Open access • Journal Article • DOI:10.1021/CT300531W

Second-Order Møller-Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach.

Mauro Del Ben, Juerg Hutter, Joost VandeVondele

Institutions: University of Zurich, ETH Zurich

Published on: 24 Sep 2012 - Journal of Chemical Theory and Computation (American Chemical Society)

Topics: Basis function, Basis set, Gaussian, Poisson's equation and Møller–Plesset perturbation theory

Second-order moller-plesset perturbation theory in the condensed phase: an efficient and massively parallel gaussian and plane waves approach

Del Ben, Mauro ; Hutter, Juerg ; VandeVondele, Joost

Abstract: A novel algorithm, based on a hybrid Gaussian and plane waves (GPW) approach, is developed for the canonical second order Moller-Plesset perturbation energy (MP2) of finite and extended systems. The key aspect of the method is that the electron repulsion integrals (ia vertical bar lambda sigma) are computed by direct integration between the products of Gaussian basis functions lambda sigma and the electrostatic potential arising from a given occupied virtual pair density ia. The electrostatic potential is obtained in a plane waves basis set after solving the Poisson equation in Fourier space. In particular, for condensed phase systems, this scheme is highly efficient. Furthermore, our implementation has low memory requirements and displays excellent parallel scalability up to 100 000 processes. In this way, canonical MP2 calculations for condensed phase systems containing hundreds of atoms or more than 5000 basis functions can be performed within minutes, while systems up to 1000 atoms and 10 000 basis functions remain feasible. Solid LiH has been employed as a benchmark to study basis set and system size convergence. Lattice constants and cohesive energies of various molecular crystals have been studied with MP2 and double-hybrid functionals.

DOI: https://doi.org/10.1021/ct300531w

Second Order Møller-Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach

Mauro Del Ben,*,† Jürg Hutter,*,† and Joost VandeVondele*,‡

Institute of Physical Chemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and Department of Materials, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland

E-mail: delben@pci.uzh.ch; hutter@pci.uzh.ch; Joost.VandeVondele@mat.ethz.ch

Abstract

A novel algorithm, based on a hybrid Gaussian and Plane Waves (GPW) approach, is developed for the canonical second-order Møller-Plesset perturbation energy (MP2) of finite and extended system. The key aspect of the method is that the electron repulsion integrals $(ia|\lambda\sigma)$ are computed by direct integration between the products of Gaussian basis functions $\lambda\sigma$ and the electrostatic potential arising from a given occupied-virtual pair density ia. The electrostatic potential is obtained in a plane waves basis set after solving the Poisson equation in Fourier space. In particular for condensed phase systems, this scheme is highly efficient. Furthermore, our implementation has low memory requirements and displays excellent parallel scalability up to 100 000 processes. In this way, canonical MP2 calculations for condensed phase systems containing hundreds of atoms or more than 5000 basis functions can be performed

*To whom correspondence should be addressed

†Institute of Physical Chemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

‡Department of Materials, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland

within minutes. Lattice constants and cohesive energies of various molecular crystals have been studied with MP2 and double hybrid functionals.

1 Introduction

The second-order Møller-Plesset (MP2) energy is a relatively cheap and effective correction to the Hartree-Fock ground state energy that accounts for electron correlation effects. It is obtained from Rayleigh-Schrödinger perturbation theory, for which the zero order Hamiltonian is chosen as the sum of the one electron Fock operators[1,2]. MP2 is appealing because it recovers a relatively large part of the dynamic correlation, and maintains an easy and compact formulation. Most notably, MP2 introduces dispersion, which is an essential non-covalent interaction. With the introduction of Double-Hybrid Density Functionals[3,4], MP2-like correlation has also established itself in Density Functional Theory (DFT). In Double-Hybrids, an MP2-like term obtained from the Kohn-Sham orbitals and eigenvalues is mixed into the correlation energy. However, the advantages of MP2 come at a computational cost that is high compared to that of Hartree-Fock or traditional DFT. In its canonical formulation, MP2 scales a $O(n^5)$, where n represents the number of basis functions, and a large amount of memory is needed to store the intermediates of the calculation. Furthermore, MP2 calculations need larger basis sets than DFT to reach a similar convergence. In order to extend the applicability of MP2 to large system these limitations have to be overcome[5].

Various reformulations of the MP2 energy expression, and new algorithms, have been proposed to address these limitations. Reducing the formal $O(n^5)$ scaling is achieved with methods such as local MP2[6–15] (LMP2) and Laplace-Transformed MP2[16–22]. The prefactor of the various terms that dominate for smaller systems can be reduced with the resolution of identity approximation[23–28] (RI-MP2), while explicitly correlated methods speedup the convergence of the MP2 energy with respect to basis set size[29] (F12-MP2). Despite this progress, calculations with good basis sets on systems containing fifty or more heavy atoms remain computationally demanding with MP2 or double hybrid DFT. In order to perform such calculations with acceptable time to solution,

massively parallel computing is becoming an indispensable tool. A variety of MP2 algorithms suitable for parallel architectures has been proposed[30–40] and these algorithms have demonstrated good scalability up to a few hundred cores.

The development of efficient parallel algorithms is of prime interest in the case of MP2 energy calculation with periodic boundary condition[12,22,27,28,41–43]. In fact, in this case, due to the considerably high cost, the practical applications are limited to periodic systems with small unit cell. XXXXXXXXXXXXX we need to say a word on periodic MP2 as well, including citations to Kresse XXXXXXXXXXXXX

Here, we present a novel MP2 algorithm that is particularly suitable for the condensed phase and has been designed to achieve excellent scalability on modern massively parallel architectures having 1000s-100000s of cores. The prefactor of the $O(n^5)$ term is minimal and the memory usage per core is small. It is based on the Gaussian and Plane Wave (GPW) approach[44], which allows for avoiding the computation of four center electron repulsion integrals (ERI) over Gaussian basis functions $(\mu\nu|\lambda\sigma)$. In conventional canonical MP2 algorithms, the computation of these integrals and their transformation into the molecular orbital basis is usually the most time-consuming step. This step is furthermore difficult to parallelize efficiently, involving significant communication and difficult load balancing issues. In the Gaussian and Plane Wave MP2 (GPW-MP2) method, half transformed ERIs of the type $(ia|\lambda\sigma)$ are directly computed in a communication free way. This is achieved by the direct computation of the electrostatic potential of the pair density $\rho^{ia} = \psi_i \cdot \psi_a$ in an auxiliary plane waves basis by means of Fast Fourier Transforms (FFTs) and the numerical integration of this potential in real space over products of pairs of Gaussian basis functions $\lambda\sigma$. With this strategy, only fully transformed ERIs $(ia|jb)$ are communicated for the calculation of the exchange like part of the MP2 energy. The efficiency derives from the use of regular auxiliary grids and FFTs, which distinguishes the method from other approaches employing numerical integration.[45,46] We report parallel scalability up to 100 000 processes with 80% efficiency, allowing calculations on molecular crystals containing more than 5000 basis functions within minutes. We further validate the GPW-MP2 method by performing calculations on molecular crystals with extended basis sets.

