

# Atomic Partitioning of the MP<sub>n</sub> (n=2,3,4) Dynamic Electron Correlation Energy by the Interacting Quantum Atoms (IQA) Method: a Fast and Accurate Electrostatic Potential Integral Approach

Mark A. Vincent, Arnaldo F. Silva and Paul L. A. Popelier

Manchester Institute of Biotechnology, The University of Manchester,  
Manchester, M1 7DN, Great Britain  
and

School of Chemistry, The University of Manchester, Manchester, M13 9PL, Great Britain

**Abstract:** Recently the quantum topological energy partitioning method called Interacting Quantum Atoms (IQA) has been extended to MP<sub>n</sub> (n=2,3,4) wavefunctions. This enables the extraction of chemical insight related to dynamic electron correlation. The large computational expense of the IQA-MP<sub>n</sub> approach is compensated by the advantages that IQA offers compared to older non-topological energy decomposition schemes. This expense is problematic in the construction of a machine learning training set to create kriging models for topological atoms. However, the algorithm presented here markedly accelerates the calculation of atomically partitioned electron correlation energies. Then again, the algorithm cannot calculate pairwise interatomic energies because it applies analytical integrals over *whole* space (rather than over atomic volumes). However, these pairwise energies are not needed in the quantum topological force field FFLUX, which only uses the energy of an atom interacting with all remaining atoms of the system that it is part of. Thus it is now feasible to generate accurate and sizeable training sets at MP<sub>n</sub> level of theory.

## 1. Introduction

Atomistic force fields continue to be developed<sup>1-10</sup> given their growing use and impact in biological and materials science. Extensive comparisons<sup>11-14</sup> between popular, traditional force fields can offer guidance in this development, as well as comparisons<sup>15,16</sup> between machine-learned potentials. Within this grand scheme of force field development the modelling of dispersion<sup>17</sup> energy has received special attention over the last decade and more.

Dispersion is normally defined within the context<sup>18</sup> of long-range Rayleigh-Schrödinger perturbation theory, which is typically used in the description of two or more interacting molecules. However, it is possible to circumvent perturbation theory and focus directly on the electron correlation, from a single supermolecular system. This is the route we follow here. We calculate dynamical correlation energy extracted from the two-particle density-matrix (2PDM) of this single supermolecular (quantum) system. This energy encompasses dispersion effects. We thus apply an alternative method to obtain dispersion-like interaction energy between molecules.

A real-space partitioning, such as the quantum topological one<sup>19</sup>, of the energy associated with electron correlation allows dispersion to be quantified within a single atom. As a result, *intra*-molecular dispersion is also defined and quantifiable, which is not the case within a perturbation approach. The capacity to look inside an atom and assess the extra stability it experiences due to electron correlation yields deeper insight in the phenomenon of dispersion, beyond that offered by the traditional Lennard-Jones intermolecular dispersion. In this work we adopt the (topological) real-space method of Interacting Quantum Atoms (IQA)<sup>20</sup> to provide dispersion energies, both within a given atom and between *any pair of* atoms.

Already in 2005, quantum topological correlation energies were calculated<sup>20</sup> for a few very small molecules at full CI level (for H<sub>2</sub> and He<sub>2</sub>) and for complete-active-space multiconfigurational

wavefunctions. Much later the first IQA-partitioned electron correlation energies were computed in the context of coupled cluster theory<sup>21</sup>, coupled-cluster Lagrangian densities<sup>22</sup> and CCSD(T) wave functions<sup>23</sup>. Very recently we made possible IQA-MPn for the first time<sup>24,25</sup>, applied<sup>26</sup> the machine learning method Gaussian Process Regression (also known as kriging) to the energy contributions, quantified<sup>25</sup> electron correlation of the chemical bond (showing that ionicity and covalency are not each other's opposite), demonstrated the transferability of topologically partitioned electron correlation energies in water clusters<sup>27</sup>, studied<sup>28</sup> the effects of higher orders of perturbation theory on the correlation energy of atoms and bonds in molecules, and finally calculated<sup>29</sup> dynamic electron correlation in a wider variety of systems including glycine...water (hydration), the ethene dimer ( $\pi$ - $\pi$  interactions), benzene (aromaticity), cyclobutadiene (antiaromaticity), and  $\text{NH}_3\text{BH}_3$  (dative bond).

The work just described focuses on chemical insight and hence the most detailed information possible, i.e. obtained at full atomic resolution. Indeed, IQA-MPn then gives access to all interatomic energies (denoted A,B) and all intra-atomic energies (denoted A,A) where A and B are (topological) atoms. In other words, the interatomic energies are known between *any two* atoms A and B wherever they are in the system; the atoms do not need to be bonded. Secondly, it should be emphasised that this information is separated from the intra-atomic one. This high resolution contrasts with the coarser atomic resolution that is the subject of the current article. Here we only discern the *total* interaction of a given atom with all other atoms in the system (denoted AA'), including its self-interaction (denoted AA). This restriction is the price paid for the faster algorithm presented here, and the reason for this price will become clear in the next section. However, such overall interaction between a given atom and its environment is sufficient for the development of our quantum topological based force field, called FFLUX<sup>30-34</sup>. This force field uses kriging to deliver fully polarizable multipolar electrostatics alongside intra- and interatomic non-electrostatic energies and charge transfer as a function of flexible molecular geometries. After all, FFLUX only needs to know the extent by which a given atom interacts with *all* other atoms. This atomic information suffices to calculate the force<sup>35</sup> on the nucleus of the atom in question, which then allows the propagating of the system in time, leading to optimised geometries or structural and thermodynamic properties in the case of a molecular dynamics simulation. Indeed, a force field always sums all the individual pairwise interactions that a given atom is involved in. It is only this sum that is needed. Knowing the individual terms would only be useful if detailed chemical insight were being kept track of, which can be done<sup>34</sup> in principle but this is computationally expensive.

In summary, we present a fast computational procedure in order to equip the topological force field FFLUX with dynamic electron correlation and thus any dispersion effects. This procedure involves an analytical integration over whole space and thereby reduces the expensive six-dimensional (6D) integration to a three-dimensional (3D) integration.

