Theor Chim Acta (1993) 86: 279–284

Theoretica Chimica Acta

The use of symmetry in direct Møller–Plesset second-order calculations

Christopher W. Murray, Jamie S. Andrews, and Roger D. Amos
University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, UK

Received October 26, 1992/Accepted December 14, 1992

Summary. An algorithm for utilising abelian point group symmetry in direct MP2 energy calculations is presented. This is based upon the direct MP2 method of Head-Gordon, Pople and Frisch. The method uses the petite atomic orbital integral list as in conventional transformations coupled with a symmetry adaption of the three quarter transformed integrals. Representative calculations for ethylene and benzene are presented which demonstrate the potential of the method.

Key words: Second-order calculations – MP2 – Abelian point group symmetry – Ethylene – Benzene

1 Introduction

Large-scale Self-Consistent Field (SCF) calculations can today be carried out routinely and cheaply thanks to the development of 'Direct methods' as pioneered by Almlöf and co-workers [1, 2]. As a result SCF calculations with up to 1000 basis functions have now appeared in the literature [3]. In essence a 'Direct' code simply recomputes the two-electron repulsion integrals (ERIs) when needed, thereby avoiding quartic disk requirements. Despite the greater computational cost associated with these algorithms, Direct methods have become highly efficient and competitive due to improvements in integral evaluation [4, 5] and the rapid developments in CPU technology.

As is well understood the SCF method does not describe electron correlation and post Hartree–Fock treatments are generally required for the accurate description of many chemical problems. For those systems where the SCF method affords a reasonable description of the ground electronic state, the MP2 method (Møller–Plesset Theory [6] to Second-Order) has been found to be extremely reliable in recovering the bulk of the correlation. Because MP2 calculations are size consistent and scale as N^5, where N is the number of basis functions, this method has historically been the most attractive approach for a correlated description of large chemical systems. Consequently there has been much work in extending the Direct approach to the evaluation of MP2 energies [7, 8] and gradients [9, 10]. Efficient evaluation of the MP2 energy needs a restricted four index transformation to give the Molecular Orbital (MO) integrals $(ia \mid jb)$, where i and j are occupied, and a and b virtual orbitals. The expression for the MO integrals is:

$$(ia \mid jb) = \sum_s C_{sb} \sum_r C_{rj} \sum_q C_{qa} \sum_p C_{pi}(pq \mid rs) \tag{1}$$

where $(pq \mid rs)$ are the Atomic Orbital (AO) ERIs and C represent the appropriate molecular orbitals. The cost of each summation in Eq. (1) working from the inner sum outwards is ON^4, OVN^3, O^2VN^2 and O^2V^2N [11], where N denotes the number of basis functions, V the number of virtual orbitals and O the number of occupied orbitals. Unlike conventional disk-based codes, a Direct MP2 algorithm establishes the MP2 energy without storing the AO ERIs and without using external storage for the partly transformed integrals. Another approach has been termed Semi-Direct, where external storage is used to store partly transformed integrals. The choice between these two approaches is dictated by the balance between core and disk space. In this paper we shall be concentrating on extending the Direct MP2 method to larger systems.

The direct algorithm suggested by Head-Gordon, Pople and Frisch [7] is depicted in Fig. 1 and will be referred to as Algorithm A. In this method the four

```
   Loop over $i$ molecular orbitals
       Loop over $R$, $S$ shells
       Loop over $P$, $Q$ sections
       Loop over $pqrs \in PQRS$
           Form integrals $(pq \mid rs)$ for $p \geqslant q, r \geqslant s$
       End $pqrs$
       Loop over $rs$
       Loop over $pq$
1          $(iq \mid rs) = (iq \mid rs) + C_{pi}(pq \mid rs)$
           $(ip \mid rs) = (ip \mid rs) + C_{qi}(pq \mid rs)$
       End $pqrs$
       End $PQ$
       Loop over $rs$
       Loop over $a, q$
2          $(ia \mid rs) = (ia \mid rs) + C_{qa}(iq \mid rs)$
       End $a, q$
       Loop over $a, j$ $(j \leqslant i)$
3          $(ia \mid js) = (ia \mid js) + C_{rj}(ia \mid rs)$
           $(ia \mid jr) = (ia \mid jr) + C_{sj}(ia \mid rs)$
       End $a, j$
       End $rs$
       End $RS$
       Loop over $j$
       Loop over $a, b, s$
4      $(ia \mid jb) = (ia \mid jb) + C_{sb}(ia \mid js)$
       End $a, b, s$
       Loop over $a, b$
       Find contribution from $i, j$ to MP2 energy
       End $a, b$
       End $j$
   End $i$
```

Fig. 1. Structure of algorithm A

index transformation is related to the one used in the GAUSSIAN series of programs as described by Hehre et al. [12], but differs in that the loop over the RS shells is not closed after the first quarter transformation (step 1), thereby avoiding $\mathcal{O}(N^3)$ storage for the $(iq \mid rs)$ integrals. Instead the RS loop is closed only after the transformation of the three-quarter transformed integrals, $(ia \mid js)$, at step 3. As a consequence the $(pq \mid rs) = (rs \mid pq)$ permutational symmetry cannot be exploited and the AO integrals are effectively done twice. However the reduction in memory makes this worthwhile especially in the context of the Direct SCF, where the integrals are recalculated anyway. The main storage requirement of the scheme is the array that holds the three quarter transformed integrals at step 3 which requires a maximum of OVN storage. In fact, the outer loop is done over batches of 'i' molecular orbitals, the size of each batch determined by how much memory is available. This feature is not shown in the figure so as to give greater clarity. In effect then, the method involves recalculating the integrals $2I$ times where I is the number of batches of i molecular orbitals. The method is therefore still only N^5 in cost although in the limit as $I \to O$ the integral re-evaluation will dominate the time taken, since many more operations are involved in this step.

We have used Direct MP2 algorithms of this type to perform large basis set MP2 calculations on acetylene trimer [13] and benzene [14]. With the computing resources available to us these calculations would not have been possible using conventional methods. As the size of the system and number of basis functions increase there will come a point where even this algorithm is no longer applicable and the OVN storage is prohibitive. In this work we present a Direct MP2 energy algorithm based on Algorithm A which significantly reduces the memory requirement for molecules with point group symmetry, thereby extending the Direct MP2 method to the study of larger systems. Within a fixed core allocation this method will reduce the number of batches (and therefore integral re-evaluations) required, saving on CPU time.

2 Use of symmetry

The number of batches is primarily controlled by the amount of memory taken up by the array holding the $(ia \mid js)$ integrals and the storage used is OVN times the number of batches. Symmetry can be used to reduce the storage of this array as well as to decrease the amount of time spent in every stage of the calculation.

In many *ab initio* codes, point group symmetry is used to reduce the computational task by the method of Dupuis and King [15]. In this method only the 'petite' set of symmetry unique integrals are formed and these are multiplied by the number of integrals they are related to by symmetry. The four index transformation can then be performed in the usual way, Pitzer's theorem [16] ensuring that the symmetric integrals have the correct value. This method ensures that less work is done for those molecules where the atoms are transformed into other atoms by the group operations. In CADPAC [18] the above procedure is standard for symmetries of D_{2h} and its subgroups. The analogous method for non-abelian point groups is more involved but has now been successfully implemented by Häser and co-workers [19].

In a Direct approach to the evaluation of the MP2 energy, such as algorithm A, the above scheme ensures that less work is required in the evaluation of the AO ERIs. However symmetry can also be used to reduce the storage problems associated with the $(ia\,|\,js)$ integrals formed in step 3. One such approach would be to perform a transformation of the AO ERIs to symmetry orbital integrals [20, 21] prior to the four index transformations. The transformed and partly transformed integral lists would then only include integrals non-zero by symmetry, thereby reducing the memory requirement at every stage. However this approach would be expensive (note that a symmetry transformation would also inhibit the sparsity of the integral list and the effectiveness of integral prescreening) and would require substantial re-coding. A more practical alternative is a mixture of the two approaches that keeps the desirable properties of the petite integral list but saves on store at the crucial stage of the transformation [22]. Simply put, the first two stages of the transformation are obtained from the petite set[a] of AO ERIs as in Algorithm A, but at the critical third stage of the transformation the final basis function is simultaneously symmetry adapted to give $(ia\,|\,js')$. Of course, the use of only the petite set of integrals is still valid for this mixed integral since the symmetry orbitals and the molecular orbitals share the same symmetry properties. In the next section we present some calculations which demonstrate the significant savings in memory and CPU time that are achieved with this new method.

3 Demonstrative calculations and discussion

As an illustration of our method a series of calculations was performed on ethylene and benzene utilising specific point groups. Table 1 presents the timings and relative timings for the six ethylene calculations performed. A $(8s6p3d/6s3p)$ basis set was used (giving a total of 148 basis functions) and all MOs were active. Similar calculations for benzene were performed and the results are given in Table 2. In this set of calculations a $(5s4p2d/3s2p)$ basis was used (giving a total of 228 basis functions) and the six carbon $1s$ MOs were frozen. For the calculations on both molecules the memory was restricted so that the D_{2h}

Table 1. Performance of MP2 symmetry energy algorithm for C_2H_4 calculated at different point group symmetries

Point group	Order of group	Number of unique atoms	Number of AO integral evaluations	Time[a]/s	Relative timings
D_{2h}	8	2	1	520	1
C_{2v}	4	2	2	913	1.8
C_{2v}	4	4	2	1661	3.2
C_s	2	3	3	1970	3.8
C_2	2	4	3	2385	4.6
C_1	1	6	6	6447	12.4

[a] Calculations performed on a Convex C220 machine

---

[a] In fact for our purposes, since the permutational symmetry $(pq\,|\,rs) = (rs\,|\,pq)$ is not being taken into account, the normal implementation must be amended slightly but this does not represent a problem.

Table 2. Performance of MP2 symmetry energy algorithm for benzene calculated at different point group symmetries

Point group	Order of group	Number of unique atoms	Number of AO integral evaluations	Time[a]/s	Relative timings
D_{2h}	8	4	1	1043	1
C_{2v}	4	6	2	2697	2.6
C_{2v}	4	8	2	2862	2.7
C_s	2	6	3	3689	3.5
C_s	2	8	3	3977	3.8
C_s	2	12	3	7449	7.1
C_1	1	12	6	12945	12.4

[a] Calculations performed on a Convex C340 machine

calculation could be performed in one batch and the lower symmetry calculations required more batches. For benzene this meant that 2.5 MW of memory were allocated to the program. This calculation would have been difficult with usual computing resources and conventional MP2 codes, and is a better indication of how our algorithm extends the applicability of the Direct MP2 method to larger systems.

The first point to note from the tables is that using symmetry as described above, we can obtain an order of magnitude saving in the MP2 calculation. The savings achieved, as indicated by the relative timings, are significant. We can come to some understanding of these savings by considering how symmetry reduces the cost of the calculations. An obvious reduction occurs from only calculating the petite list of AO integrals during the transformation. This gives a saving dictated by the order of the group (g) and the number of symmetry unique atoms since, in the Dupuis and King formulation, the grande list of integrals is at most g times larger than the petite list. In practice the actual saving is less because the method only takes advantage of group operations that move atoms onto other atoms, and therefore offers no saving for molecules where the number of symmetry unique atoms equals the total number of atoms. Furthermore the routines that calculate symmetry unique integrals are slower than those that evaluate all integrals since the symmetry tests are difficult to vectorise. As an example, the SCF part of the D_{2h} calculation on benzene was only 3.3 times faster than the C_1 calculation.

Another saving is achieved during the transformation part of the algorithm. At the first step of the transformation (step 1) there is a reduction in the cost of up to the order of the group because only the petite list of atomic integrals are processed. In fact, since most of the saving in the petite list occurs in the loops over the *RS* shells, the saving in the 2nd and 3rd is of about the same order. The fourth step of the transformation is only performed for totally symmetric integrals and therefore there is a possible saving in this part of about g^2. Consequently the overall saving in the transformation is at most the order of the group.

Finally there is a further saving associated with the reduction in memory of the *OVN* array. At step 3 the symmetry transformation reduces the amount of memory required for the $(ia\,|\,js)$ integrals by an approximate factor of g. As a result more occupied MOs (i) can be treated in any one batch leading to a possible reduction in the total number of batches (and AO integral evaluations) by up to this factor.

Summarising then, there is a possible saving of up to a factor of g^2 associated with the repeated AO integral evaluations and a saving of up to a factor of g associated with the transformation. Depending on the balance of time between the recalculation of AO integrals and the transformation of those integrals, any saving up to a factor of g^2 could be observed. This explains why we obtain relative speed ups that are larger than the order of the group in Tables 1 and 2. In fact for the D_{2h} calculation on benzene the MP2 part of the calculation runs over 12 times faster. The calculations on ethylene indicate that similar savings can also be expected for other molecules. The results also suggest that the cost of large MP2 calculations using this algorithm is dictated by the cost of the repeated calculation of the AO integrals. As a final point we note that for the benzene C_s calculation with 12 symmetry unique atoms, there is still a speed up of a factor of 1.7, even though the Dupuis and King symmetry formulation offers no advantage. In this case, the saving occurs only through the reduction in the number of batches.

In conclusion, the method outlined here offers significant saving over the Direct MP2 approach when applied to symmetrical molecules. By reducing the amount of memory used by the large OVN array formed at step 3 of the algorithm, we have extended the applicability of the method to larger systems. An important final point is that this method is simply implemented in existing Direct MP2 codes.

References

1. Almlöf J, Faegri K, Korsell K (1982) J Comput Chem 3:385
2. Häser M, Ahlrichs R (1989) J Comput Chem 10:104
3. Price SL, Harrison RJ, Guest MF (1989) J Comput Chem 10:552
4. Obara S, Saika A (1986) J Chem Phys 84:3963
5. Head-Gordon M, Pople JA (1988) J Chem Phys 89:5777
6. Møller C, Plesset MS (1934) Phys Rev 46:618
7. Head-Gordon M, Pople JA, Frisch MJ (1989) Chem Phys Lett 153:503
8. Sæbø S, Almlöf J (1989) Chem Phys Lett 154:83
9. Frisch MJ, Head-Gordon M, Pople JA (1990) Chem Phys Lett 166:275
10. Frisch MJ, Head-Gordon M, Pople JA (1990) Chem Phys Lett 166:281
11. In fact, if $V > O$ the summation can be written slightly more efficiently by reversing the order of the middle two summations. However the order in (1) is more pertinent to the succeeding discussion
12. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) *Ab initio* molecular orbital theory. Wiley-Interscience, NY
13. Bone RGA, Murray CW, Amos RD, Handy NC (1989) Chem Phys Lett 161:166
14. Handy NC, Maslen PE, Amos RD, Andrews JS, Murray CW, Laming GL (1992) Chem Phys Lett 197:506
15. Dupuis M, King HF (1977) Int J Quantum Chem 11:613
16. Pitzer RM (1973) J Chem Phys 59:3308
17. Amos RD, Rice JE (1989) Comp Phys Rep 10:147
18. Amos RD, Alberts IL, Andrews JS, Colwell SM, Handy NC, Jayatilaka D, Knowles PJ, Kobayashi R, Koga N, Laidig KE, Maslen PE, Murray CW, Rice JE, Sanz J, Simandiras ED, Stone AJ, Su MD (1992) CADPAC5: The Cambridge Analytic Derivatives Package Issue 5 (Cambridge)
19. Häser M, Almlöf J, Feyereisen W (1991) Theor Chim Acta 79:115
20. Almlöf J (1974) Int J Quantum Chem 8:915
21. Takada T, Sasaki F (1980) Int J Quantum Chem 18:1157
22. Murray CW (1989) PhD Thesis. Cambridge University