2 The Gaussian and Plane Wave MP2 method

In the canonical orbital formalism, the closed shell MP2 correlation energy $E^{(2)}$ is obtained as

$$E^{(2)} = -\sum_{ij,ab}^{occ,vir} \frac{(ia|jb)[2(ia|jb)-(ib|ja)]}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j}. \tag{1}$$

Indices i, j refer to occupied and a, b to virtual canonical orbitals, and ε_p to the corresponding orbital energy. The ERIs over molecular orbitals (MO ERI) are given by

$$(ia|jb) = \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\psi_j(\vec{r}_2)\psi_b(\vec{r}_2)d\vec{r}_1 d\vec{r}_2 \tag{2}$$

and conventionally computed by a four index transformation from ERIs over atomic orbitals (AO ERI) $(\mu\nu|\lambda\sigma)$ as

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

where $C_{\mu i}$ represent elements of the MOs coefficient matrix and Greek indices refer to AOs, a linear combination of Gaussian basis functions in our approach. For systems described by periodic boundary conditions (PBC), Brillouin sampling is implicitly implied for Eq. (1), but here we will assume that Γ-point sampling is sufficient for systems with a sufficiently large unit cell and band gap.[41,42,47] In the periodic case, AOs and the integrals in Eq. (2) must take the PBC into account.[44]

The basis of the Gaussian and Plane Wave MP2 method (GPW-MP2) is the direct formulation of the half transformed ERIs based on the electrostatic potential v^{ia} of the pair density ρ^{ia}

$$\begin{aligned}
(ia|\lambda\sigma) &= \int\int \psi_i(\vec{r}_1)\psi_a(\vec{r}_1)\frac{1}{\vec{r}_{12}}\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_1 d\vec{r}_2 &(4)\\
&= \int\left[\int \frac{\psi_i(\vec{r}_1)\psi_a(\vec{r}_1)}{\vec{r}_{12}}d\vec{r}_1\right]\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2 &(5)\\
&= \int\left[\int \frac{\rho^{ia}(\vec{r}_1)}{\vec{r}_{12}}d\vec{r}_1\right]\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2 &(6)\\
&= \int v^{ia}(\vec{r}_2)\phi_\lambda(\vec{r}_2)\phi_\sigma(\vec{r}_2)d\vec{r}_2 &(7)
\end{aligned}$$

The form of the last equation is essentially identical to the one used in the GPW method[44] to compute matrix elements of the Hartree potential. Thus, the highly efficient implementation of that operation in CP2K[48] can be directly used and we refer to Ref.[49] for a detailed discussion.

Central in the GPW method is the representation of the density $\rho^{ia}(\vec{r})$ on a regular grid, which can be considered equivalent to an expansion of the density in an auxiliary basis of plane waves (PW). The expansion is given by

$$\rho^{ia}(\vec{r}) \approx \frac{1}{\Omega} \sum_{|\vec{G}| \leq G_c} \rho^{ia}(\vec{G}) e^{i\vec{G}\cdot\vec{r}} \tag{8}$$

where the sum over the reciprocal lattice vectors $\vec{G}$ is determined by the resolution of the grid. $\rho^{ia}(\vec{G})$ are the Fourier coefficients of the density, and Ω is the volume of the simulation cell. Conventionally, the resolution of the grid is specified as the energy cutoff $\frac{1}{2}G_c^2$ that limits the kinetic energy of the PWs. Fast Fourier transforms (FFTs) efficiently change representation between real space ($\rho^{ia}(\vec{R})$) and reciprocal space ($\rho^{ia}(\vec{G})$). In particular, for a grid with S grid points, the transformation can be performed in linear scaling time ($O(S\log S)$). In reciprocal space, it becomes straightforward to solve the Poisson equation for the potential v^{ia}

$$v^{ia}(\vec{G}) = \frac{4\pi}{G^2}\rho^{ia}(\vec{G}) \tag{9}$$

and an additional back FFT (FFT^{-1}) will yield the potential in real space. The orthonormality of the orbitals implies that $\rho^{ia}(\vec{G}=0) = 0$ and divergence at $\vec{G}=0$ is thus avoided.[50] Note that the PW auxiliary basis is a natural choice for periodic systems, but it can equally be used for gas phase or surface calculations. Indeed, once the density is specified on a regular grid efficient methods are available for solving the Poisson equation with free (for example cluster or slab) boundary conditions[51–53]. The simplicity of the GPW method has as a drawback that all-electron calculations are not possible, and that pseudopotentials have to be employed in order to have densities that are smooth. The Gaussian and Augmented Plane Wave (GAPW) scheme[54,55] overcomes this limitation and is suitable for all-electron calculations. However, whereas this method is available in CP2K for

all-electron DFT calculations, our MP2 implementation is currently limited to the GPW method only.

Once the potential v^{ia} is available on a regular real space grid, the numerical integration over the basis functions is performed by summing the product of the value of the potential and the primitive Gaussian functions (PGFs) over the grid points. Within a given threshold (ε_{grid}), all non-zero matrix elements for a given pair *ia* can be obtained in linear scaling time. This is possible since only pairs of overlapping Gaussians need to be considered, and only a finite number of grid points within a spherical region around the center of the PGF is required. A further gain in efficiency is obtained by employing a multi-grid technique that represents the potential v^{ia} on grids with increasingly coarser grid spacing. Depending on the smoothness or width of the PGF, the appropriate grid is selected so that the number of points employed for the integration is essentially independent of the exponent of the PGF. The accuracy of the multi-grid scheme is fixed by specifying a relative cutoff (E_{cut}^{rel}) that specifies the E_{cut} of the grid that will be employed for a PGF with exponent 1.0.

Finally, $(ia|\lambda\sigma)$ integrals are transformed into MO ERIs using (sparse) matrix multiplication. Introducing for a given pair *ia* the matrix of half transformed ERIs $\mathbf{B}^{ia}$ ($(ia|\lambda\sigma) = B^{ia}_{\lambda\sigma}$), the matrix of MO ERIs $\mathbf{V}^{ia}$ is obtained by two index transformations as $\mathbf{V}^{ia} = \mathbf{C}_o^{\dagger}\mathbf{B}^{ia}\mathbf{C}_v$, where $\mathbf{C}_o$ and $\mathbf{C}_v$ represent the coefficient matrices of the occupied and virtual orbitals. The multiplication by $\mathbf{C}_v$ can exploit the sparsity of $\mathbf{B}^{ia}$, implying an $O(nv)$ scaling per *ia* pair, while the final multiplication can not exploit sparsity and is asymptotically dominant, scaling as $O(onv)$. o, v, and n refer to the number of occupied, virtual, and total orbitals respectively. The thresholding in the sparse matrix multiplication is enforced using a threshold $\varepsilon_{filter} \approx \varepsilon_{grid}$. As we will show below, the overall accuracy of the MP2 energy can be well controlled, and for the systems tested here, is on the order of $10^{-7} - 10^{-8}$ a.u. per heavy atom for $E_{cut} = 300$ Ry, $E_{cut}^{rel} = 50$ Ry, $\varepsilon_{filter} = \varepsilon_{grid} = 10^{-8}$. E_{cut} depends on the largest exponent of the basis used, while the other parameters are system independent.

3 Implementation of the Gaussian and Plane Wave MP2 method

3.1 Serial algorithm

Loop over all i occupied orbitals

  Calculate wavefunction $\psi_i(\vec{R})$ on the real space grid

  Store $\psi_i(\vec{R})$

End i Loop

Loop over all a virtual orbitals

  Calculate wavefunction $\psi_a(\vec{R})$ on the real space grid

  Loop over all i occupied orbitals

    Compute $\rho^{ia}(\vec{R}) = \psi_i(\vec{R}) * \psi_a(\vec{R})$ on the real space grid

    Transfer $\rho^{ia}(\vec{R}) \rightarrow \rho^{ia}(\vec{G})$: $\rho^{ia}(\vec{G}) = \mathbf{FFT}[\rho^{ia}(\vec{R})]$

    Solve the Poisson equation: $\rho^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{G})$

    Transfer $v^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{R})$: $v^{ia}(\vec{R}) = \mathbf{FFT^{-1}}[v^{ia}(\vec{G})]$

    Integrate potential in real space: $(ia|\lambda\sigma) = \mathrm{B}^{ia}_{\lambda\sigma} = \int v^{ia}(\vec{R})\phi_\lambda(\vec{R})\phi_\sigma(\vec{R})d\vec{R}$

    Index transformations $\mathbf{V}^{ia} = \mathbf{C}_o^\dagger(\mathbf{B}^{ia}\mathbf{C}_v)$

    Store $\mathbf{V}^{ia}$

  End i Loop

  $E^{(2)} = E^{(2)} + \sum_{ij,b} \frac{(ia|jb)[2(ia|jb)-(ja|ib)]}{\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b}$

End a loop

Figure 1: Pseudocode for the serial implementation of the GPW-MP2 energy.