## 2. Background and Method

### 2.1 Derivation

We have presented our original approach in detail elsewhere<sup>24</sup>, which we briefly summarise here. The key equation is eq 1,

$$V_{ee,corr}^{AB} = \sum_{j=1}^{N_{basis}} \sum_{k=1}^j K_{jk} \sum_{l=1}^{N_{basis}} \sum_{m=1}^l K_{lm} d_{jklm}^{corr} \int_{\Omega_A} d\mathbf{r}_1 G_{jk}(\mathbf{r}_1 - \mathbf{R}_{jk}) \int_{\Omega_B} d\mathbf{r}_2 \frac{1}{r_{12}} G_{lm}(\mathbf{r}_2 - \mathbf{R}_{lm}) \quad (1)$$

where  $V$  is the (electron) correlation energy between any two atoms A and B (note that the intra-atomic correlation energy, i.e. when  $A=B$ , is also covered),  $d$  is the correlated part of the 2PDM, the  $G$  functions are Gaussians arising from products of Gaussian primitives originally centred at nuclei,  $K$  refers to the product pre-factor (and should of course not be confused with summation index  $k$ ), and there are  $N_{basis}$  Gaussian primitives. There are two consecutive 3D integrations, each over the volume of a topological atom  $\Omega$ , introducing its own integration variable ( $\mathbf{r}_1$  or  $\mathbf{r}_2$ ). Traditionally, the resulting 6D integration is carried out completely numerically, after installing a quadrature grid (a grid, in short) over each of the respective atomic 3D volumes.

Now we show how an analytical integration over whole space can be introduced. The starting point of the derivation below is the calculation of  $V_{ee,corr}^{AA'}$ , which is obtained as the sum of the contributions from all atoms  $B$ . Substituting eq 1 into this sum leads to the derivation shown in eq 2,

$$\begin{aligned}
V_{ee,corr}^{AA'} &= \sum_B V_{ee,corr}^{AB} \\
&= \sum_B \sum_{j=1}^{N_{basis}} \sum_{k=1}^j K_{jk} \sum_{l=1}^{N_{basis}} \sum_{m=1}^l K_{lm} d_{jklm}^{corr} \int_{\Omega_A} d\mathbf{r}_1 G_{jk}(\mathbf{r}_1 - \mathbf{R}_{jk}) \int_{\Omega_B} d\mathbf{r}_2 \frac{1}{r_{12}} G_{lm}(\mathbf{r}_2 - \mathbf{R}_{lm}) \\
&= \sum_{j=1}^{N_{basis}} \sum_{k=1}^j K_{jk} \sum_{l=1}^{N_{basis}} \sum_{m=1}^l K_{lm} d_{jklm}^{corr} \int_{\Omega_A} d\mathbf{r}_1 G_{jk}(\mathbf{r}_1 - \mathbf{R}_{jk}) \sum_B \int_{\Omega_B} d\mathbf{r}_2 \frac{1}{r_{12}} G_{lm}(\mathbf{r}_2 - \mathbf{R}_{lm}) \quad (2) \\
&= \sum_{j=1}^{N_{basis}} \sum_{k=1}^j K_{jk} \sum_{l=1}^{N_{basis}} \sum_{m=1}^l K_{lm} d_{jklm}^{corr} \int_{\Omega_A} d\mathbf{r}_1 G_{jk}(\mathbf{r}_1 - \mathbf{R}_{jk}) \underbrace{\int_{full\ space} d\mathbf{r}_2 \frac{1}{r_{12}} G_{lm}(\mathbf{r}_2 - \mathbf{R}_{lm})}_{analytical} \\
&= \sum_{j=1}^{N_{basis}} \sum_{k=1}^j K_{jk} \sum_{l=1}^{N_{basis}} \sum_{m=1}^l K_{lm} d_{jklm}^{corr} \int_{\Omega_A} d\mathbf{r}_1 G_{jk}(\mathbf{r}_1 - \mathbf{R}_{jk}) V_{lm}^{ESP}(\mathbf{r}_1)
\end{aligned}$$

where the summation index  $B$  is allowed to be equal to  $A$  (and indeed must be in one term).

The transformation from the second to the third line in the above derivation is justified by the linearity of integration (i.e. the sum of an integral is the integral of the sum, in short) and the distributivity of summation versus multiplication.

It is clear in the third line that the partition over atom  $B$  disappears and thus the integral is carried out over whole space. Hence, the integral can be calculated analytically, using ideas reminiscent of the calculation of the electrostatic potential (ESP). This integral technology has matured some time ago culminating in the elaborate PRISM algorithm<sup>36</sup>, which has been implemented in the program GAUSSIAN09<sup>37</sup> (or G09, in short) used in this work. Note that, although we use the label ESP, we actually do *not* calculate the electrostatic potential itself. The latter involves the electron density (in the integrand) rather than a Gaussian, as is the case in eq 2. Indeed, the electrostatic potential is a molecular property but here we face an integral involving only atomic orbitals. Note also that the penultimate expression in eq 2 cannot be re-written such that the actual ESP appears. In summary, the net advantage is that a numerical grid, one for each atom, has disappeared, and with it an expensive integral.

