

# MP2[V] – A Simple Approximation to Second-Order Møller-Plesset Perturbation Theory

Jia Deng, Andrew T.B. Gilbert, and Peter M.W. Gill\*

*Research School of Chemistry, The Australian National University, ACT 0200, Australia*

(Dated: March 17, 2015)

We propose a simplified variant of the dual-basis MP2[K] scheme [Journal of Chemical Physics, **12**, 081103 (2011)] that bootstraps a small-basis MP2 result to a large-basis one. This simplified method, which we call MP2[V], assumes the occupied orbitals are adequately described by the smaller basis and, therefore, only the relaxation of the virtual orbitals is considered when shifting to the larger basis. Numerical tests on several organic reactions and non-covalent interactions show that MP2[V] yields absolute and relative energies that are in excellent agreement with the conventional large-basis MP2 calculations, but at a small fraction of the time.

## I. INTRODUCTION

In order to calculate most molecular properties accurately, it is vital to account for the effects of electron correlation, *i.e.*, to go beyond the Hartree-Fock approximation [1]. Second-order Møller-Plesset (MP2) perturbation theory is one of the least expensive wave function-based electronic structure methods that includes such effects [2]. Compared to the more economical Kohn-Sham density functional theory (DFT) methods, MP2 has the advantage of naturally and properly accounting for medium- and long-range correlation effects. Furthermore, the appearance of scaled MP2 [3, 4] and double-hybrid DFT [5] (which includes a MP2-like term) has helped to highlight the importance of MP2 theory in quantum chemistry. However, whether one uses conventional MP2, scaled MP2, or double-hybrid DFT, the steep computational cost associated with calculating the MP2 correction term, and the need for large basis sets for reliable results, pose significant obstacles to its application in large molecular systems.

The bottleneck in an MP2 calculation is the transformation of the two-electron integrals from the atomic basis to the molecular orbital basis, and this step scales as the fifth power of the number of basis functions. There have been numerous attempts to reduce the cost of this transformation including local MP2 (LMP2) [6, 7], cutoff-based Laplace-transformed MP2 [8–10], atomic-orbital-based LMP2 [11] and scaled-opposite-spin MP2 [4]. All these methods have costs that scale more slowly as the system size is increased. Other approaches, such as those based on density fitting [12–14], Cholesky decomposition [15], or the pseudospectral method [16], dramatically reduce the cost prefactor of the integral transformation, but still retain the fifth-order scaling. More recently, extraordinary speed-ups of MP2 calculations have been achieved by exploiting graphics processing units (GPUs) [17–19].

Despite the impressive improvements offered by these

new methods, they are not without their limitations. Methods that rely on spatial cutoffs only exhibit reduced scaling when applied to relatively large structures with modest basis sets. When applied to more compact structures, they exhibit the same high-order scaling, particularly as the basis set size is increased. Furthermore, cutoff-based methods neglect contributions from distant electron pairs, leading to the underestimation of dispersion interactions, obviating one of the key advantages of MP2 over the much cheaper DFT methods. Because of these limitations, there remains a need for developing MP2 alternatives that are cheaper, yet still maintain the accuracy of conventional MP2 calculations.

The use of dual basis sets has provided useful efficiency gains in both Hartree-Fock [20, 21] and DFT [22, 23] calculations. More recently, the dual basis strategy has been extended to the calculation of MP2 energies [24, 25] and we have proposed a hierarchy of such dual basis MP2 schemes (denoted MP2[ $x$ ], where  $x = 1, 2, J, K, 3$ ) [25]. These schemes all improve the energy of a small primary basis MP2 calculation by including various subsets of the orbital corrections obtained from a larger secondary basis HF calculation. Preliminary results showed that these schemes yield energies that are in excellent agreement with the target secondary basis and, in principle, promise significant computational savings. The MP2[K] scheme, which neglects all the three- and four-orbital corrections and includes only some of the two-orbital corrections, was found to offer a particularly attractive trade-off between cost and accuracy.

In the course of our investigations we became interested in a simplified version of the MP2[K] scheme which takes advantage of the fact that, for correlated calculations, the basis set demands of the occupied and virtual orbitals are very different. Because the relaxation of occupied orbitals due to basis set extension is small, it might be possible to neglect this relaxation altogether without sacrificing much accuracy. In this Paper, we present our simplified scheme, which we call MP2[V], and provide accuracy and timing results that demonstrate its efficacy.

---

\*E-mail: peter.gill@anu.edu.au

## II. THEORY

We begin by briefly outlining the MP2[ $x$ ] family of approximations of which MP2[V] is member. For more details the reader is referred to the original paper [25].