The pseudocode for the serial algorithm is presented in Figure 1. In a first step, the wavefunctions of all occupied orbitals ($\psi_i(\vec{R})$) are precalculated on real space grids, which speeds up the calculation of the pair density ρ^{ia} in the main loop. In the next step, the outer loop iterates over all virtual orbitals a, while the inner one loops over occupied orbitals i. For each value of a, the MO ERIs $(i.|jb)$ for all other indices are available after the inner loop, making it possible to compute both Coulomb and exchange contributions to $E^{(2)}$ with $O(o^2v)$ memory usage, which compares favorably with the $O(on^2)$ memory required in the standard direct canonical algorithm. Other objects, such as the stored grids and molecular orbital coefficients, require memory scaling no worse than quadratically with system size. In the inner loop, all the grid operations and integral transformations for a given pair ia are performed. The density $\rho^{ia}(\vec{R})$ on the real space grid is simply obtained by multiplying

the values for $\psi_i(\vec{R})$ and $\psi_a(\vec{R})$, and is transformed by **FFT** to reciprocal space, where the Poisson equation is solved. An inverse FFT yields the potential in real space, which is used in the numerical integration procedure to yield the $(ia|\lambda\sigma)$ integrals. The latter are stored as a sparse matrix, which can directly be employed in the following index transformations that yield the MO ERIs $(ia|jb)$. Note again that the first transformation can exploit the sparsity of the $\mathbf{B}^{ia}$ matrix, while the second transformation can not. The main features of the serial GPW-MP2 algorithm are summarized in Table 1.

Table 1: Features of the serial and parallel GPW-MP2 energy algorithm expressed as "order of" the calculation parameters. n number of basis functions, o and v number of occupied and virtual orbitals, S grid size, N_i and N_a number of occupied and virtual divisions, N_G and N_w number of groups and group size, N_p number of processes. N_i, N_a, N_G, N_w and N_p are related by $N_G = N_i N_a$ and $N_p = N_G N_w$.

	Serial		Parallel	
	Memory	Execution Time	Memory	Execution Time
Wave functions calculation	oS	nS	$\frac{oS}{N_i N_w}$	$\left(\frac{o}{N_i}+\frac{v}{N_a}\right)\frac{S}{N_w}$
FFT and Poisson solver	S	$ovS\log S$	$\frac{S}{N_w}$	$\frac{ovS\log S}{N_p}$
v^{ia} integration	S	ovn	$\frac{S}{N_w}$	$\frac{ovn}{N_p}$
1^{st} quarter transformation	on	o^2vn	$\frac{on}{N_w}$	$\frac{o^2vn}{N_p}$
2^{nd} quarter transformation	ov	o^2v^2n	$\frac{ov}{N_w}$	$\frac{o^2v^2n}{N_p}$
MP2 energy contraction	o^2v	o^2v^2	$\frac{o^2v}{N_i N_w}$	$\frac{o^2v^2}{N_p}$
Communication	-	-	$\frac{o^2v}{N_i^2 N_w}$	$\frac{N_i-1}{N_i}\frac{o^2v^2}{N_p}$

3.2 Parallel algorithm

The parallel algorithm for the GPW-MP2 energy calculation has been designed to enable calculations on large systems and to display excellent scalability. This implies that the computational load and the amount of data communicated per process decreases linearly as the number of processes

Assign each process its coordinate-triplet (n_i, n_a, n_w)

Create ranges $[i_{start}^{n_i}, i_{end}^{n_i}]$, $[a_{start}^{n_a}, a_{end}^{n_a}]$, $[b_{start}^{n_w}, b_{end}^{n_w}]$

Loop over i occupied orbitals ($i_{start}^{n_i} \leq i \leq i_{end}^{n_i}$)

Calculate wavefunction $\psi_i(\vec{R})$ on the real space grid

Store $\psi_i(\vec{R})$

End i Loop

Loop over a virtual orbitals ($a_{start}^{n_a} \leq a \leq a_{end}^{n_a}$)

Calculate wavefunction $\psi_a(\vec{R})$ on the real space grid

Loop over i occupied orbitals ($i_{start}^{n_i} \leq i \leq i_{end}^{n_i}$)

Compute $\rho^{ia}(\vec{R}) = \psi_i(\vec{R}) * \psi_a(\vec{R})$ on the real space grid

Transfer $\rho^{ia}(\vec{R}) \rightarrow \rho^{ia}(\vec{G})$: $\rho^{ia}(\vec{G}) = \mathbf{FFT}[\rho^{ia}(\vec{R})]$

Solve Poisson's Equation: $\rho^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{G})$

Transfer $v^{ia}(\vec{G}) \rightarrow v^{ia}(\vec{R})$: $v^{ia}(\vec{R}) = \mathbf{FFT^{-1}}[v^{ia}(\vec{G})]$

Integrate Potential in real space: $(ia|\lambda\sigma) = \mathrm{B}_{\lambda\sigma}^{ia} = \int v^{ia}(\vec{R})\phi_\lambda(\vec{R})\phi_\sigma(\vec{R})d\vec{R}$

Index transformation $\mathbf{V}^{ia} = \mathbf{C}_o^\dagger(\mathbf{B}^{ia}\mathbf{C}_v)$

Redistribute and store $\mathbf{V}^{ia}$ (all j, $b_{start}^{n_w} \leq b \leq b_{end}^{n_w}$)

End i Loop

Loop over N_i processes with same n_a and n_w but different n_i

determine the coordinates n_i^S and n_i^R of sending and receiving process.

Receive $(ja|ib)$ from (n_i^R, n_a, n_w)

$(i_{start}^{n_i^R} \leq j \leq i_{end}^{n_i^R}, i_{start}^{n_i} \leq i \leq i_{end}^{n_i}, b_{start}^{n_w} \leq b \leq b_{end}^{n_w})$

Send $(ia|jb)$ to (n_i^S, n_a, n_w)

$(i_{start}^{n_i^S} \leq j \leq i_{end}^{n_i^S}, i_{start}^{n_i} \leq i \leq i_{end}^{n_i}, b_{start}^{n_w} \leq b \leq b_{end}^{n_w})$

$E^{(2)} = E^{(2)} + \sum_{ij,b} \frac{(ia|jb)[2(ia|jb)-(ja|ib)]}{\varepsilon_i+\varepsilon_j-\varepsilon_a-\varepsilon_b}$

$(i_{start}^{n_i^R} \leq j \leq i_{end}^{n_i^R}, i_{start}^{n_i} \leq i \leq i_{end}^{n_i}, b_{start}^{n_w} \leq b \leq b_{end}^{n_w})$

End loop over processes

End a loop

Global summation of $E^{(2)}$

Figure 2: Pseudocode of the parallel GPW-MP2 energy.

(N_p) increases. Furthermore, the computational load is easily balanced and the communication pattern involves relatively large messages between a small subset of all possible pairs of processes. Additionally, no significant data is fully replicated and memory usage can be decreased as the number of processes is increased. This is achieved by a multi-level parallelization scheme, and a careful process layout. The first level of parallelization corresponds to the work performed for a given *ia* pair. The second level of parallelization corresponds to a distribution of the nearly independent calculations for each of the *ia* pairs. The N_p processes are therefor split in N_G groups, each group working on a given *ia* pair and consisting of N_w processes ($N_p = N_G N_w$). The first level of parallelization is complicated, involving parallel FFTs, halo-exchanges, and sparse matrix multiplications over N_w processes. However, this level is readily available, as it corresponds to the standard parallelization scheme for DFT calculations in CP2K.[49] As a rule of thumb, reasonable speedups are observed as long as $N_w \leq o$, while memory usage benefits from the nearly perfect distribution of the grids and sparse matrices. Nevertheless, the best performance is obtained if groups do not communicate across nodes, ideally, if memory permits, $N_w = 1$. The second level of parallelization is more straightforward, as it only requires inter-process communications of fully transformed ERIs. In order to distribute the *ia* pairs efficiently, the occupied orbitals i and the virtual orbitals a are split into N_i and N_a disjoint ranges respectively. A 2D Cartesian layout with dimensions $N_i \times N_a$ is considered for the N_G groups ($N_G = N_i N_a$) giving each group coordinates (n_i, n_a) and the corresponding index ranges $[i^{n_i}_{start}, i^{n_i}_{end}]$ and $[a^{n_a}_{start}, a^{n_a}_{end}]$. Additionally, each of the N_w processes within a group is given an index n_w, so that a processes is uniquely identified by its coordinate triplet (n_i, n_a, n_w). Finally, the b index is split in N_w ranges $[b^{n_w}_{start}, b^{n_w}_{end}]$, while a splitting of j is not necessary.