The purpose of this article is to quantify the improvement over the original 6D algorithm of the new ESP-based algorithm (discussed in detail below). A simple measure that helps is achieving this is the so-called *recovery error*, which is defined as the difference between the system's correlation energy directly obtained from G09 and that obtained (i.e. recovered) from the sum of atomic contributions  $V_{ee,corr}^{AA'}$ , or

$$\Delta E_{ee,corr} = V_{ee,corr}^{G09} - \sum_A V_{ee,corr}^{AA'} \quad (3)$$

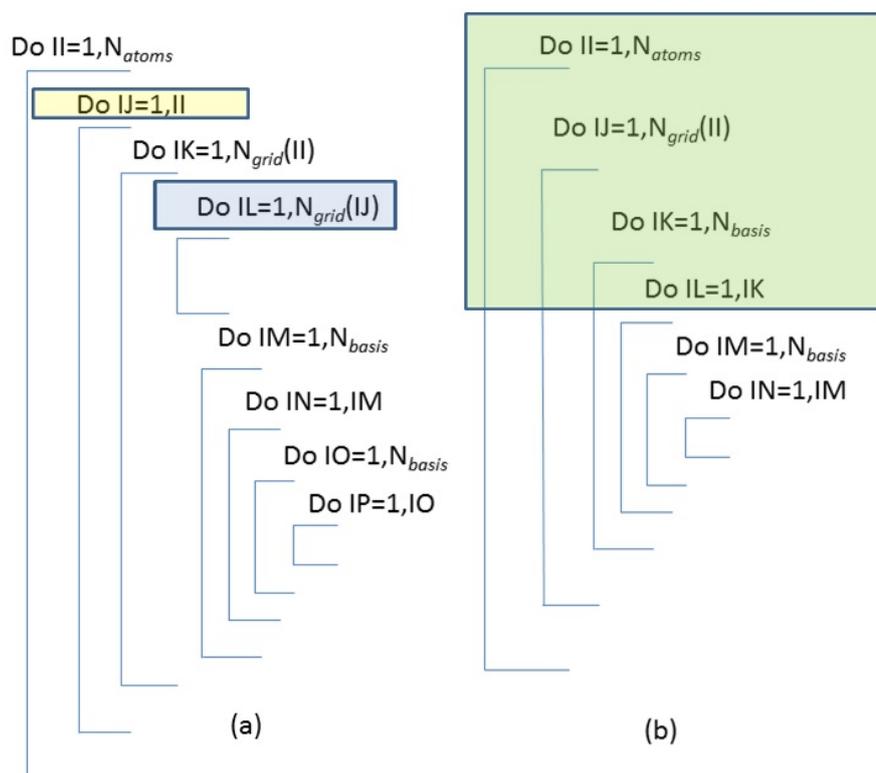
We will report this recovery error for a wide variety of systems, alongside with estimates of CPU timings.

## 2.2 Implementation of the 3D ESP algorithm

In order to evaluate eq 2 we have modified a link of G09 (called L1111) to generate the correlated part of the 2PDM (denoted  $d$ ). In addition, the standard wavefunction file ("wfn") is outputted from G09. These two pieces of information are then passed to our in-house code called MORFI, which generates a grid for the molecule under study, atom by atom. This grid (over  $\Omega_A$ ) is then passed back to G09 to generate the ESP integrals<sup>38</sup> at each grid point  $\mathbf{r}_1$  (see eq 2). The nuclear part of an ESP integral is not needed. Because only

the electronic part is required we have further modified G09 to obtain these values. The latter are then used as input for another MORFI run, where the atomic energies are calculated from  $d$  via eq 2.

Figure 1 compares the implementation of the original *6D algorithm* (left,  $V_{ee,corr}^{AB}$ ) with that of the new *3D ESP algorithm* (right,  $V_{ee,corr}^{AA'}$ ). Both implementations consist of a nested DO-loop structure, introducing three numbers:  $N_{atoms}$ ,  $N_{grid}$  and  $N_{basis}$ . These counters respectively refer to the number of atoms, the number of (quadrature) grid points of each atom and the number of Gaussian primitive basis functions. The original 6D algorithm is computationally expensive because it involves a two-fold nested DO-loop over the number of atoms covering all atom-atom pairs (including AA self-interaction) ( $N_{atoms}(N_{atoms}+1)/2$  or approximately  $N_{atoms}^2$ ), one DO-loop over the grid points installed on each atom's volume ( $N_{grid}$ ) (hence two nested DO-loops in total), and four DO-loops over the number of basis functions ( $\left[ \frac{N_{basis}(N_{basis}+1)}{2} \times \frac{N_{basis}(N_{basis}+1)}{2} \right] \approx \frac{1}{4} N_{basis}^4$ ). As a result the number of executions of the innermost DO-loop (over index IP in Figure 1) scales roughly as  $N_{atoms}^2 N_{grid} N_{basis}^4$ . In addition, there is another DO-loop (over index IL in Figure 1) over the grid that does not include the  $N_{basis}^4$  DO-loops. This large number of DO-loops has confined the application of  $V_{ee,corr}^{AB}$  to small systems only, with some of our larger systems studied being the water pentamer<sup>27</sup>, glycine<sup>28</sup> and glycine...water<sup>29</sup>.



**Figure 1.** A schematic representation of the DO-loops involved in the running of the program MORFI by the (a) original 6D algorithm and the (b) new 3D ESP algorithm. The two DO-loops missing in the latter (but present in the former), are indicated in (a) as yellow and blue. In addition, the DO-loops indicated in green in (b) actually correspond to the total number of ESP integrals. Hence the green part can be thought of as a single repeated action consisting of reading and processing the ESP integrals.

In summing over the atoms, one obtains a *total* energy of an atom in the molecule rather than a self-energy plus all the pairwise energies of that atom with all the other atoms in the molecule. Thus, the decrease in computational effort demands as a price the loss of detailed energy contributions. Naively, one would expect a massive speed-up, by a factor of several thousands, due to the elimination of a grid per atom (i.e.  $N_{atom}N_{grid}$ ) and  $N_{grid}$  is of the order of a thousand. However, this is not quite the case. Indeed, there are actually approximately  $N_{atom}N_{grid}N_{basis}^2/2$  ESP integrals to be calculated, which is a significant number. Then again, we have replaced one numerical 3D volume integration by, in theory, an analytic one, which means we expect increased numerical accuracy.

We have employed the uncontracted 6-31++G(2d,2p) basis set with the MP4SDQ (including core orbitals) method to study a series of molecules and complexes. The basis sets are in their uncontracted form because MORFI works in primitive functions and therefore a contracted basis set offers no advantage in the most time-consuming stage of the calculation. Indeed, it takes longer to generate the MP4SDQ wavefunction using the uncontracted basis and to optimize the geometry but generally this extra effort lengthens the overall calculation time by a relatively small amount. Although we employ the MP4SDQ 2PDM in this study, any correlation method's 2PDM can actually be employed. We have used the MP2 and MP3 methods to generate the respective 2PDM in previous studies<sup>25,28</sup>. In these cases the density matrices are the same size as the MP4SDQ 2PDM. Only the matrix elements differ, meaning that the calculation time in MORFI is the same irrespective of the wavefunction. However, MP2 can sometimes have more zero matrix elements than MP3 and MP4SDQ.

