\chi\mu)], \tag{A12}$$

$$\underline{R}_{\lambda\lambda'} = (\bar{\nu}\underline{\mu}|\lambda'\underline{\chi})\,[2(\nu\mu|\lambda\chi) - (\nu\chi|\lambda\mu)], \tag{A13}$$

$$\begin{aligned} \bar{D}^{\xi}_{\nu\mu} &= (C_{\nu a}\,C_{\mu i} + C_{\nu i}\,C_{\mu a})\exp(\varepsilon_i t)\,U^{\xi}_{ai} \\ &\quad + (C_{\nu j}\,C_{\mu i} + C_{\nu i}\,C_{\mu j})\exp(\varepsilon_i t)\,U^{\xi}_{ji} \\ &\quad + C_{\nu i}\,C_{\mu i}\exp(\varepsilon_i t)\,\varepsilon^{\xi}_{j} t, \end{aligned} \tag{A14}$$

and

$$\begin{aligned} \underline{D}^{\xi}_{\nu\mu} &= (C_{\nu a}\,C_{\mu i} + C_{\nu i}\,C_{\mu a})\exp(-\varepsilon_a t)\,U^{\xi}_{ia} \\ &\quad + (C_{\nu a}\,C_{\mu b} + C_{\nu b}\,C_{\mu a})\exp(-\varepsilon_a t)\,U^{\xi}_{ba} \\ &\quad - C_{\nu a}\,C_{\mu a}\exp(-\varepsilon_a t)\varepsilon^{\xi}_{a} t. \end{aligned} \tag{A15}$$

Matrix $\bar{R}$ has already been constructed in context with the AO-MP2 in-core multiple pass algorithm, Eq. (42). Matrix $\underline{R}$ can be calculated similarly and simultaneously, provided the half-transformed integrals $(\bar{\nu}\underline{\mu}|\chi'\lambda')$ are transformed both to $(\bar{\nu}\underline{\mu}|\bar{\chi}\lambda')$ and to $(\bar{\nu}\underline{\mu}|\lambda'\underline{\chi})$ as well.

If the CPHF equations for Ξ were now solved, Eqs. (A1, A2), we could determine $\bar{D}^{\xi}$ and $\underline{D}^{\xi}$, Eqs. (A14, A15). Since the other quantities on the right side of Eq. (A10) have been shown to be available by the AO-MP2 ansatz we would thus obtain $e^{\xi}(t)$. However, we wanted to avoid the CPHF equations for the set Ξ of perturbations, and thus proceed towards Eq. (A5). On our way it will become necessary to distinguish active and frozen orbitals, but this is a mere technicality adding some extra terms to the equations. For clarity we assume that all SCF MOs will be included in the correlation treatment.

We re-write Eq. (A10) with the (symmetrized) matrices $\bar{R}$ and $\underline{R}$ transformed into the MO basis, and we use Eq. (A2) and some index symmetries to obtain:

$$\begin{aligned} e^{\xi}(t) = e^{(\xi)}(t) &+ 4(\underline{R}_{ai}\exp(\varepsilon_i t) - \bar{R}_{ai}\exp(-\varepsilon_a t))\,U^{\xi}_{ai} - 4\bar{R}_{ai}\exp(-\varepsilon_a t)\,S^{(\xi)}_{ai} \\ &+ 2\underline{R}_{ji}((\exp(\varepsilon_i t) - \exp(\varepsilon_j t))\,U^{\xi}_{ji} + \delta_{ji}\exp(\varepsilon_i t)\,\varepsilon^{\xi}_{i} t) \\ &+ 2\bar{R}_{ba}((\exp(-\varepsilon_a t) - \exp(-\varepsilon_b t))\,U^{\xi}_{ba} - \delta_{ba}\exp(-\varepsilon_a t)\,\varepsilon^{\xi}_{a} t) \\ &- 2\underline{R}_{ji}\exp(\varepsilon_j t)\,S^{(\xi)}_{ji} - 2\bar{R}_{ba}\exp(-\varepsilon_b t)\,S^{(\xi)}_{ba}. \end{aligned} \tag{A16}$$

Preparing for the removal of U^{ξ}_{ji} (and U^{ξ}_{ba}) from Eq. (A16) we introduce a function:

$$g(\varepsilon_i, \varepsilon_j, t) = \frac{\exp(\varepsilon_i t) - \exp(\varepsilon_j t)}{\varepsilon_i - \varepsilon_j}$$

$$= t \exp((\varepsilon_i + \varepsilon_j)t/2) \frac{\sinh((\varepsilon_i - \varepsilon_j)t/2)}{((\varepsilon_i - \varepsilon_j)t/2))} \tag{A17}$$

with

$$g(\varepsilon_i, \varepsilon_i, t) = t \exp(\varepsilon_i t)$$

which is continuous for near-degenerate orbitals, $\varepsilon_i \approx \varepsilon_j$, and avoids numerical problems associated with small differences between large numbers. Aided by Eqs. (A1, A17) we now remove U^{ξ}_{ji}, U^{ξ}_{ba}, ε^{ξ}_i, and ε^{ξ}_a from Eq. (A16), and obtain Eq. (A5) with:

$$\varphi_{ji}(t) = 2\underline{R}_{ji}\, g(\varepsilon_i, \varepsilon_j, t), \tag{A18}$$

$$\varphi_{ba}(t) = -\, 2\bar{R}_{ba}\, g(-\,\varepsilon_a, \,-\,\varepsilon_b, t), \tag{A19}$$

$$\varphi_{ai}(t) = \varphi_{ia}(t) = 0, \tag{A20}$$

$$\begin{aligned}\sigma_{ji}(t) = {} & 2\underline{R}_{ji}(\exp(\varepsilon_j t) + \varepsilon_j g(\varepsilon_i, \varepsilon_j, t)) + 2A_{jilk}\, \underline{R}_{lk}\, g(\varepsilon_k, \varepsilon_l, t)\\ & -\, 2A_{jibc}\, \bar{R}_{bc}\, g(-\,\varepsilon_b, \,-\,\varepsilon_c, t),\end{aligned} \tag{A21}$$

$$\sigma_{ba}(t) = 2\bar{R}_{ba}(\exp(-\,\varepsilon_b t) - \varepsilon_b g(-\,\varepsilon_a, \,-\,\varepsilon_b, t)), \tag{A22}$$

$$\sigma_{ai}(t) = \sigma_{ia}(t) = 2\bar{R}_{ai} \exp(-\,\varepsilon_a t), \tag{A23}$$

and

$$\begin{aligned}\omega_{ai}(t) = {} & 4(\underline{R}_{ai} \exp(\varepsilon_i t) - \bar{R}_{ai} \exp(-\,\varepsilon_a t)) + 2A_{aijk}\, \underline{R}_{jk}\, g(\varepsilon_k, \varepsilon_j, t)\\ & -\, 2A_{aibc}\, \bar{R}_{bc}\, g(-\,\varepsilon_b, \,-\,\varepsilon_c, t).\end{aligned} \tag{A24}$$

All terms containing the A-matrix, Eq. (A4), can of course be evaluated in the CGTF basis at the price of one (direct) SCF iteration.

We summarize: Within the framework of the continuous Laplace transform AO-MP2 approach we have derived a gradient formula, Eq. (A5). The quantities $e^{(\xi)}(t)$, $\varphi_{pq}(t)$, $\sigma_{pq}(t)$, and $\omega_{pq}(t)$ have been shown to be available by some modifications of the AO-MP2 algorithms, Sects. 3.1 and 3.3, and may increase computational expenses by about 50% as compared to the calculation of $e(t)$ alone (the multiple pass algorithm in Sect. 3.3 may become significantly more costly if derivative integrals are evaluated repeatedly to obtain $e^{(\xi)}(t)$ in-core with only N^2 memory requirements).

The only non-trivial step left is the t-integration of the matrices defined in Eqs. (A18–A24). Other than with the discretized integration of $e(t)$, which only necessitates a good approximation of $1/x$ in some finite positive interval by a basis of τ exponentials $\exp(-\, x t_\alpha)$, the presence of the function g, Eq. (A17), in $e^{\xi}(t)$ will also require all difference quotients $(1/x_1 - 1/x_2)/(x_1 - x_2)$, including differential quotients, to be well approximated by respective difference quotients of the exponential expansion, Eq. (7). While this requirement can probably be met by a sufficiently large number τ of exponential terms in the least-squares fit, Eq. (8), it may be more efficient to modify Eq. (8) so that the deviation between the first derivative of $1/x$ and the derivative of its exponential approximation is minimized, too. While this has yet to be tested one can probably expect a modest increase in τ if one aims at a similar accuracy in E^{ξ} as in E.

With regard to *second derivatives* we point out that in the AO-MP2 ansatz one can completely avoid any transformation of perturbed basis functions ν^{ξ} while they form part of four-index quantities. This removes a major obstacle to the calculation of vibrational modes at the MP2 level of theory [39].

Acknowledgements. I would like to thank Jan Almlöf, who gave me the opportunity to explore the potential of the Laplace transform MP2 ansatz. The other main ingredient of this work (integral bounds) originates from my earlier research with Reinhart Ahlrichs whom I thank for his continuing support. Michael Ehrig kindly helped me with hardware problems, and Jürgen Gauss proofread part of my manuscript. Computational resources have been supplied by the Fonds der Chemischen Industrie.

References

1. Møller C, Plesset MS (1934) Phys Rev 46:618
2. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) Ab initio molecular orbital theory. Wiley, NY
3. Gauss J (1992) Chem Phys Lett 191:614
4. Häser M, Almlöf J, Scuseria GE (1991) Chem Phys Lett 181:497
   Häser M, Almlöf J, Feyereisen MW (1991) Theoret Chim Acta 79:115
5. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Chem Phys Lett 162:165
6. Dyczmons V (1973) Theoret Chim Acta 28:307
7. Almlöf J, Faegri Jr K, Korsell K (1982) J Comput Chem 3:385
8. Head-Gordon M, Pople JA, Frisch MJ (1988) Chem Phys Lett 153:503
9. Edmiston C, Krauss M (1965) J Chem Phys 42:1119; (1966) 45:1833
10. Jungen M, Ahlrichs R (1970) Theoret Chim Acta 17:339
11. Meyer W (1977) in: Schaefer HF III (ed) Methods of electronic structure theory. Plenum, NY, p 413
12. Pulay P (1983) Chem Phys Lett 100:151; Saebø S, Pulay P (1985) Chem Phys Lett 113:13; Saebø S, Pulay P (1987) J Chem Phys 86:914
13. Saebø S, Pulay P (1988) J Chem Phys 88:1884
14. Pulay P, Saebø S (1986) Theoret Chim Acta 69:357
15. Stollhoff G, Fulde P (1977) Z Physik B26:257
    Stollhoff G, Vasilopoulos P (1986) J Chem Phys 84:2744
16. Stollhoff G, Fulde P (1980) J Chem Phys 73:4548
17. Almlöf J (1991) Chem Phys Lett 176:319
18. Häser M, Almlöf J (1992) J Chem Phys 96:489
19. Whitten JL (1973) J Chem Phys 58:4496
20. Häser M, Ahlrichs R (1989) J Comput Chem 10:104
21. Saebø S, Almlöf J (1989) Chem Phys Lett 154:83
22. Dupuis M, Rys J, King HF (1976) J Chem Phys 65:111
23. des Cloizeaux J (1964) Phys Rev 135:A685
24. These results most likely also hold for crystals with degenerate band structure; proofs, however, become significantly more technical and involved. The reader may want to consult the paper by Lix B (1971) Phys Stat Sol B44:411
25. Pisani C, Dovesi R, Roetti C (1988) Hartree Fock ab initio treatment of crystalline systems. Springer, Berlin
26. Bulyanitsa DS, Svetlov YE (1962) Soviet Phys-Solid State 4:981 (the proof given is for a one-electron operator with local potential)
27. Monkhorst HJ (1979) Phys Rev B20:1504
28. Bary NK (1964) Treatise on trigonometric series. Pergamon, Oxford
29. Pitzer RM (1973) J Chem Phys 59:3308;
    Davidson ER (1975) J Chem Phys 62:400
30. Heinzmann R, Ahlrichs R (1976) Theoret Chim Acta 42:33;
    Ehrhardt C, Ahlrichs R (1985) Theoret Chim Acta 68:231
31. Raffenetti RC (1973) J Chem Phys 58:4452

32. Almlöf J, Taylor PR (1987) J Chem Phys 86:4070
33. Van Alsenoy C (1988) J Comput Chem 9:620
34. see, for example, Ringnalda MN, Belhadj M, Friesner RA (1990) J Chem Phys 93:3397
35. Pople JA, Krishnan R, Schlegel HB, Binkley JS (1979) Int J Quant Chem Quant Chem Symp 13:225
36. Pulay P (1987) Adv Chem Phys 69:241, and references therein
37. Handy NC, Schaefer HF III (1984) J Chem Phys 81:5031
38. Frisch MJ, Head-Gordon M, Pople JA (1990) Chem Phys Lett 166:275
39. Head-Gordon M, Trucks GW, Frisch MJ (1992) Chem Phys Lett 196:624 and references therein.