The pseudo-code of the parallel algorithm is shown in Figure 2 and it follows the serial algorithm closely (but with restricted index ranges for i and a), until the end of the inner loop over i. Here, as a last step of the inner loop, the matrix of fully transformed integrals $\mathbf{V}^{ia}$ is redistributed within the group, so that the full range of j and the restricted range of b, corresponding the process' index n_w, is stored locally. After completion of the loop over i the inter-group communication takes

place, and here the benefit of the process layout becomes apparent. Indeed, only processes with identical coordinates n_a and n_w need to exchange data, i.e. only within small subgroups of size N_i communication takes place. This is due to the fact that for a given pair ab, only $(ia|jb)$ and $(ja|ib)$ need to be simultaneously available to a process. Since each process stores only ERIs $(ia|jb)$ for i in $[i_{start}^{n_i}, i_{end}^{n_i}]$ and all j, the locally held integrals can be contracted if integrals are received from all the other processes that store the ERIs for which $i \notin [i_{start}^{n_i}, i_{end}^{n_i}]$ for the current ab pair. This communication step is easily accomplished by employing the standard message passing interface (MPI) point-to-point communication protocol.[56] The fact that the computational effort for every ia pair is essentially the same, implies that processes arrive well synchronized at the communication step, contributing to the scalability of the algorithm. The size of each message send is $O\left(\frac{vo^2}{wN_i^2}\right)$, while the number of message exchanges, including the loop over a, is $O\left((N_i-1)\frac{v}{N_a}\right)$, yielding an expected communication time $O\left(\frac{N_i-1}{N_i}\frac{o^2v^2}{p}\right)$ in the bandwidth-limited regime. Note that if only the spin opposite (SO) component of the MP2 energy is required, the MO ERIs do not need to be communicated among processes, yielding an essentially communication free algorithm with reduced $O(on/N_w)$ memory usage. The main features of the parallel GPW-MP2 algorithm are summarized in Table 1.

4 Benchmark Calculations

4.1 Computational Details

Basis Sets, thresholds and pseudopotentials

The GPW-MP2 method as implemented in CP2K[48] has been employed for all calculations in this manuscript. The MP2 calculations are based on pseudopotentials of the form suggested by Goedecker, Teter and Hutter (GTH) in Ref.[57] but specifically parameterized for Hartree-Fock (HF) calculations. In this way, core states do not need to be represented and valence orbitals are smooth, as required by the GPW method. Valence-only basis sets have been generated for use

with these pseudopotentials, and are suitable for MP2, i.e. of the correlation consistent type. The basis sets have been labeled as cc-DZVP, cc-TZVP and cc-QZVP, denoting increasing quality. A family basis set scheme has been adopted, using the same set of exponents for each primitive with *s* and *p* l-quantum number. The number of primitive Gaussian functions (for the valence only) has been chosen to be 4, 5 and 6 for cc-DZVP, cc-TZVP and cc-QZVP respectively. The three primitive Gaussian functions with higher exponent have been used to generate a contracted Gaussian, while the others have been kept uncontracted as in the split valence scheme. For the cc-QZVP basis the most diffuse functions have exponents H=0.13906, C=0.0597, O=0.10700. The basis sets have been augmented by the polarization functions taken from the all-electron basis set cc-pVXZ (X=D,T,Q) of Dunning[58,59], up to g-functions. Despite the relatively diffuse nature of the Gaussian primitives, our robust implementation of Hartee-Fock exchange[60,61] allows for stable calculations in the condensed phase.[61,62] The Schwartz screening threshold for the HF calculations is in the range $10^{-8} - 10^{-10}$, periodic calculations used a truncated Coulomb operator[61], using approximately half the length of the smallest edge of the simulation cell as truncation radius. The threshold for the SCF convergence was 10^{-7} or tighter. The PW cutoff for the DFT calculations was $E_{cut} = 1200Ry$ to guarantee convergence, at small cost compared to the MP2 calculation. The MP2 calculation employed high quality PW cutoffs of $E_{cut} = 300$ Ry, $E_{cut}^{rel} = 50$ Ry, $\varepsilon_{filter} = 10^{-7}$, and $\varepsilon_{grid} = 10^{-6}$, unless mentioned otherwise. Gas phase systems have been computed using cluster boundary conditions for solving the Poisson equation[53].

Geometries

For all crystals, supercells have been generated by replicating the unit cell, so that the smallest edge was larger than 9Å, in order for the Γ-point approximation to be reasonable. The geometries of the NH_3 and CO_2 crystals have been built from the experimental lattice parameter and the space group ($a = 5.048$ Å , $P2_13$, for NH_3; $a = 5.55$ Å , Pa-3, for CO_2) as detailed in Ref.[63]. The experimental geometries of the other molecular crystals have been retrieved from the Cambridge Structural Database (CSD)[64]. These positions of the hydrogen atoms of these geometries have been