Our aim is twofold: (a) to understand how the size of the grid affects the recovery error when the 3D ESP approach is employed, and (b) for a fixed recovery error, to find out how much faster the 3D ESP approach is than the original 6D approach. For this purpose we conducted two types of calculations, one for each aim, respectively discussed in Sections 3a and 3b.

### 3. Results and Discussion

#### 3.1 Grid size to achieve a given recovery error

We first consider the size of grid necessary to achieve a given recovery error in the correlation energy of the whole system. Note that the FFLUX force field aims for the recovery error be of the order of  $1 \text{ kJmol}^{-1}$  prior to training. Table 1 gives some grid sizes and recovery errors for the water dimer,  $F_2$  and the neon dimer, which have all been geometry-optimized.

Grid $n_{\text{ang}}-n_{\text{rad}}^{\text{a}}$	No grid points <sup>b</sup>	MP4SDQ Correlation energy of G09	6D recovery error (MORFI)	3D ESP recovery error (MORFI)
<b>Water Dimer</b>				
1-10	120	-1508.3	+772.8	+2.5
3-10	520	-1508.3	-31.6	-0.5
5-10	1000	-1508.3	-68.0	+0.2
10-10	3400	-1508.3	-39.0	+0.1
5-20	2000	-1508.3	+37.8	-0.1
10-20	6800	-1508.3	-17.4	-0.1
<b><math>F_2</math></b>				
1-10	120	-1992.4	+274.2	+15.4
3-10	520	-1992.4	-57.2	-0.05
5-10	1000	-1992.4	-59.1	+0.5
10-10	3400	-1992.4	-11.8	+0.2
<b>Neon Dimer</b>				
1-10	120	-1388.7	-586.1	+20.3
3-10	520	-1388.7	-233.7	+20.2
10-10	3400	-1388.7	43.9	+20.2
15-10	7000	-1388.7	102.0	+20.2
10-20	6800	-1388.7	-49.2	+0.4
15-20	14,000	-1388.7	-22.5	+0.4
10-30	10,200	-1388.7	-40.1	-0.0
15-30	21,000	-1388.7	-24.2	-0.0
20-40	47,200	-1388.7	-14.6	+0.0

<sup>a</sup> In the notation  $n_{\text{ang}}-n_{\text{rad}}$  the first number is the Lebedev angular part of the grid while the second number is the radial part of the grid, given by Gauss-Legendre quadrature. The number  $n_{\text{ang}}$  refers to a grid shorthand adopting the values 1, 3, 5, 10, 15 or 20, which respectively designate angular grids of 6, 26, 50, 170, 350 and 590 points. The number  $n_{\text{rad}}$ , on the other hand, directly designates the number of radial points in the Gauss-Legendre grid.

<sup>b</sup> The total number of grid points is the number of Gauss-Legendre radial points multiplied by the number of Lebedev angular points, multiplied by two (one grid inside the  $\beta$ -sphere and one outside the  $\beta$ -sphere).

**Table 1.** The recovery errors (calculated by MORFI) of the MP4SDQ energy of the water dimer,  $F_2$  and the neon dimer. Units:  $\text{kJmol}^{-1}$ .

Table 1 contrasts the recovery errors obtained with the new 3D ESP integration with those of the original 6D method, for a variety of grids. It is clear that the errors are dramatically reduced in going from 6D to 3D, by one to two orders of magnitude. Table 1 also shows that, for the neon dimer, the recovery error of the 3D ESP method follows the radial part of the grid and not the angular part (so  $n_{\text{rad}}$  in  $n_{\text{ang}}-n_{\text{rad}}$ ). Indeed, the recovery errors are nearly constant within each of the  $n_{\text{ang}}-10$ ,  $n_{\text{ang}}-20$  and  $n_{\text{ang}}-30$  sets of grids. However, this observation is not true for the water dimer or for  $F_2$ . Nevertheless the error generally becomes smaller for all three systems, for both 6D and 3D ESP integration, with increasing angular grid. For the neon dimer the 10-20 grid, with the 3D ESP approach, yields energies with an error of  $0.4 \text{ kJmol}^{-1}$ , while for the water dimer and  $F_2$  errors of -0.5 and -0.05 are respectively observed for the 3-10 grid. These are acceptable errors in the context of the FFLUX force field. However, they do imply that the electronic

structure of different molecules call for different grids in order to obtain a similar absolute accuracy in the recovery error with the 3D ESP method. For example, 3-10 for (H<sub>2</sub>O)<sub>2</sub> as opposed to 10-20 for Ne<sub>2</sub>.

In Table 2 we present the results for a variety of molecules studied with the 10-20 grid only. This grid was selected because for the neon dimer it was the smallest grid that already gave an acceptable recovery error by the 3D ESP algorithm (Table 1, 0.4 kJmol<sup>-1</sup>). The 3D ESP approach significantly reduces the recovery error compared to the original 6D method for all systems, except for BH<sub>3</sub> and BeH<sub>3</sub><sup>-</sup>. This reduction is typically at least one order of magnitude but often two orders. However, for BH<sub>3</sub>, BeH<sub>3</sub><sup>-</sup>, there is barely a reduction. In absolute terms, 4 systems show an error of more than 1 kJmol<sup>-1</sup> (BH<sub>3</sub>, BeH<sub>3</sub><sup>-</sup>, AlH<sub>3</sub> and MgH<sub>3</sub><sup>-</sup>), and 5 systems (Ne<sub>2</sub>, CO, SiH<sub>4</sub>, HCl and FHF<sup>-</sup>) show medium sized errors (0.4 to 0.7 kJmol<sup>-1</sup>).

We now look into the reason for the abnormal behaviour of BH<sub>3</sub> and BeH<sub>3</sub><sup>-</sup>, starting with BH<sub>3</sub>. On studying BH<sub>4</sub><sup>-</sup>, a similar situation arose as observed for BH<sub>3</sub>. A test where BH<sub>4</sub><sup>-</sup> was given the geometry and the basis set of CH<sub>4</sub>, resulted in the same small error reduction seen for BH<sub>3</sub> whereas CH<sub>4</sub> behaves normally. Thus, the behaviour of boron is intrinsic to this element. In order to consider this behaviour further we increased the grid size. It turns out that the recovery error decreases significantly (from 0.44 to -0.02 kJmol<sup>-1</sup>) with the larger 15-30 grid. This reduction (22-fold) is typical and seen in all regular systems. The situation of the iso-electronic BeH<sub>3</sub><sup>-</sup> is more complicated than that of BH<sub>3</sub>. If the 15-30 grid is employed then the 6D result is marginally more accurate than the 3D ESP result (+0.59 *versus* +0.60 kJmol<sup>-1</sup>). In fact one has to employ an angular grid larger, than any used before, and coded "23" (770 angular grid points) before the 3D ESP approach gives an order of magnitude reduction in the recovery error over the 6D method. This great sensitivity to the angular grid size suggests that the elements *Be* and *B* show anisotropy in their description when bonded.

System (Molecule or complex)	MP4SDQ energy of G09	6D recovery error (MORFI)	3D ESP recovery error (MORFI)
LiH	-177.4	-3.5	-0.1
BeH <sub>3</sub> <sup>-</sup>	-383.3	+2.9	+2.4
BH <sub>3</sub>	-423.5	+2.1	+1.1
CH <sub>4</sub>	-635.3	-1.5	-0.0
NH <sub>3</sub>	-707.5	-5.9	-0.0
H <sub>2</sub> O	-752.0	-9.5	-0.0
HF	-751.9	-13.1	+0.1
N <sub>2</sub>	-1130.4	-14.0	-0.0
O <sub>2</sub>	-1266.1	-14.2	-0.0
F <sub>2</sub>	-1456.6	-23.8	-0.1
Ne <sub>2</sub>	-1388.7	-49.2	+0.4
NO	-1186.0	-16.1	-0.0
CO	-1087.7	-17.3	-0.4
OF	-1302.5	-19.7	-0.1
NaH	-575.0	-15.4	+0.0
MgH <sub>2</sub>	-643.8	-15.0	+0.1
MgH <sub>3</sub> <sup>-</sup>	-730.9	-13.5	+1.1
AlH <sub>3</sub>	-753.5	-10.9	+1.5
SiH <sub>4</sub>	-884.7	-10.4	+0.6
PH <sub>3</sub>	-930.8	-16.6	+0.0
SH <sup>-</sup>	-936.8	-24.0	+0.2
HCl	-952.9	-21.6	-0.7
FHF <sup>-</sup>	-1536.5	-30.6	+0.6
(H <sub>2</sub> O) <sub>2</sub>	-1508.3	-17.4	-0.1
(H <sub>2</sub> O) <sub>3</sub>	-2271.6	-24.4	-0.1

**Table 2.** The recovery errors (kJmol<sup>-1</sup>) for the 6D and 3D ESP integration for a variety of systems. All results have been calculated with the 10-20 grid (170 angular Lebedev points and 20 radial Gauss-Legendre points, or 6800 points in total, for the volume inside and outside the β-sphere).

Table 3 gives the CPU timings of the 6D and 3D ESP method, as well as the number of ESP integrals, for all systems given in Tables 1 and 2. The ratio of the timings between the two methods is only indicative (in principle) because (i) our computer cluster is heterogeneous in terms of its processors and (ii) even if the jobs run on the same type of processor, that processor may have a different load due other jobs running. Both these factors affect the timing. Table 3 shows that the ratio of the timings varies from 3 to 52, with most occurring in the 5 to 10 range. The time taken to calculate the ESP integrals has not been taken into account. These calculations typically take from a few minutes to half an hour. For example, for the water trimer there were just short of a billion ESP integrals to calculate (934,418,898) and this task was carried out in 92 groups (4 groups running concurrently on the current batch system). All batches calculated the same number of integrals (10,264,674), other than the last one, which calculated 333,564 ESP integrals. The fastest group (excluding the last one), employing a single core, took 1 min 16 sec of CPU time. However, the overall calculation can be even faster. Indeed, in calculating these ESP integrals, the G09 run includes the calculation of a Hartree-Fock wavefunction, prior to the computation of the required integrals. Technically, it is the ESP integrals over atomic basis functions that are needed. In other words, we do not require a wavefunction but only the atomic orbital basis and the grid positions. Thus, the wavefunction calculation is superfluous and clocks up unnecessary CPU time.

System	CPU Time: 6D MORFI	CPU Time: 3D ESP MORFI	CPU time ratio: 6D/3D	Number of ESP INTEGRALS
LiH	118	3	39	16,951,998
BeH <sub>3</sub> <sup>-a</sup>	963 (4599)	77 (89)	13 (52)	71,136,312 (220,443,132)
BH <sub>3</sub> <sup>a</sup>	306 (2027)	28 (106)	11 (19)	71,238,888 (220,258,488)
CH <sub>4</sub>	766	148	5	117,963,345
NH <sub>3</sub>	358	26	14	71,538,648
H <sub>2</sub> O	92	10	9	38,663,685
HF	31	4	8	17,519,376
N <sub>2</sub>	101	14	7	41,050,836
O <sub>2</sub>	83	17	5	40,669,980
F <sub>2</sub>	109	17	6	40,248,318
Ne <sub>2</sub>	109	17	6	40,248,318
NO	83	17	5	40,756,716
CO	90	17	5	41,010,030
OF	109	14	8	40,533,960
NaH	108	28	4	37,881,570
MgH <sub>2</sub>	290	32	9	74,103,696
MgH <sub>3</sub> <sup>-</sup>	745	89	8	124,947,972
AlH <sub>3</sub>	822	316	3	124,947,972
SiH <sub>4</sub>	1358	197	7	193,862,505
PH <sub>3</sub>	1520	200	8	125,600,868
SH <sup>-</sup>	413	13	32	38,327,170
HCl	84	15	6	37,881,570
FHF <sup>-</sup>	323	39	8	77,000,922
(H <sub>2</sub> O) <sub>2</sub>	3971	328	12	283,381,584
(H <sub>2</sub> O) <sub>3</sub> <sup>b</sup>	~24,000	3072	8	934,418,898

a) The number that is not bracketed refers to the 10-20 grid while the bracketed number is for the 15-30 grid, which is only employed for B and Be.

b) The size of the water trimer meant that the full MORFI run had to be split into several runs. Variation in processor speed and other jobs running concurrently on the node affect the timings of each of these jobs. Thus this timing is approximate and is a lower bound.

**Table 3.** The CPU timings (minutes) for each of the systems studied by the 6D and 3D approach, the ratio of the two timings rounded to the nearest integer and the number of ESP integral. All results are for the 10-20 grid unless stated otherwise.

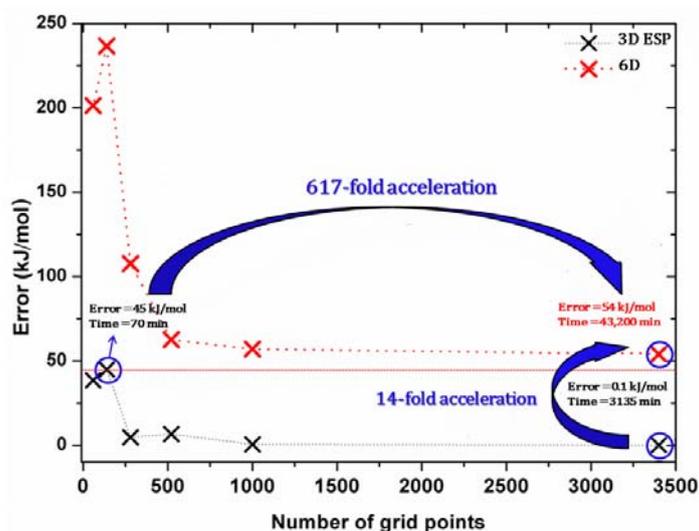
### 3.2 Algorithm speed up (from 6D to 3D ESP) for a fixed recovery error

As mentioned in Section 2.2 there are two ways of quantifying the progress made by the 3D ESP algorithm: (a) how the size of the grid affects the recovery error and (b) how much faster 3D ESP is compared to 6D for a fixed recovery error. Here we look at the second way.

The timings shown in Table 3 are impressive but still do not do justice to the improvements achieved by changing the algorithm from a 6D to a 3D ESP integration. Comparing the results using *the same grid* for both methods is fair and useful. Indeed, the very small recovery errors, often obtained by the 3D ESP method with small to medium grids, are impressive. However, for the purpose of constructing a training set for FFLUX, larger recovery errors are acceptable. Hence, much computational time can be saved by employing smaller grids by increasing the recovery error to a fixed value that we are happy with. Thus we analyse the 3D ESP and 6D method by equalizing their respective recovery errors and then comparing the 2 CPU times. By doing this it becomes possible to really appreciate the improvements achieved by the 3D ESP version of the code in terms of speed-up. We carry out an analysis on a non-trivial system and one that FFLUX benefits from, given its ultimate aim to work for peptides in aqueous solution. Looking at the list of systems at hand the choice then falls on the water trimer. We fix an acceptable recovery error ( $0.1 \text{ kJmol}^{-1}$  for example) and search for the smallest grid available that can achieve this accuracy.

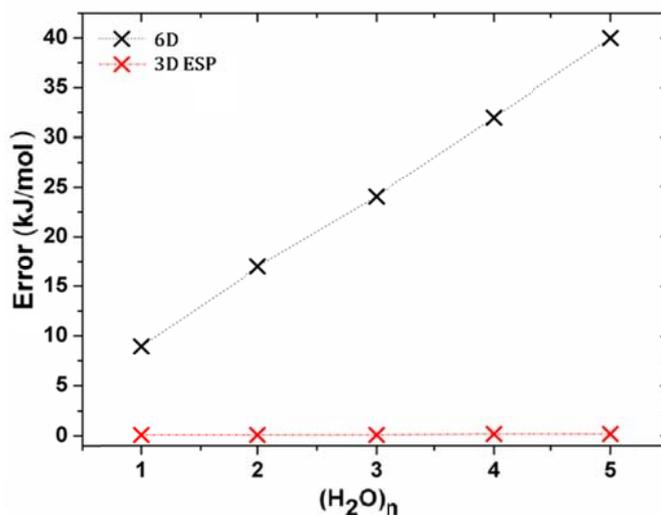
Figure 2 contrasts the performance of the 6D and 3D ESP methods. Unfortunately the 6D method cannot reach a recovery error of  $\sim 50 \text{ kJmol}^{-1}$  (see red dotted line in Figure 2) with less than the maximum of 3500 grid points tested. Thus, this is the practical threshold (rather than  $1 \text{ kJmol}^{-1}$ ) allowing a direct comparison between the two methods. For the 6D method it takes 43,200 minutes to reach this recovery error while for the 3D ESP method it takes 70 minutes, or 617 times less. These timings are obtained for two very different numbers of grid points: 6D needs many points while 3D ESP needs few. However, if the same grid size is employed by both methods then the speed-up is about 14 (43,200 *versus* 3135 minutes). This grid optimization also shows that while it is possible to achieve a recovery error as small as  $0.1 \text{ kJmol}^{-1}$  using  $\sim 3500$  grid points, a still respectful error of  $0.6 \text{ kJmol}^{-1}$  can be achieved with only 1000 grid points, in about 150 minutes.

In principle, the grid size can be varied from atom to atom, as we have reported<sup>27</sup> for water clusters by showing that hydrogen atoms required smaller grids in order to obtain good recovery errors compared to oxygen atoms. However, the implementation of the ESP 3D integration in MORFI had an unforeseen secondary effect: the required number of grid points to achieve small recovery errors is significantly lower when compared to the 6D integration. In fact the ESP 3D approach makes the use of different grids for different atoms obsolete, since the required grids for both heavy and hydrogen atoms can be pushed to their lower limit.



**Figure 2.** A comparison between the 6D and 3D ESP methods implemented in MORFI for the water trimer showing the recovery error (summed over all 9 atoms), employing different grid sizes. The straight red dotted line marks the approximate recovery error at which the 6D and 3D ESP methods are compared.

There is one final test to be done, which will further show the power of the 3D ESP method. We applied both the 6D and 3D ESP method to investigate a series of growing water clusters ( $\text{H}_2\text{O}$ )<sub>n</sub> (n=1, 2, 3, 4 and 5). Figure 3 plots the recovery error for each of these five water clusters. While the recovery error increases almost linearly with the number of atoms for the 6D method, this is not the case for the 3D ESP method. Once a sufficiently robust grid (10-10, 3400 grid points) is used, the error remains more or less constant for the latter method. While the 6D recovery error starts off about 9 kJmol<sup>-1</sup> for the water monomer, it becomes as large as 40 kJmol<sup>-1</sup> for the water pentamer. However, for 3D ESP, the error remains around 0.1-0.2 kJmol<sup>-1</sup> throughout the series, regardless of the system size. This result is superb for the incorporation of dynamic electron correlation (and thus dispersion) in FFLUX.



**Figure 3.** The recovery error for the 6D method increases linearly with the size of the system, while it remains constant for the 3D ESP method. The grid used was 10-10 (3400 grid points).

The 6D version of MORFI will remain relevant within the context of the Quantum Chemical Topology literature, because it provides much physical insight, especially related to the nature of chemical bonds. Nonetheless, the fact that the recovery error obtained through the 6D algorithm heavily depends on the system size creates an unsustainable situation to transition FFLUX from small molecules to large biological systems. However, this is not the case for the 3D ESP algorithm because the recovery error is largely independent of the system size. One can use this size independency to the benefit of FFLUX, by first carrying out a grid optimisation for, say, an amino acid and then apply the knowledge of this optimal grid to also obtain excellent recovery errors for polypeptides.

#### 4. Conclusions

The Møller-Plesset dynamic electron correlation of topological atoms can now be calculated 3 to 50 times faster for a given fixed size of quadrature grid. This is possible by replacing the numerical integration over the atomic volume by an analytical integration over whole space. The price paid is the loss of pairwise inter-atomic electron energies. However, these are fortunately not required for the force field FFLUX, which only needs the correlation energy of a given atom interacting with all other atoms by summation. The total recovery error for all atoms in the water pentamer now amounts to only 0.6 kJmol<sup>-1</sup> for a grid of only 3400 points. Moreover, for this system the recovery error is largely independent of the system size.

## AUTHOR INFORMATION

### Corresponding Author

\*Phone: +44 161 3064511. E-mail: [pla@manchester.ac.uk](mailto:pla@manchester.ac.uk)

### Funding Sources

P.L.A.P. acknowledges the EPSRC for funding through the award of an Established Career Fellowship (grant EP/K005472) and Dr A. F. Silva thanks the Rutherford Fund Strategic Partner Award, the Brazilian government's FAPESP and CAPES for the award of postdoctoral position number 2014/21241-9.

### References

- (1) Choi, J.-M.; Pappu, R. V. Improvements to the ABSINTH Force Field for Proteins Based on Experimentally Derived Amino Acid Specific Backbone Conformational Statistics *J.Chem.Theor.Comput.* **2019**, *15*, 1367.
- (2) Visscher, K. M.; Geerke, D. P. Deriving Force-Field Parameters from First Principles Using a Polarizable and Higher Order Dispersion Model *J.Chem.Theor.Comput.* **2019**, *15*, 1875.
- (3) Lin, F.-Y.; MacKerell, A. D. Polarizable Empirical Force Field for Halogen-Containing Compounds Based on the Classical Drude Oscillator *J.Chem.Theory Comput.* **2018**, *14*, 1083.
- (4) Wu, X.; Clavaguera, C.; Lagardère, L.; Piquemal, J.-P.; de la Lande, A. AMOEBA Polarizable Force Field Parameters of the Heme Cofactor in Its Ferrous and Ferric Forms *J.Chem.Theor.Comput.* **2018**, *14*, 2705.
- (5) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field *J.Chem.Theor.Comput.* **2015**, *11*, 2609.
- (6) Silva, T. F. D.; Vila-Viçosa, D.; Reis, P. B. P. S.; Victor, B. L.; Diem, M.; Oostenbrink, C.; Machuqueiro, M. The Impact of Using Single Atomistic Long-Range Cutoff Schemes with the GROMOS 54A7 Force Field *J.Chem.Theor.Comput.* **2018**, *14*, 5823.
- (7) Das, A. K.; Demerdash, O. N.; Head-Gordon, T. Improvements to the AMOEBA Force Field by Introducing Anisotropic Atomic Polarizability of the Water Molecule *J.Chem.Theor.Comput.* **2018**, *14*, 6722.
- (8) Shannon, R. J.; Hornung, B.; Tew, D. P.; Glowacki, D. R. Anharmonic Molecular Mechanics: Ab Initio Based Morse Parametrizations for the Popular MM3 Force Field *J.Phys.Chem.A* **2019**, *123*, 2991.
- (9) Perez, A.; MacCallum, J. L.; Brini, E.; Simmerling, C.; Dill, K. A. Grid-Based Backbone Correction to the ff12SB Protein Force Field for Implicit-Solvent Simulations *J.Chem.Theor.Comput.* **2015**, *11*, 4770.
- (10) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning *J.Chem.Theor.Comput.* **2016**, *12*, 2312.
- (11) Zhang, H.; Yin, C.; Jiang, Y.; van der Spoel, D. Force Field Benchmark of Amino Acids: I. Hydration and Diffusion in Different Water Models *J.Chem.Inf.Model.* **2018**, *58*, 1037.
- (12) Zhang, H.; Jiang, Y.; Cui, Z.; Yin, C. Force Field Benchmark of Amino Acids. 2. Partition Coefficients between Water and Organic Solvents *J.Chem.Inf.Model.* **2018**, *58*, 1669.
- (13) Dean Smith, M.; Rao, S.; Segelken, E.; Cruz, L. Force-Field Induced Bias in the Structure of A $\beta$ 21–30: A Comparison of OPLS, AMBER, CHARMM, and GROMOS Force Fields *J.Chem.Inf.Model.* **2015**, *55*, 2587–2595.
- (14) Kuzmanic, A.; Pritchard, R. B.; Hansen, D. F.; Gervasio, F. L. Importance of the Force Field Choice in Capturing Functionally Relevant Dynamics in the von Willebrand Factor *J.Phys.Chem.Lett.* **2019**, 1928.
- (15) Nguyen, T. T.; Szekely, E.; Imbalzano, G.; Behler, J.; Csanyi, G.; Ceriotti, M.; Gotz, A. W.; Paesani, F. Comparison of permutationally invariant polynomials, neural networks, and Gaussian approximation potentials in representing water interactions through many-body expansions *J.Chem.Phys.* **2018**, *148*, 241725.
- (16) Li, J.; Song, K.; Behler, J. A critical comparison of neural network potentials for molecular reaction dynamics with exact permutation symmetry **2019**, Advance Article DOI: 10.1039/C8CP06919K.
- (17) Klimeš, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory *J.Chem.Phys.* **2012**, *137*, 120901.
- (18) Stone, A. J. *The Theory of Intermolecular Forces. Second Edition.*; 1 ed.; Clarendon Press: Oxford, 2013; Vol. 32.

- (19) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory.*; Oxford Univ. Press: Oxford, Great Britain, 1990.
- (20) Blanco, M. A.; Martín Pendás, A.; Francisco, E. Interacting quantum atoms: a correlated energy decomposition scheme based on the quantum theory of atoms in molecules. *J.Chem.Theor.Comput.* **2005**, *1*, 1096.
- (21) Chávez-Calvillo, R.; García-Revilla, M.; Francisco, E.; Martín Pendás, A.; Rocha-Rinza, T. Dynamical correlation within the Interacting Quantum Atoms method through coupled cluster theory *Comput.Theor.Chem.* **2015**, *1053*, 90.
- (22) Holguin-Gallego, F. J.; Chavez-Calvillo, R.; Garcia-Revilla, M.; Francisco, E.; Martin Pendas, A.; Rocha-Rinza, T. Electron correlation in the interacting quantum atoms partition via coupled-cluster lagrangian densities *J.Comput.Chem.* **2016**, *37*, 1753.
- (23) Casalz-Sainz, J. L.; Guevara-Vela, J. M.; Francisco, E.; Rocha-Rinza, T.; Martin Pendas, A. Where Does Electron Correlation Lie? Some Answers from a Real Space Partition *ChemPhysChem* **2017**, *18*, 3553.
- (24) McDonagh, J. L.; Vincent, M. A.; Popelier, P. L. A. Partitioning dynamic electron correlation energy: viewing Møller-Plesset correlation energies through Interacting Quantum Atom (IQA) energy partitioning *Chem.Phys.Lett.* **2016**, *662*, 228.
- (25) McDonagh, J. L.; Silva, A. F.; Vincent, M. A.; Popelier, P. L. A. Quantifying electron correlation of the chemical bond *J.Phys.Chem.Letts.* **2017**, *8*, 1937.
- (26) McDonagh, J. L.; da Silva, A.; Vincent, M. A.; Popelier, P. L. A. Machine Learning of Dynamic Electron Correlation Energies from Topological Atoms *J.Chem.Theor.Comput.* **2018**, *14*, 216–224.
- (27) Silva, A. F.; Vincent, M. A.; McDonagh, J. L.; Popelier, P. L. A. The Transferability of Topologically Partitioned Electron Correlation Energies in Water Clusters *ChemPhysChem* **2017**, *18*, 3360.
- (28) Vincent, M. A.; Silva, A. F.; McDonagh, J. L.; Popelier, P. L. A. The effects of higher orders of perturbation theory on the correlation energy of atoms and bonds in molecules. *Int.J.Quant.Chem.* **2018**, *118*, e25519.
- (29) Silva, A. F.; Popelier, P. L. A. MP2-IQA: upscaling the analysis of topologically partitioned electron correlation *J.Mol.Mod.* **2018**, *24*, 201 (10 pages).
- (30) Popelier, P. L. A. QCTFF: On the Construction of a Novel Protein Force Field *Int.J.Quant.Chem.* **2015**, *115*, 1005.
- (31) Popelier, P. L. A. In *Modern Charge-Density Analysis*; Gatti, C., Macchi, P., Eds.; Springer, Germany 2012; Vol. 14, p 505.
- (32) Popelier, P. L. A. Molecular Simulation by Knowledgeable Quantum Atoms *Phys.Scr.* **2016**, *91*, 033007.
- (33) Thacker, J. C. R.; Wilson, A. L.; Hughes, Z. E.; Burn, M. J.; Maxwell, P. I.; Popelier, P. L. A. Towards the simulation of biomolecules: optimisation of peptide-capped glycine using FFLUX *Mol.Simul.* **2018**, *44*, 881.
- (34) Zielinski, F.; Maxwell, P. I.; Fletcher, T. L.; Davie, S. J.; Di Pasquale, N.; Cardamone, S.; Mills, M. J. L.; Popelier, P. L. A. Geometry Optimization with Machine Trained Topological Atoms *Scientific Reports* **2017**, *7*, 12817.
- (35) Mills, M. J. L.; Popelier, P. L. A. Electrostatic Forces: formulae for the first derivatives of a polarisable, anisotropic electrostatic potential energy function based on machine learning. *J.Chem.Theory Comput.* **2014**, *10*, 3840–3856.
- (36) Gill, P. M. W.; Pople, J. A. The prism algorithm for two-electron integrals *Int.J.Quant.Chem.* **1991**, *40*, 753.
- (37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Wallingford CT, 2009.
- (38) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. Computing molecular electrostatic potentials with the PRISM algorithm *Chem.Phys.Lett.* **1993**, *206*, 239.