A self-consistent field calculation using a target (secondary) basis consisting of  $N$  functions yields  $O$  occupied and  $V$  virtual molecular orbitals (MOs). These orbitals, and their associated orbital energies,  $\epsilon_i$ , can be used to calculate the MP2 correlation energy

$$E^{(2)} = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{virt}} \frac{\langle ij || ab \rangle^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (1)$$

$$\langle ij || ab \rangle = (ia|jb) - (ib|ja)$$

$(ia|jb)$  denotes an electron repulsion integral (ERI) in the MO basis, which is obtained by contracting the atomic orbital (AO) integrals with the MO coefficients

$$(ia|jb) = \sum_{\mu\nu\lambda\sigma}^N C_\mu^i C_\nu^a C_\lambda^j C_\sigma^b (\mu\nu|\lambda\sigma) \quad (2)$$

This transformation step is the most expensive part of an MP2 calculation and, for maximum efficiency, is carried out in four quarter-transformations. The first of these quarter-transformations, which is normally the most expensive, has a cost of  $ON^4$ .

The MP2[ $x$ ] methods tackle this computational bottleneck by avoiding the construction of the exact secondary ERIs in (2). Rather, an additional HF calculation is performed using a much smaller primary basis of  $n$  functions, to obtain  $O$  occupied and  $V_p$  virtual orbitals. Each secondary MO-ERI can then be written as the sum of a primary MO-ERI ( $i_p a_p | j_p b_p$ ), and correction terms that account for one-, two-, three-, and four-orbital relaxation effects. **By including different subsets of the orbital corrections and retaining secondary orbital energies in the denominator**, a hierarchy of approximations can be established that bridge between the primary and secondary MP2 energies (and costs). For example the MP2[K] approximation can be written

$$(ia|jb)_{[K]} = (i_\delta a_p | j_p b) + (i_p a | j_\delta b_p) + (i_p a | j_p b) \quad (3)$$

where the orbital corrections,  $i_\delta$ , are given by

$$i_\delta = i - i_p \quad (4)$$

In most cases the occupied orbitals are well-described by the primary basis, from which it follows that their orbital corrections,  $i_\delta$  and  $j_\delta$ , are small. Neglecting these corrections leads to a simplification of the MP2[K] expression, which we define as the MP2[V] approximation for the secondary MO-ERI

$$(ia|jb)_{[V]} = (i_p a | j_p b) \quad (5)$$

The V subscript indicates that only relaxation of the virtual orbitals is included. The MP2[V] correlation energy

is given by

$$E_{\text{MP2[V]}}^{(2)} = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{virt}} \frac{\langle ia || jb \rangle_{[V]}^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (6)$$

and the total MP2[V] energy, which we denote by

$$\text{MP2[V]}/\text{primary basis}/\text{secondary basis}$$

is the sum of this and the HF energy in the secondary basis.

Although mathematically simple, the implementation of efficient integral transformations requires careful consideration of the available memory and disk space. Two main types of serial algorithm have surfaced in the literature, the Saebø and Almlöf algorithm [26], and the Head-Gordon and Pople [27] algorithm, both of which can be easily applied to the transformation required to produce the MP2[V] integrals. Table I summarizes the costs and relative savings of both of these algorithms when used to calculate the MP2[V] integrals compared to the cost of calculating the regular MO-ERIs.

In the Saebø and Almlöf algorithm, the integral transformation is accomplished via two half-transformations involving first the two occupied orbitals and then the two virtual orbitals. As a consequence, the computational cost for the four transformation steps scale as  $On^2N^2$ ,  $O^2nN^2$ ,  $O^2VN^2$ , and  $O^2V^2N$ . When used to calculate the MP2[V] integrals, the first two transformations are cheaper than regular MP2 by factors of  $(N/n)^2$ , and  $(N/n)$ , with the overall savings being dominated by the  $(N/n)^2$  factor. However, this algorithm requires all AO-ERI of the type  $(R\mu|S\nu)$ , for fixed secondary basis functions  $R$  and  $S$  and for all primary basis functions,  $\mu$  and  $\nu$ , prior to the transformation. These must be generated in the order  $RS\mu\nu$  (slow to fast) and this sequence is not compatible with the AO-ERI package within Q-CHEM.