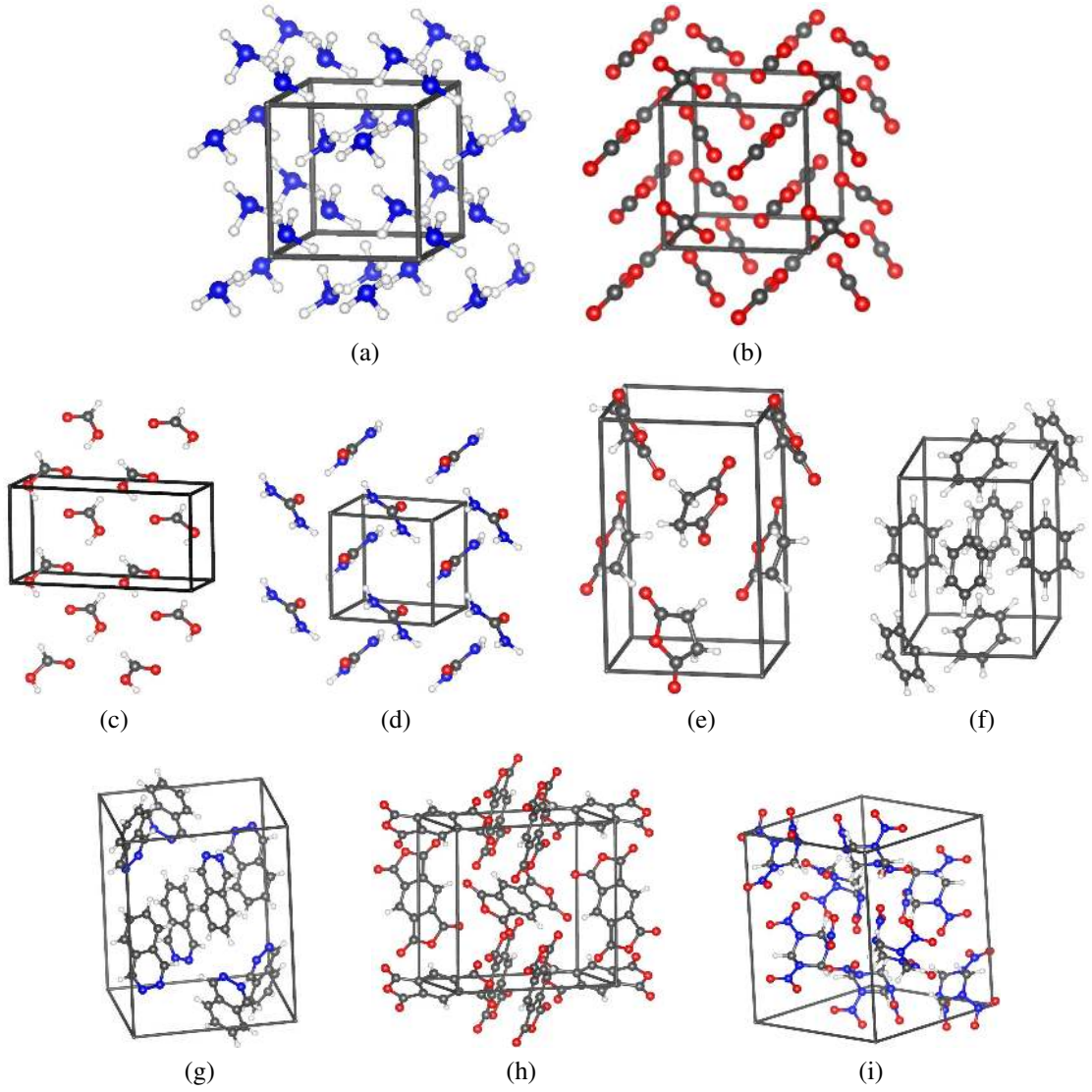


Figure 3: Unit cells of the molecular crystals under study. For all crystals $\alpha = \beta = \gamma = 90^\circ$, all lattice parameters expressed in Å. (a) NH_3, $a = b = c = 5.048$, supercell $2 \times 2 \times 2$, $N_{mol} = 32$ (b) CO_2, $a = b = c = 5.55$, supercell $2 \times 2 \times 2$, $N_{mol} = 32$ (c) Formic Acid, $a = 10.241$ $b = 3.544$ $c = 5.356$, supercell $1 \times 3 \times 2$ (refcode FORMAC01), $N_{mol} = 24$ (d) Urea, $a = b = 5.645$ $c = 4.704$, supercell $2 \times 2 \times 2$ (refcode UREAXX09), $N_{mol} = 16$ (e) Succinic Anhydride, $a = 5.4257$ $b = 6.9746$ $c = 11.7167$, supercell $2 \times 2 \times 1$ (refcode SUCANH15), $N_{mol} = 16$ (f) Benzene, $a = 7.398$ $b = 9.435$ $c = 6.778$, supercell $2 \times 1 \times 2$ (refcode BENZEN07), $N_{mol} = 16$ (g) 2,3-Diazanaphthalene, $a = 13.695$ $b = 10.557$ $c = 9.285$, supercell $1 \times 1 \times 1$ (refcode DAZNAP), $N_{mol} = 8$ (h) Pyromellitic Dianhydride, $a = b = 10.792$ $c = 7.4128$, supercell $1 \times 1 \times 2$ (refcode PYMDAN), $N_{mol} = 8$ (i) Trinitro-Triazacyclohexane, $a = 13.182$ $b = 11.574$ $c = 10.709$, supercell $1 \times 1 \times 1$ (refcode CTMTNA), $N_{mol} = 8$

further relaxed at the DFT/BLYP[65,66] level employing the cc-TZVP basis set. The main features of the structure of each crystal, together with the supercell used in the calculation and the CSD refcode, are reported in the caption of Figure 3.

Cohesive Energies and Lattice Parameter Optimization

The counterpoise (CP) corrected cohesive energy per molecule at a given volume V has been computed as

$$E_{coh}^{CP}(V) = \frac{E_{supercell}(V)}{N_{mol}} - E_{mol}^{gas} - E_{mol+ghost}^{crystal}(V) + E_{mol}^{crystal}(V). \quad (10)$$

Here, N_{mol} is the number of molecules per supercell, $E_{supercell}(V)$ the total energy of the supercell, and $E_{mol+ghost}^{crystal}(V)$, $E_{mol}^{crystal}(V)$, and E_{mol}^{gas} the total energy of an isolated molecule in either the crystal geometry ($E_{mol+ghost}^{crystal}(V)$ and $E_{mol}^{crystal}(V)$) or a gas phase geometry (E_{mol}^{gas}). $E_{mol+ghost}^{crystal}(V)$, includes ghost atoms from the 12 nearest neighbor molecules in the case of NH_3 and CO_2, while only the first coordination shell has been retained for the other crystals. This procedure has also been followed in Ref.[63] and Ref.[67]. The gas phase geometry employed for NH_3 and CO_2 corresponds to XXXXXX, consistent with Ref.[63]. The other molecular crystals have been computed twice, once with gas phase geometries relaxed at the B3LYP/cc-pTZVP level[66,68,69], and once, for direct comparison with Ref.[67], using the crystal geometry also for the gas phase geometry.

Lattice parameter optimization has been carried out for the NH_3 and CO_2 crystals. The employed procedure is approximate, as MP2 gradients and stresses are currently not available in CP2K, but is similar to the procedure in Ref.[63]. First, structures have been relaxed at a DFT/B3LYP level with the cc-TZVP basis set for various values of the lattice parameter. Next, $E_{coh}^{CP}(V)$ has been computed for each of these geometries. Finally, these results have been fitted with a third order Birch-Murnaghan equation of state in order to get the equilibrium cohesive energy and volume.

To assess the accuracy of computed cohesive energies, these values will be compared to the experimental sublimation enthalpies ($\Delta H(s)$). However, it has to be emphasized that this comparison includes theoretical bias and is subject to experimental error. Indeed, for non-volatile compounds, the sublimation enthalpies can be hard to measure, and can be subject to several kJ/mol error.

Experimental sublimation enthalpies usually are obtained at high temperature, while the cohesive energy is a zero temperature property. Only in a few cases can experimental sublimation enthalpies be extrapolated down to 0K, CO_2 being such an example[70]: 26.8kJ/mol at 0K, 25.2kJ/mol at 195K. Even at 0K, zero point energy differences should be taken into account, and in fact the anharmonicity of molecular crystals implies that also lattice parameters might need to be corrected for quantum effects to be truly comparable with experiment. Such corrections require the calculation of vibrational or phonon properties, see for example Ref.[71] for an early example based on a force field description of molecular crystals, and this has not been attempted in this work.

4.2 GPW-MP2 Accuracy

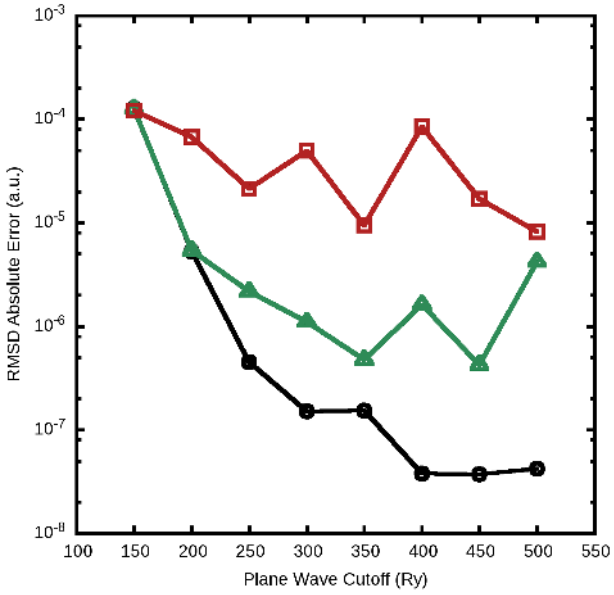
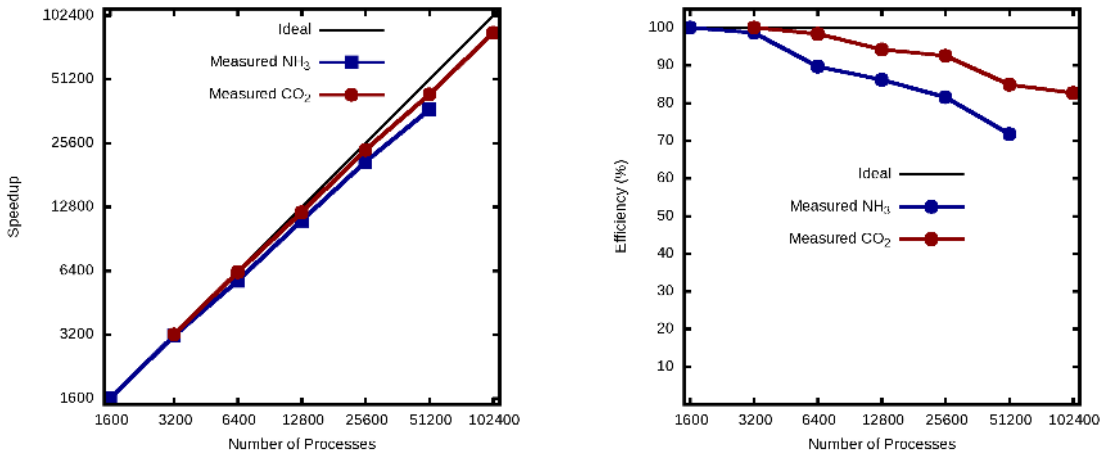


Figure 4: Root-mean-square deviation (RMSD) of the absolute error for the GPW-MP2 energy as a function of the plane wave cutoff (E_{cut}) for different values of the relative cutoff (E_{cut}^{rel}). Red Squares $E_{cut}^{rel} = 30$ Ry, Green Triangles $E_{cut}^{rel} = 40$ Ry and Black Circles $E_{cut}^{rel} = 50$ Ry (XXXXX adjust abbreviation in the figure, remove line for $E_{cut}^{rel} = 60$ XXX). The non-monotonic behavior of the curves is the result of fortuitous cancellation of errors.

In order to judge the impact of the PW cutoff (E_{cut}), and the multi-grid relative cutoff (E_{cut}^{rel}) on the accuracy of the GPW-MP2 energy, calculations with various values for these parameters have been performed and summarized in Figure 4. The benchmark set is based on 10 molecules

(from Figure 3 and H_2O) in the gas phase with the cc-TZVP basis, for which reference energies have been obtained using a traditional direct MP2 algorithm based on analytic four center integrals over atomic orbitals. The root-mean-square deviation (RMSD) of the absolute difference between the GPW-MP2 energy and the traditional MP2 implementation is used as a measure of the error. Tight values $\varepsilon_{filter} = 10^{-12}$ and $\varepsilon_{grid} = 10^{-12}$ together with a 17Å cubic cell and cluster boundary conditions have been used, to guarantee convergence with respect to these parameters. It can be observed that the GPW-MP2 energy converges rapidly with respect to both parameters, in particular if one realizes that the time for the integration in the MP2-GPW algorithm grows slowly as $E_{cut}^{\frac{3}{2}}$. Not unexpectedly, both parameters need to be increased simultaneously in order to obtain accurate results. The combinations 300/50Ry, 250/40Ry, and 200/30Ry yield errors of approximately 10^{-6}, 10^{-5}, and 10^{-4} Hartree respectively. Note again that E_{cut} depends on the largest exponent in the basis set used, while E_{cut}^{rel} is system independent.

4.3 Performance of the Parallel Algorithm



(a) (b)

Figure 5: Measured speed up and efficiency for the calculation of the GPW-MP2 energy. (a) Speed up measured with respect to 1600 processes for NH_3 crystal and with respect to 3200 processes for CO_2 crystal. (b) Efficiency measured with respect to 1600 processes for NH_3 crystal and with respect to 3200 processes for CO_2 crystal. (XXXX should we have just 2 panels, i.e. speedup and efficiency, and two curves per panel ? XXXX)

To assess the performance of the parallel algorithm, test calculations on molecular crystals of NH_3 and CO_2 have been performed. For NH_3 a cc-TZVP quality basis set has been chosen, while CO_2 has been described with a cc-QZVP basis. Each supercell contains 32 molecules, resulting in 2272 and 5184 basis functions respectively The speed up and the parallel efficiency for the two test cases are reported in Figure 5. For both benchmarks the algorithm shows very good parallel scalability in a wide range. In particular in the case of CO_2, the efficiency remains higher than 80% even for the 102400 processes run. For the NH_3 calculation, the number of *ia* pairs (274560) becomes similar to the number of processes, making an even distribution of the pairs more difficult. Additionally, the overhead of initializing grids and matrices becomes non-negligible leading to a efficiency of 70%. At full scale-out, the MP2 calculation required a wall-time of 74 and 518 seconds respectively.

4.4 System Size Scaling

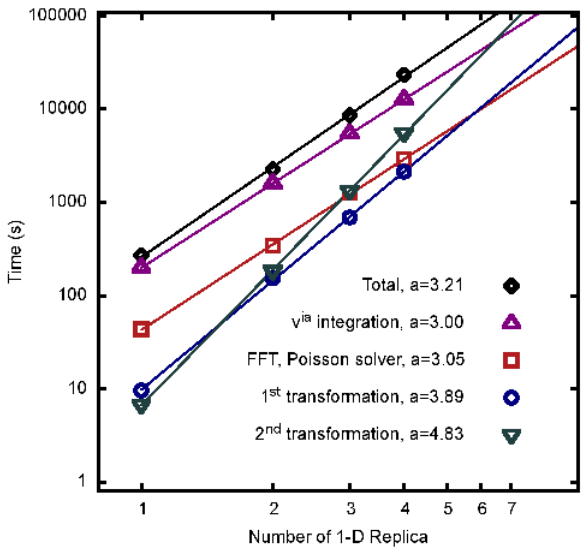


Figure 6: Shown is the time spent in the various significant parts of the GPW-MP2 energy calculation, as a function of the number of replicas of the supercell, containing 32, 64, 96 and 128 molecules of NH3 respectively. Lines represent a linear two-parameter fit of the form $y = bx^a$. The values of a for each task are reported in the legend.

In order to validate the expected timings for the important parts of the GPW-MP2 algorithm shown in Table 1, we have performed calculations on systems of increasing size. The test system is the supercell of NH_3, containing 32 molecules, with a cc-DZVP basis. This supercell has been replicated in one dimension, yielding an orthorhombic supercell containing up to 128 molecules. The timings, as obtained from runs on 1200 cores, are shown in Figure 6, together with a fit of the form $y = bx^a$. The measured exponents agree very well with the expected values as reported in Table 1, being very close to 3 for the integration and the Poisson solver, 4 for the first and 5 for the last index transformation. In addition to giving information about the scaling behavior, insight about the prefactor can be inferred from the graph. In particular, the last index transformation has a very small prefactor and an extrapolation suggest that it will only dominate for systems containing more than 200 molecules. Note that in an RI-MP2 calculation, this term has a larger prefactor, approximately given by the ratio of basis function in the auxiliary and primary basis. For all system sizes tested, the cost is currently dominated by the integration routine, yielding an apparent overall scaling exponent of the GPW-MP2 algorithm of 3.21.

4.5 NH_3 and CO_2 Molecular Crystals

The molecular crystals of NH_3 and CO_2 present two useful benchmark systems, as they differ in the nature of their interaction: NH_3 is hydrogen bonded, while CO_2 is not. Furthermore, results for these systems can be compared to results presented in Ref.[63] and obtained by LMP2 as implemented in CRYSCOR. The effect of the basis set and the supercell size on the cohesive energy at the experimental geometry are reported in Table 2. Not unexpectedly, it can be seen that cc-DZVP yields poor results for the cohesive energy, and a cc-TZVP basis is required to yield a cohesive energy within approx. 5 kJ/mol of the basis set extrapolated result. However, the smaller cc-DZVP basis allows for larger systems, and hence can be used to study the size dependence of the result. Indeed, since our results are obtained at the Γ-point only, the size of the supercell matters. Fortunately, we see that the difference between the cohesive energies obtained 2x2x2 and 3x3x3 unit cells (32 and 108 molecules respectively) is rather small, a few kJ/mol, smaller than the difference

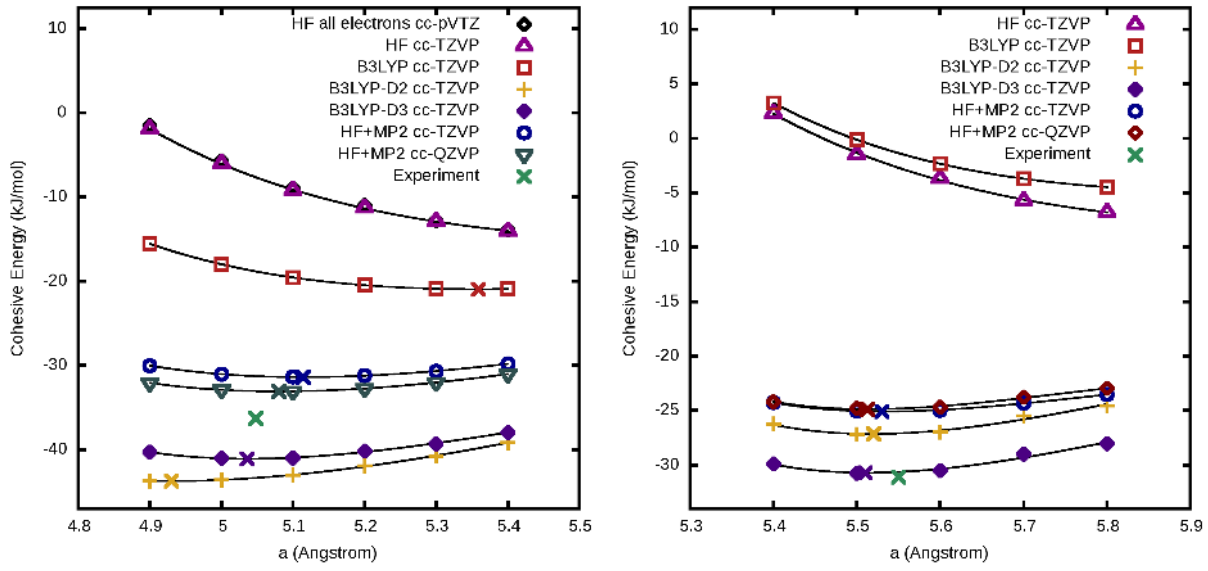
Table 2: Counterpoise corrected cohesive energy (E_{coh}^{CP}) in kJ/mol for the NH_3 and CO_2 crystals. Results have been calculated employing different basis sets and system sizes at the experimental geometry. Values in parenthesis refer to the HF+LMP2 results from Ref.[63] as obtained with the CRYSCOR program, using basis sets of comparable quality.

	NH_3		CO_2	
	HF	HF+MP2	HF	HF+MP2
cc-DZVP	-8.43	-25.78 (-22.6)	-7.48	-8.92 (-8.9)
cc-DZVP ($3 \times 3 \times 3$) [a]	-8.80	-27.89	-7.47	-11.46
cc-TZVP	-5.98	-30.93 (-29.6)	-5.86	-20.95 (-19.3)
cc-QZVP	-5.60	-32.76 (-32.3)	-5.91	-23.28 (-24.4)
Extrapolated [b]	-5.52	-33.93	-5.99	-26.09
Experiment [c]		-36.3		-31.1

[a] Calculation performed with supercell $3 \times 3 \times 3$ instead of $2 \times 2 \times 2$.

[b] (T-Q) extrapolation toward the basis set limit[72].

[c] Values from Ref.[63], see also Ref.[73] for NH_3 and Ref.[74,75] for CO_2. (XXXX check this, or provide references to the original literature XXXX).



(a) (b)

Figure 7: Lattice parameter optimization curves for NH_3 (a) and CO_2 (b), computed at different level of theory and different basis sets. The crosses represent the location of the minimum point for each curve.

Table 3: Equilibrium counterpoise corrected cohesive energy (E_{coh}^{CP} in kJ/mol per molecule) and lattice parameter (a in Å) for the NH_3 and CO_2 crystals calculated employing different methods. The basis set is cc-TZVP, except when specified otherwise.

	NH_3		CO_2	
	a	E_{coh}	a	E_{coh}
HF	5.98	-15.7	6.05	-5.00
B3LYP	5.36	-20.9	6.20	-8.20
B3LYP-D2	4.93	-43.7	5.52	-27.2
B3LYP-D3	5.04	-41.1	5.51	-30.7
HF+MP2	5.11	-31.4	5.53	-25.1
HF+MP2 [a]	5.08	-33.1	5.51	-24.9
Exp.	5.048	-36.3	5.55	-31.1

[a] Calculated with cc-QZVP basis set.

between cc-TZVP and basis extrapolated results. Note that the HF results are less sensitive to both system size and basis set effects. Of course, the Hartree-Fock level of theory is a poor description for these systems, since the MP2 correlation contributes almost 80% of the cohesive energy. Careful estimates of the MP2 cohesive energy could combine the size effect as computed with smaller basis sets with basis set extrapolated results for smaller supercells. Furthermore, our results are in good agreement with those reported in Ref.[63], despite the various differences in methodology, such as the use of pseudopotentials, corresponding basis sets, local MP2, etc.

In Figure 7, lattice parameter optimization curves are reported as obtained at various levels of theory. In the case of NH_3, the quality of the pseudopotential approximation has been verified as illustrated by the excellent agreement between the HF results as obtained with pseudopotentials and the corresponding cc-TVZP basis, and an all-electron calculation employing the Dunning cc-pVTZ basis. For both NH_3 and CO_2, the DFT/B3LYP and HF results predict an equilibrium lattice parameter much larger than the experimental one, and a cohesive energy that is far from the experimental one. The HF results are significantly improved by the MP2 correction, and the same can be said for the dispersion corrections in the case of B3LYP. The calculated equilibrium values of a and E_{coh} are summarized in Table 3. The MP2 lattice constants are within approx. 1% of the

experimental results, and while the same holds for B3LYP-D3[76], a larger error is observed with B3LYP-D2[77]. The cc-TZVP and cc-QZVP MP2 results are very similar, suggesting that the former might be a cost effective choice for these calculations.

4.6 Cohesive Energy of Molecular Crystals from MP2 and double hybrid DFT

Finally, cohesive energies for the remaining molecular crystals reported in Figure 3 are presented in Table 4, in all cases using the experimental crystal cell, and employing the cc-TZVP basis. Here, a wider set of theoretical methods is employed, including MP2 as well as double hybrid functionals. These values are compared to the experimentally measured sublimation enthalpies, with the caveats mentioned in Figure 3. The RMSD, measured over this relatively small test set, is reported as well. First, HF and DFT/B3LYP methods display the largest RMSD, demonstrating that these methods are not suitable for computing, even to a qualitative level, the lattice energy of molecular crystals. Small molecules that interact mostly via hydrogen-bonds and dipole-dipole electrostatics are described best, but van der Waals dominated complexes can have even a negative computed lattice energy. On the other hand, the D2 and D3 dispersion corrections appreciably improve the bare B3LYP results, for all cases the computed lattice energy goes closer to the experimental value, decreasing significantly the RMSD. However, both B3LYP-D2 and B3LYP-D3 tend to overbind the crystals, B3LYP-D3 being slightly worse. With an RMSD of 16.9 kJ/mol, the lattice energies calculated at the MP2 level agree outperform the DFT results. The molecules that display the largest deviations, Benzene and Pyromellitic Dianhydride, have a large π electronic delocalization, which is known to be unfavorable for the performance of MP2. Relaxation of the gas phase geometry has a large effect on the cohesive energies for those systems with strong interactions in the crystal, and is only negligible for benzene. Also in this case, there good agreement between the computed GPW-MP2 energies (non relaxed case) and the LMP2 calculations reported in Ref.[67].

The spin component scaled variants of the MP2 method tend to improve slightly over the performance of standard MP2, the MP2-SCF method of Grimme[78] yielding the smallest RMSD.

Table 4: Counterpoise corrected cohesive energy ($-E_{coh}^{CP}$) in kJ/mol for the the molecular crystals of B = Benzene (fig Figure 3f), FA = Formic Acid (fig Figure 3c), SA = Succinic Anhydride (fig Figure 3e), D = 2,3-Diazanaphthalene (fig Figure 3g), PD = Pyromellitic Dianhydride (fig Figure 3h), U = Urea (fig Figure 3d), CT = Cyclotrimethylene-Trinitramine (fig Figure 3i). The sign of the E_{coh}^{CP} has been changed in order to be compared with the experimental sublimation enthalpies $\Delta H(s)$. (C) means that the gas phase geometry has not been relaxed but was constrained to the crystal geometry.

	B	FA	SA	D	PD	U	CT	RMSD
Exp $\Delta H(s)$ [a]	45	68	81	83	83	92	112	
LMP2 (C) [b]	56.6	63.2	87.0			108.6		
GPW-MP2 (C)	58.7	64.9	84.9	93.3	127.3	106.6	126.6	19.7
GPW-MP2	58.8	55.5	81.2	79.7	123.4	94.6	113.7	16.9
HF	-21.2	26.3	38.6	-5.7	31.3	55.8	49.8	58.1
B3LYP	-12.5	34.1	27.5	-7.0	15.8	64.2	17.3	65.1
B3LYP-D2	56.9	69.1	91.8	84.0	122.0	110.5	128.4	18.5
B3LYP-D3	60.5	71.9	94.7	87.3	128.1	111.6	131.0	21.5
B2PLYP [c]	14.7	44.8	51.0	74.2	60.0	78.4	58.8	29.2
B2PLYP-D3 [c]	53.4	64.7	86.4	77.3	119.8	103.7	119.9	15.6
DSD-BLYP [c]	31.6	54.4	67.6	50.6	88.8	90.9	90.8	17.2
DSD-BLYP-D3 [c]	56.9	67.4	90.8	82.8	128.1	107.4	130.9	20.2
MP2-SCS [d]	38.6	52.2	75.6	62.3	106.4	89.2	112.0	13.6
MP2-SOS [e]	28.5	50.6	72.9	53.6	97.9	86.7	111.2	15.8
MP2-SCS(MI) [f]	47.9	47.3	70.1	63.9	104.6	84.8	90.2	16.5
MP2-SCSN [g]	52.3	44.9	67.1	64.3	103.4	82.3	79.3	19.7

[a] Taken from the supporting information of Ref [67], see also `http://webbook.nist.gov/chemistry/`.

[b] Local MP2 calculation performed with the CRYSCOR program reported from Ref [67].

[c] The parameter for these double hybrid functional and their relative D3 corrections have been taken from Ref [4].

[d] Spin Component Scaled (SCS) [78] ($p_S = 1.2$, $p_T = 0.333$).

[e] Scaled Opposite Spin (SOS) [79] ($p_S = 1.3$, $p_T = 0$).

[f] Spin Component Scaled (Molecular Interaction) (SCS(MI)) [80] ($p_S = 0.4$, $p_T = 1.29$).

[g] Spin Component Scaled for Nucleobases (SCSN) [81] ($p_S = 0$, $p_T = 1.76$).

This is largely due to the better agreement observed for Pyromellitic Dianhydride, for which the discrepancy is reduced from ∼40 to ∼20 kJ/mol. Interestingly, the MP2-SCS(MI) of Distasio *et al.*[80], which has been parameterized explicitly for molecular interactions using the S22 database[82] as training set, does not perform best for the molecular interactions in these crystals. It is the best method for benzene, which is also present in the S22 database in a similar configuration as the one found in the crystal, but it is less accurate for formic acid, which is present in S22 in a fairly different geometry. This suggests that a database of accurate cohesive energies for molecular crystals would complement the S22 set by providing a wider range of molecular geometries, and thus provide valuable input for the development of improved methods.

Two different double hybrid functionals have been tested: B2PLYP and DSD-BLYP with and without dispersion corrections.[4] The main difference between these two functionals is that the MP2-like term is scaled with a single parameter in B2PLYP, while independent parameters are used for the same spin and opposite spin MP2-like terms in DSD-BLYP, as is done in MP2-SCS. B2PLYP without dispersion correction shows a large RMSD (XXXX please add to the table, for relaxed gas geo XXXXX). B2PLYP-D3, with an added D3 correction, improves slightly over MP2. Finally, the DSD-BLYP functional yields the best results for Pyromellitic Dianhydride, which is the source of large errors for other methods. Surprisingly, adding the D3 correction actually increases the RMSD, increasing the errors for some compounds significantly.

5 Conclusions

In the present work a novel method for the calculation of the canonical MP2 energy of finite and extended systems has been presented. The crucial aspect of the method is that half-transformed electron repulsion integrals (ERIs) $(ia|\lambda\sigma)$ are directly calculated. This is possible using a mixed Gaussian and Plane Wave approach, which allows for computing the electrostatic potential v^{ia} of the occupied-virtual pair density ρ^{ia} in an auxiliary basis, and numerically integrating over products of basis functions $\lambda\sigma$. The method is naturally suited and robust for periodic systems and the

numerical accuracy can be easily controlled. The corresponding algorithm shows excellent parallel performance up to 100000 processes, and allows for MP2 calculations of systems containing hundreds of atoms and thousands of basis functions in minutes. Benchmark calculations on molecular crystals have been performed to validate the GPW-MP2 method, and good agreement with literature results and, for most benchmarks, with experiment is obtained. These calculations also suggest that a database with reliable reference cohesive energies for molecular crystals could complement existing gas phase databases, and contribute to the development of improved methods and functionals for weak interactions. We believe that the GPW-MP2 method can now be used to study condensed phase systems with a few hundred atoms per unit cell, including not only crystals but also systems without symmetry such as molecules on surfaces and liquids. Advanced techniques, such as RI-MP2 and local MP2, and adaptation to new hardware, such as accelerators, are likely to further improve upon the method presented here.

6 Acknowledgements

We thank Ruyman Reyes Castro and Iain Bethune from the EPCC, UK, supported by PRACE project PA0723, for application analysis and optimization. JV acknowledges financial support by the European Union FP7 in the form of an ERC Starting Grant under contract no. 277910. Calculations were enabled by 2011 INCITE awards on the CRAY XK6 using resources of the National Center for Computational Sciences at Oak Ridge National Laboratory (ORNL), which is supported by the Office of Science of the U.S. DOE under Contract No. DE-AC05-00OR22725, and by the Swiss National Supercomputer Centre (CSCS). The research leading to these results has received funding from the Swiss University Conference through the High Performance and High Productivity Computing (HP2C) Programme.

XXXXXXXXXXXX the references need careful checking, some of the special charaters are messed up, short journal names are needed, doi should not appear XXXXXXXXXXXX

References

(1) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

(2) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; McGraw Hill: New York, 1982.

(3) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.

(4) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.

(5) Cremer, D. *WIREs Comput. Mol. Sci.* **2011**, *1*, 509–530.

(6) Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.

(7) Pulay, P.; Saebø, S. *Theor. Chim. Acta* **1986**, *69*, 357–368.

(8) Rauhut, G.; Pulay, P.; Werner, H.-J. *J. Comput. Chem.* **1998**, *19*, 1241–1254.

(9) Schütz, M.; Hetzer, G.; Werner, H.-J. *J. Chem. Phys.* **1999**, *111*, 5691–5705.

(10) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H.-J. *J. Chem. Phys.* **2000**, *113*, 9443–9455.

(11) Saebø, S.; Pulay, P. *J. Chem. Phys.* **2001**, *115*, 3975–3983.

(12) Pisani, C.; Busso, M.; Capecchi, G.; Casassa, S.; Dovesi, R.; Maschio, L.; Zicovich-Wilson, C.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 094113.

(13) Pisani, C.; Maschio, L.; Casassa, S.; Halo, M.; Schütz, M.; Usvyat, D. *J. Comput. Chem.* **2008**, *29*, 2113–2124.

(14) Maslen, P. *Chem. Phys. Lett.* **1998**, *283*, 102–108.

(15) Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 7093–7099.

(16) Almlöf, J. *Chem. Phys. Lett.* **1991**, *181*, 319–320.

(17) Häser, M.; Almlöf, J. *J. Chem. Phys.* **1992**, *96*, 489–494.

(18) Häser, M. *Theor. Chim. Acta* **1993**, *87*, 147–173.

(19) Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 3660–3671.

(20) Lambrecht, D. S.; Doser, B.; Ochsenfeld, C. *J. Chem. Phys.* **2005**, *123*, 184102.

(21) Doser, B.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2009**, *130*, 064107.

(22) Ayala, P. Y.; Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2001**, *115*, 9698–9707.

(23) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

(24) Weigend, F. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(25) Bernholdt, D. E.; Harrison, R. J. *J. Chem. Phys.* **1998**, *109*, 1593–1600.

(26) Werner, H.-J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149–8160.

(27) Izmaylov, A. F.; Scuseria, G. E. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3421.

(28) Katouda, M.; Nagase, S. *J. Chem. Phys.* **2010**, *133*, 184103.

(29) Klopper, W.; Manby, F. R.; Ten-No, S.; Valeev, E. F. *Int. Rev. Phys. Chem.* **2006**, *25*, 427–468.

(30) Limaye, A. C.; Gadre, S. R. *J. Chem. Phys.* **1994**, *100*, 1303–1307.

(31) Marquez, A. M.; Dupuis, M. *J. Comput. Chem.* **1995**, *16*, 395–404.

(32) Nielsen, I. M. B.; Seidl, E. T. *J. Comput. Chem.* **1995**, *16*, 1301–1313.

(33) Baker, J.; Pulay, P. *J. Comput. Chem.* **2002**, *23*, 1150–1156.

(34) Ishimura, K.; Pulay, P.; Nagase, S. *J. Comput. Chem.* **2006**, *27*, 407–413.

(35) Katouda, M.; Nagase, S. *Int. J. Quantum Chem.* **2009**, *109*, 2121–2130.

(36) Doser, B.; Lambrecht, D. S.; Ochsenfeld, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3335.

(37) Valeev, E. F.; Janssen, C. L. *J. