

# MP2 perturbation theory for large molecules

## Haser

## 1 Introduction

MP theory most economical way of including dynamical electronic correlation in ab initio electronic structure calculations of molecules. Standard MP2 applied to large molecular systems more demanding than SCF techniques. Current limit is about 100 valence electrons. For large molecules the computationally demanding step is the transformation of ERI's from given basis set to molecular orbitals. The computational effort can be reduced if near zero integrals are neglected in the course of the first three quarter transformation steps. The asymptotic operation count for traditional MP2 algorithm scales with the fifth power of the size of the system. Screening fails when applied to last quarter transformation step since scf mo's have non-zero amplitudes everywhere in a molecule.

In the SCF loop no integral transformation reduces operations for construction of fock from N4 to N2. The comp disparity b/n mp2 and scf is thus bound to grow.

To reduce comp labour for correlated wfn many have suggested localized occupied MOs with local correlation spaces. Latter represented by a truncated expansion of PNO. This approach has been helpful in variational and inf order calc of correlation energy but is too costly for MP2.

Unoptimized local correlation spaces from subsets of gaussians are more economical and successfully applied to MP2 theory. Problem is deliberate truncation of the local correlation spaces on the grounds of empirical rules. Need to be a systematic and efficient approach devised.

Much the same for local ansatz where similar or larger error bars apply. Not using localized MO but rely on choice of 2 electron excitation operators which define terms of empirically constructed local regions. This paper MP2 theory in AO basis, no restriction to particular choice of basis set.

## 2 Theory

Starting point is the expression for the MP2 energy of closed shell systems in terms of canonical SCF orbitals.

Typical closed shell formalism.

### 2.1 The laplace transform MP2 ansatz

The most expensive step in conventional algorithms is the transformation. It is the orbital energy denominator which necessitates ERI formed over canonical SCF MOs. Way to remove this denominator is Laplace transform ansatz using scaled canonical orbitals. This ansatz depends on the HOMO LUMO gap. The ansatz holds if the t-dependent orbitals are localized by some orthogonal transformation but approach not in this work.

First obstacle is the t-integration. This problem has been solved using quadrature.

$$1/x = \int_0^\infty dt e^{-xt} \approx \sum_{\alpha=1}^{\tau} w_{\alpha} e^{-xt_{\alpha}}$$

Useful approx if orbital energy is greater than 0 and confined to some finite interval and if parameters w and t are determined by a least squares condition.

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

where f(x) is the distribution density of  $x = \epsilon_{pa} + \epsilon_{pb} - \epsilon_{pi} - \epsilon_{pj}$

The number of exponentials  $\tau$  determined this way depends on the width of the interval but does not depend of the size of the molecular system.

This sums up some of the Laplace MP2 theory provided by Almlöf.

## 2.2 The AO-MP2 ansatz

Relate conventional MO basis MP2 to new approach need the expansion of the MOs in terms of AO (gaussians) Lets define the 2 symmetric matrices  $D^{(\alpha)}$  and  $D_{(\alpha)}$  as defined in the paper Each D relates to occupied occupied and virtual virtual blocks of the exponential of the fock matrix, respectively. Some MOs are deliberately be excluded from the summation and thus correlation treatment (frozen orbitals) as is often done for core orbitals (frozen core) in conventional MP2

Substituting these in forms a simple expression for MP2 energy with matrices and ERIs in gaussian basis. There is an eightfold summation that can be removed with special definition.

For each alpha four N5 transformation steps are necessary to establish the orbital step according to the new definition. Computationally this is worse than the 4 steps of conventional integral transformation for MP2 energies.

In typical application conventional mp2 scales 100 times more efficient than ao-mp2 Later will discuss how operation count will be sig reduced in applications to large molecules.

## 2.3 Integral screening in the AO-MP2 ansatz

ao-mp2 can be competitive if an efficient strategy used to reduce comp effort in evaluation of  $e_\alpha$  by a factor of 100 or more. Large molecules can be screened for integrals with near zero contributions to the MP2 energy in each quarter transformation step and in the final assembly of e. Outlined how screening procedure defining large and near 0.

Will use Schwarz inequality for interacting charge distributions. This innequality holds for any type of orbital provided the right hand side exists.

For the screenings to be useful the matrices have to be calculated and stored beforehand. Some of the matrices require N3 storage but only take N4 operations with direct evaluation.

To sum of this section: For each individual partial transformation the author has devised a bound to its contribution to the MP2 energy. The bound rely on 4 N2 matrices. The 4 matrices can be obtained by at most N3 integral evaluations and a direct N4 integral transformation step. The bounds are derived from Schwarz inequality and can thus be expected to be very efficient.

# 3 Algorithms

two algorithms presented: first a 2 step out of core transformation, later discovered a superior algorithm with solves the memory bottleneck problem in MP2 calculations

## 3.1 two-step out-of-core algorithm

AO-MP2 uses semi-direct two step out of core transformation

1. get basic data
2. determine the dist function  $f(x)$
3. calc from  $f(x)$  tau, w, talpha
4. Choose threshold parameter for screening partial transformed integrals
5. Calc the integral bounds Q
6. for each alpha = 1, ..., tau evaluate a contribution to the MP2 energy
7. output the MP2 energy E.

In their imp exploited symmetry. integral neglect in one quarter transform not independent from neglect at other stages of AO-MP2 calc. effects contribute in second order to the total error, insignificant as screening threshold is small.

If the disc reqs are larger than available switch to a mutiple pass algorithm.

### 3.2 note on multiple pass algorithms

Devised in conventional MP2 schemes with (semi-)direct ERI transformations in order to cut disc space demands for (partially) transformed integrals. Accomplished by partitioning the occupied space into  $m \leq n$  subspaces  $M$  and performing a partial MP2 calculation  $m$  times, each time evaluating quarter transformed integrals. During the complete MP2 calc the orbital integrals need to be recalculated from each occupied sub-space. Integral screening reduces the scaling factor.

Using the screening developed in this paper expect to reach scaling limit earlier than conventional direct multiple pass MP2.

$m$  passes through the AO evaluation does not imply an increase of integral evaluation cost by a factor of  $m$ .

### 3.3 in core multiple pass algorithm

Another algorithm is obtained if use multiple pass type integral re-evaluation at beginning and end of the AO-MP2 transformation. without integral screening this calculates by  $N^5/\text{dim}(M) = mN^4$  plus an  $N^5$  operation count for the transformation. The largest matrices to be kept in memory have small dimensions and each contribution can be evaluated independently one different nodes of a parallel computer.

## 4 applications: efficiency and over-all accuracy of AO-MP2 ansatz

2 parameters in AO-MP2 determine the accuracy and the computational cost of a calculation the number of exponentials and the screening threshold.

Any pre-determined accuracy can be accomplished by proper choice of  $t$  and threshold.

in p-chloro-phospha-benzene performance due to integral screening increased by 5 and  $N^5$  dependency moved to  $N^4$  though it will still be slower than conventional MP2 algorithm for half transformed integrals. To conclude MP2 correlation energies have been evaluated by AO-MP2 ansatz and an operation count of  $N^4$  larger system (430 basis functions). However AO-MP2 is a factor of 100 more costly in small molecules than conventional MP2. Current implementations of AO-MP2 ansatz not competitive if micro hartree accuracy required. Expect the break point to be at about 200 atoms..

## 5 further improvements

AO-MP2 could be more efficient if  $D$  matrices were sparse even in medium sized molecules. This depends on the basis set used. In canonical SCF diagonalization of the  $D$  matrices is back to original laplace transform MP2 ansatz.

There is a balance between AO-MP2 and laplace transform. Symmetry adapted basis functions are one example.

Choose to partition the basis functions by ones centered at the same atom, or small groups of atoms. New basis in the atom centered subspace  $M$  which is orthonormal in  $M$ . Basis functions from different subspaces would not usually be orthogonal. The AO-MP2 would now be formulated in terms of new molecular basis set  $B$  which thus consists of modified atomic orbital or group orbitals.

The modified AO approach was not tested in this paper.

referred to a similar consideration in a different context is modified atomic orbitals (MAO) previously used in population analyses where carefully optimized small number of MAOs at each atom can be a good approx serve as a basis of a scf one-electron density.

Calculation of Mixed gaussian/MAO ERI not a problem similar to ERI have efficiently been implemented in context with atomic natural orbital generalized contraction basis sets.

## 6 conclusions

With AO-MP@ ansatz presented a novel formulation of MP2 theory. Starting with laplace transform MP2 moved into atomic orbital basis representation and eliminated the need to calculate electronic repulsion

integrals for MOs. Aim has been to remove all bottlenecks which hinder an application of MP2 to larger molecules. Show that AO-Mp2 grows at most with fourth power of size of the system. AO-MP2 does not outperform a conventional semi-direct MP2 program in molecular systems with 100 correlated electron pairs or less. AO-MP2 for molecular systems of 200 atoms or more.