The alternative is the cubic-memory algorithm of Head-Gordon and Pople and this is the one we have adopted. It performs the transformation in the order of occupied - virtual - occupied - virtual, and accumulates the results of the half-transformed integrals,  $(i_p a | S\nu)$ , on disk. It uses only cubic memory and disk storage and, where there is insufficient memory, the algorithm performs the transformation in multiple batches, each requiring the recalculation of the entire set of AO integrals. The MP2[V] modification of this algorithm reduces the first three quarter-transformations by factors of  $(N/n)^2$ ,  $(N/n)$ , and  $(N/n)$  when compared to the calculation of the secondary MO-ERIs. We note that the second quarter transformation step dominates the cost and therefore expect the cost reduction to be around  $(N/n)$ . However, because of the reduced memory requirements, calculating the  $(i_p a | j_p b)$  integrals can require fewer batches which introduces additional savings as the AO-ERI do not have to be recomputed as often. Pseudocode showing the modified algorithm for calculating the  $(i_p a | j_p b)$  integrals is shown in Algorithm 1.

TABLE I: Costs (multiplies and adds) of the quarter integral transformations in the MP2[V] approximations.  $n$  = size of primary basis set;  $N$  = size of secondary basis set;  $O$  = number of occupied orbitals; orbital indices with a p subscript refer to MOs in the primary basis; unsubscripted orbital indices represent secondary MOs;  $\mu$  and  $\nu$  represent primary AO basis functions;  $R, S$  represent secondary AO basis functions.

	Head-Gordon and Pople			Saebø and Almlöf		
	Transformed integral	Cost of quarter transformation	Cost reduction	Transformed integral	Cost of quarter transformation	Cost reduction
1 <sup>st</sup> quarter	$(i_p R   \nu S)$	$O n^2 N^2$	$(N/n)^2$	$(i_p R   \nu S)$	$O n^2 N^2$	$(N/n)^2$
2 <sup>nd</sup> quarter	$(i_p a   \nu S)$	$O n V N^2$	$(N/n)$	$(i_p R   j_p S)$	$O^2 n N^2$	$(N/n)$
3 <sup>rd</sup> quarter	$(i_p a   j_p S)$	$O^2 n V N$	$(N/n)$	$(i_p a   j_p S)$	$O^2 V N^2$	1
4 <sup>th</sup> quarter	$(i_p a   j_p b)$	$O^2 V^2 N$	1	$(i_p a   j_p b)$	$O^2 V^2 N$	1

---

**Algorithm 1** Pseudocode of the MP2V integral transformation program.

---

```

loop { over  $K$  batches (batch size  $k$ ) }
  loop { over  $\nu S$  batches,  $\nu$  primary,  $S$  secondary AO functions }
    loop { over  $\mu R$  batches,  $\mu$  primary,  $R$  secondary AO functions }
      loop {  $\mu R \nu S$  }
        calculate  $(\mu R | \nu S)$  AO ERIs
      end loop
      loop { over  $\mu R, i_p \in k$  }
         $(i_p R | \nu S) += (\mu R | \nu S) C_{\mu i}^p$ 
      end loop
    end loop
    loop { over  $R, a, \nu S, i_p \in k$  }
       $(i_p a | \nu S) += (i_p R | \nu S) C_{Ra}$ 
    end loop
    write  $(i_p a | \nu S)$  to disk
  end loop
  loop { over  $i_p \in k$  }
    read  $(i_p a | \nu S)$  for all  $a, R$ , and  $\nu$ 
    loop { over  $a$  }
      transform  $(i_p a | \nu S)$  to  $(i_p a | j_p b)$ 
    end loop
    loop { over  $j_p$  }
      evaluate  $i_p j_p$  contributions to obtain  $E_{\text{MP2[V]}}^{(2)}$ 
    end loop
  end loop
end loop

```

---

### III. RESULTS

#### A. Computational details

To demonstrate the accuracy and efficiency of the MP2[V] method, we have calculated all-electron MP2[V] energies and examined the errors in both absolute and reaction energies with respect to MP2 benchmark calculations using the full secondary basis. Specifically we considered the set of 34 isomerization reactions featured in our previous studies and also the S22 set of non-covalent interaction systems. For all these systems, the energy differences are sensitive to basis set quality and therefore

provide a stringent test of our method. The geometries for the structures were taken from the literature [28, 29] and all self-consistent field (SCF) calculations, either in the primary or secondary basis set, were converged to a DIIS (direct inversion in the iterative subspace) error of  $10^{-7}$ . All calculations were performed within a development version of the Q-CHEM program [30].

#### B. Performance Analysis

In our previous work, we found that cc-pVDZ is a cost-effective primary basis when the target secondary basis is cc-pVQZ. We therefore computed MP2[V]/cc-pVDZ/cc-pVQZ energies for the set of isomerization reactions shown in Figure 1. Table II summarizes the MP2[V] deviations from the benchmark MP2/cc-pVQZ reaction energies, along with the mean absolute error (MAE), maximum error (MAX) and the percentage of the benchmark reaction energy captured by the MP2[V] approximation.

While the secondary orbitals are usually obtained by conventional SCF calculations, it is beneficial to invoke economical SCF approximations for systems where the underlying large-basis HF calculation represents a significant overhead. One such method is the Hartree-Fock Perturbative Correction (HFPC)[21] which obtains approximate secondary orbitals from a single diagonalization of a small basis Fock operator in the larger secondary basis set.

The results in columns 2–5 are for MP2[V] coupled with conventional HF/cc-pVQZ SCF calculations whilst the results in columns 6–9 show the performance of MP2[V] combined with the unconverged orbitals obtained using HFPC. In both cases, the summary statistics for the reactants and products have been combined.

When coupled with conventional HF calculations, the average error in the MP2[V] total energies is only 2.0 kJ/mol and only 5 out of the 68 errors exceed 4 kJ/mol. Unlike the HFPC method, none of the MP2[ $x$ ] variants, including MP2[V], is variational with respect to the true MP2 energy and therefore the errors in the total energies of the reactants and products can be of either sign. Inter-

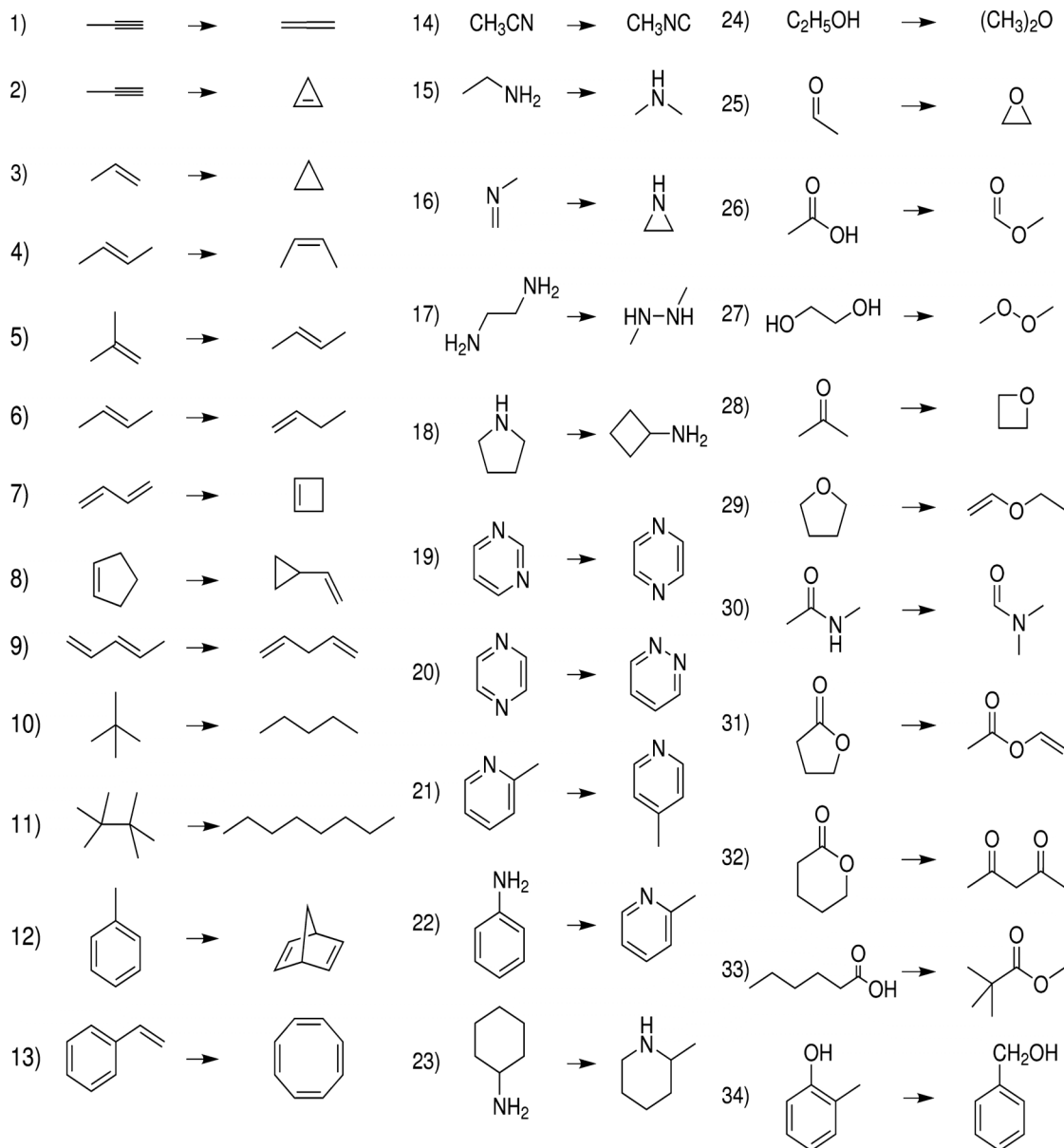


FIG. 1: Isomerization reactions included in the test set

estingly, however, when MP2[V] is coupled with HFPC the energies of individual reactants and products are consistently over-estimated leading to errors in the total energies that are all positive.

The errors for the reaction energies are, on average, significantly lower than those of the total energies. This reflects the fact that a large portion of the error in the total energies arises from core electrons, and this error cancels when considering energy differences. In particular, it is worth noting that the reactants and products with the largest errors (reactions 10, 11 and 23) yield some of the smallest reaction energy errors. Also, despite

the use of orbitals from HFPC calculations giving much higher errors for the reactants and products, they give very similar errors for the reaction energies. This suggests additional savings may be achieved by combining the two perturbative approaches when calculating reaction energies, although such benefits become less efficacious as the system size increases.

Since MP2 is frequently used to describe non-covalent interactions, it is desirable for any approximate MP2 method to maintain its accuracy for such systems. To test if this is the case for the MP2[V] method, we applied it to the S22 set of weakly bound systems and have

TABLE II: Mean absolute errors MAE, kJ/mol), maximum absolute errors (MAX, kJ/mol) and the percentage recovery of the MP2/cc-pVQZ reaction energy for MP2[V] coupled with HF/cc-pVQZ (cols. 2-5) and HFPC/cc-pVDZ/cc-pVQZ (cols. 6-9).

Reaction	MP2[V]/cc-pVDZ/cc-pVQZ				HFPC+MP2[V]/cc-pVDZ/cc-pVQZ			
	Reactant	Product	Reaction	%	Reactant	Product	Reaction	%
1	-0.3	0.1	0.5	97.7	3.2	2.5	-0.8	103.8
2	-0.3	-0.7	-0.3	100.3	3.2	3.1	-0.2	100.2
3	-1.4	-3.5	-2.1	110.3	3.7	5.9	2.2	89.3
4	-3.1	-2.7	0.4	90.7	5.8	5.3	-0.5	110.0
5	-2.3	-3.1	-0.8	112.7	5.3	5.8	0.5	92.4
6	-3.1	-2.2	0.9	92.3	5.8	5.2	-0.6	105.4
7	0.1	-1.0	-1.1	102.8	2.7	3.4	0.6	98.4
8	-2.2	-2.7	-0.6	100.6	5.2	6.2	1.1	98.9
9	-1.7	-0.6	1.2	96.2	4.9	4.1	-0.8	102.7
10	-6.2	-6.1	0.1	99.4	10.1	9.6	-0.5	102.5
11	-9.8	-9.4	0.4	97.8	15.5	14.4	-1.0	105.7
12	-2.2	-0.5	1.7	99.2	5.7	4.3	-1.4	100.7
13	-1.3	0.3	1.6	99.1	5.4	3.8	-1.6	100.9
14	0.4	1.5	1.0	99.1	2.7	2.9	0.2	99.8
15	-1.9	-2.1	-0.2	100.6	6.4	6.1	-0.3	100.8
16	0.4	-1.2	-1.5	104.5	3.0	4.7	1.7	95.0
17	-1.2	-2.0	-0.8	100.6	8.3	7.7	-0.6	100.5
18	-2.4	-2.1	0.3	99.3	7.5	7.4	-0.1	100.2
19	2.8	1.6	-1.2	107.6	1.6	3.8	2.2	85.3
20	1.6	2.2	0.6	99.3	3.8	2.8	-1.0	101.3
21	-0.4	-0.3	0.1	98.0	5.1	4.9	-0.2	104.7
22	-0.3	-0.4	-0.1	100.6	5.6	5.1	-0.5	102.8
23	-3.9	-4.3	-0.4	101.5	10.3	10.2	-0.1	100.4
24	-0.6	-0.6	-0.0	100.0	4.7	4.2	-0.5	100.9
25	1.5	0.9	-0.5	100.5	1.1	1.6	0.5	99.5
26	3.1	3.2	0.2	99.8	0.3	0.2	-0.2	100.2
27	1.6	0.7	-0.8	100.3	4.8	4.6	-0.2	100.1
28	0.9	0.6	-0.4	100.3	2.0	3.3	1.3	99.0
29	-1.0	-0.8	0.3	99.6	5.6	5.3	-0.3	100.6
30	1.4	0.6	-0.7	101.8	2.9	3.1	0.2	99.5
31	3.3	1.8	-1.5	102.1	0.8	2.1	1.3	98.2
32	2.5	2.9	0.5	98.6	2.2	1.9	-0.4	101.1
33	-2.3	-2.9	-0.7	101.9	7.8	8.4	0.6	98.4
34	-1.9	-0.5	1.4	95.5	7.4	6.1	-1.3	104.2
MAE		2.0	0.7			5.0	0.7	
MAX		9.8	2.1			15.5	2.2	

summarized the results in Table III. The results have not been corrected for basis set superposition error, although such calculations are ideal candidates for dual-basis methods [31]. Because diffuse functions are important for accurately describing non-covalently bonded systems, we adopted aug-cc-pVTZ as the target secondary basis.

Columns 2–6 of Table III show the results obtained when using 6-31G(*d*) as the primary basis set. On average, MP2[V] does a good job by recovering 93% of

the MP2/aug-cc-pVTZ interaction energies. However, in the worst case, ethene dimer, it recovers only 84%. Adding a single set of diffuse functions to the primary basis (columns 7–11) significantly improves the results with even the worst case recovering almost 95% of the interaction energy which amounts to an error of only 1.3 kJ/mol. These results suggest that it is important that the primary basis contains some diffuse functions when performing calculations on weakly bound systems.

TABLE III: Errors (kJ/mol) in absolute and interaction energies using MP2[V] with respect to MP2/aug-cc-pVTZ benchmark energies.

Complex	MP2[V]/6-31G(d)/aug-cc-pVTZ					MP2[V]/6-31+G(d)/aug-cc-pVTZ				
	Dimer	Monomer	Monomer	Int	%	Dimer	Monomer	Monomer	Int	%
Ammonia dimer (C <sub>2h</sub> )	-3.3	-1.7	-1.7	-0.1	99.3	3.2	2.0	2.0	0.7	94.9
Water dimer (C <sub>s</sub> )	-3.4	-1.8	-1.7	-0.1	99.5	5.0	2.8	2.9	0.7	97.2
Formic acid dimer (C <sub>2h</sub> )	-4.0	-3.8	-3.8	-3.6	95.8	8.5	4.0	4.0	-0.5	99.5
Formamide dimer (C <sub>2h</sub> )	-9.2	-6.4	-6.4	-3.5	95.3	5.8	2.8	2.8	-0.3	99.6
Uracil dimer (C <sub>2h</sub> )	-15.4	-9.8	-9.8	-4.1	96.0	14.3	6.6	6.6	-1.2	98.8
2-Pyridoxine 2-aminopyridine (C <sub>1</sub> )	-18.3	-11.5	-10.2	-3.5	96.0	8.0	4.1	4.2	0.3	99.7
Adenine thymine WC (C <sub>1</sub> )	-15.9	-11.9	-9.4	-5.4	93.8	13.6	7.0	6.3	0.3	99.6
Methane dimer (D <sub>3d</sub> )	-2.4	-1.3	-1.3	-0.1	97.2	-1.3	-0.6	-0.6	0.1	97.1
Ethene dimer (D <sub>2d</sub> )	-10.8	-6.2	-6.2	-1.6	84.4	-0.5	-0.2	-0.2	0.1	99.2
Benzene methane (C <sub>3</sub> )	-10.1	-10.0	-1.2	-1.1	92.7	2.1	3.0	-0.6	0.3	97.6
Benzene dimer (C <sub>2h</sub> )	-15.2	-10.0	-10.0	-4.8	85.6	5.2	3.0	3.0	0.9	97.4
Pyrazine dimer (C <sub>s</sub> )	-10.0	-7.1	-7.1	-4.2	89.9	6.8	3.6	3.6	0.5	98.8
Uracil dimer stack (C <sub>2</sub> )	-14.2	-9.5	-9.5	-4.9	92.7	12.6	6.6	6.6	0.6	99.1
Indole benzene stack (C <sub>1</sub> )	-18.8	-14.2	-9.9	-5.3	89.9	6.1	4.3	3.0	1.3	97.6
Adenine thymine stack (C <sub>1</sub> )	-14.6	-11.2	-9.1	-5.8	93.7	11.0	6.9	6.4	2.3	97.5
Ethene ethyne (C <sub>2v</sub> )	-10.6	-6.2	-6.0	-1.6	86.5	1.6	-0.2	1.9	0.2	98.6
Benzene water (C <sub>s</sub> )	-9.2	-10.0	-1.7	-2.4	88.7	5.7	3.0	2.9	0.2	99.1
Benzene ammonia (C <sub>s</sub> )	-9.9	-10.0	-1.5	-1.7	90.7	4.7	3.0	2.0	0.3	98.2
Benzene HCN (C <sub>s</sub> )	-11.6	-10.0	-3.3	-1.7	95.1	4.1	3.0	1.8	0.7	97.8
Benzene dimer (C <sub>2v</sub> )	-17.6	-10.0	-10.1	-2.5	91.4	5.7	3.0	3.0	0.4	98.7
Indole benzene T-shape (C <sub>1</sub> )	-20.2	-14.2	-10.0	-4.1	91.3	6.9	4.3	3.0	0.4	99.2
Phenol dimer (C <sub>1</sub> )	-15.8	-8.9	-8.9	-2.0	95.5	9.4	4.9	4.9	0.4	99.1
MAE				2.9					0.6	
MAX				5.8					2.3	

### C. Efficiency Study

Table I indicates the theoretical speed-ups that are possible when using MP2[V]. However, there are many factors that can affect the efficiency of the implementation of such a method and the real advantage of MP2[V] can only be appreciated by considering actual timing data. Figure 2 summarizes the timing results for all the calculations performed for the isomerization reactions in Section III B. The calculations were carried out using a 2.93 GHz Intel Nehalem processor, 3 GB of DDR3 RAM, and 20 GB of scratch disk space.

The plot shows the speed-ups relative to the cost of the conventional MP2 calculation as a function of the number of secondary basis functions. Calculations where the conventional MP2 calculation had insufficient disk, and therefore had to recalculate the AO-ERIs, are indicated with a cross. The speed-up is roughly a factor of five, trending upwards (as seen from the linear regression) for larger systems due to the conventional algorithm having to recalculate the integrals. The cc-pVDZ/cc-pVQZ basis set pairing typically has a  $N/n$  ratio of 4.5 which, in principle, should translate into a cost reduction factor of

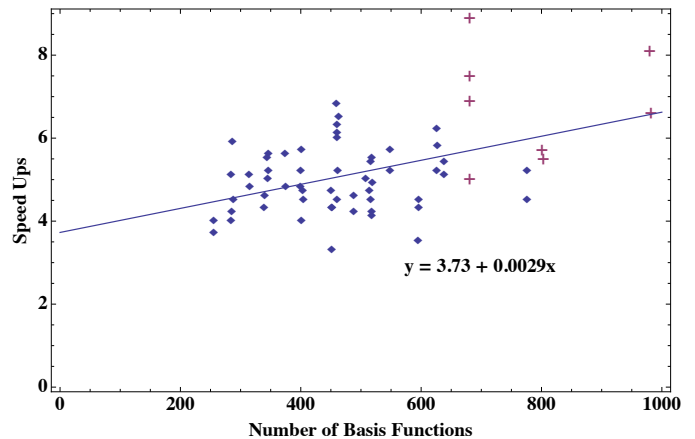


FIG. 2: The speed-ups of MP2[V] relative to conventional MP2 for the molecules in the isomerization set. Crosses indicate cases where the conventional MP2 algorithm required two batches and therefore two evaluations of the AO-ERIs.

the  $(N/n)^2$ . However, integral cutoffs and prescreening, which are key to efficient integral packages, reduces this

to ( $N/n$ ).

For much larger systems, our modified algorithm will also need to recompute the AO-ERIs. However, the disk storage requirements for MP2[V] are lower by a factor of  $N/2n$  and therefore it will never require more batches than the conventional algorithm.

#### IV. CONCLUDING REMARKS

We have proposed a simplification to our previously published dual-basis MP2[K] scheme for bootstrapping a small basis MP2 calculation to a larger basis. This new method, which is termed MP2[V], accounts for the relaxation effects of virtual orbitals upon basis set enlargement, while neglecting such effects for occupied-orbitals.

We have tested the accuracy and efficiency of the MP2[V] scheme on sets of organic reactions and non-covalent interactions and our results show that, with an adequate choice of primary basis, MP2[V] yields excellent agreement with large secondary basis MP2 benchmarks at approximately 20% of the cost. The accuracy of the results obtained using MP2[V] justifies the underlying

assumption of the method - that the relaxation of the occupied orbitals can be neglected when expanding the basis set.

Density Fitting is another popular technique for improving the efficiency of MP2 calculation and one may wonder if it can be used in combination with MP2[V] for even greater speed-ups. Our initial investigations suggest that the potential speed-ups are insubstantial.

While our assessment of both accuracy and efficiency focused on conventional MP2 calculations, we expect the computational savings to carry over to other methods which utilize second-order perturbative energies, such as the popular "double-hybrid" DFT functionals.

#### Acknowledgments

We thank Pierre-François Loos for useful comments on the manuscript. P.M.W.G thanks the Australian Research Council for funding (Grant Nos. DP1094170 and DP120104740) and APAC for a generous allocation of supercomputer resources. J.D thanks the ANU/RSC for a PhD scholarship.

- 
- [1] A. Szabo and N. S. Ostlund, *Modern quantum chemistry* (McGraw-Hill, New York, 1989).
- [2] D. Cremer, *WIREs Comput. Mol. Sci* **1**, 509530 (2011).
- [3] S. Grimme, *J. Chem. Phys* **118**, 9095 (2003).
- [4] Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, *J. Chem. Phys* **121**, 9793 (2004).
- [5] S. Grimme, *J. Chem. Phys* **124**, 034108 (2006).
- [6] S. Saebo and P. Pulay, *Ann. Rev. Phys. Chem* **44**, 213 (1993).
- [7] M. Schutz, G. Hetzer, and H. J. Werner, *J. Chem. Phys* **111**, 5691 (1999).
- [8] P. Ayala and G. E. Scuseria, *J. Chem. Phys* **110**, 3660 (1999).
- [9] P. Ayala, K. Kudin, and G. E. Scuseria, *J. Chem. Phys* **115**, 9698 (2001).
- [10] B. Doser, D. Lambrecht, J. Kussmann, and C. Ochsenfeld, *J. Chem. Phys* **130**, 064107 (2009).
- [11] M. S. Lee, P. E. Maslen, and M. Head-Gordon, *J. Chem. Phys* **112**, 3592 (2000).
- [12] M. Feyereisen, G. Fitzgerald, and A. Komornicki, *chem. phys. lett* **208**, 359 (1993).
- [13] F. Weigend, M. Haser, H. Patzelt, and R. Ahlrichs, *chem. phys. lett* **294**, 143 (1998).
- [14] U. Bozkaya, *J. Chem. Phys* **141**, 124108 (2014).
- [15] F. Aquilante, T. B. Pedersen, and R. Lindh, *J. Chem. Phys* **126**, 194106 (2007).
- [16] R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnald, W. T. Pollard, B. D. Dunietz, and Y. X. Cao, *J. Phys. Chem. A* **103**, 1913 (1999).
- [17] L. Vogt, R. Olivares-Amaya, Y. S. S. Kermes, C. Amador-Bedolla, and A. Aspuru-Guzik, *J. Chem. Theory Comput.* **112**, 2049 (2008).
- [18] R. Olivares-Amaya, M. A. Watson, R. G. Edgar, L. Vogt, Y. Shao, and A. Aspuru-Guzik, *J. Chem. Theory Comput.* **6**, 135 (2010).
- [19] M. A. Watson, R. Olivares-Amaya, R. G. Edgar, and A. Aspuru-Guzik, *Computing in Science and Engineering* **12**, 40 (2010), ISSN 1521-9615.
- [20] J. Deng, A. T. B. Gilbert, and P. M. W. Gill, *J. Chem. Phys.* **130**, 231101 (2009).
- [21] J. Deng, A. T. B. Gilbert, and P. M. W. Gill, *J. Chem. Phys.* **133**, 044116 (2010).
- [22] T. Nakajima and K. Hirao, *J. Chem. Phys.* **124**, 184108 (2006).
- [23] J. Deng, A. T. B. Gilbert, and P. M. W. Gill, *Phys. Chem. Chem. Phys.* **12**, 10759 (2010).
- [24] R. P. Steele, Y. Shao, R. A. DiStasio, Jr, and M. Head-Gordon, *J. Phys. Chem. A.* **110**, 13915 (2006).
- [25] J. Deng and P. M. W. Gill, *J. Chem. Phys.* **12**, 081103 (2011).
- [26] S. Saebo and J. Almlof, *Chem. Phys. Lett* **154**, 83 (1989).
- [27] M. J. Frisch, M. Head-Gordon, and J. A. Pople, *Chem. Phys. Lett* **166**, 281 (1990).
- [28] S. Grimme, M. Steinmetz, and M. Korth, *J. Org. Chem.* **72**, 2118 (2007).
- [29] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).
- [30] Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, et al., *Mol. Phys* **113**, 184 (2015).
- [31] J. Deng, A. T. B. Gilbert, and P. M. W. Gill, *J. Chem. Phys* **135**, 081105 (2011).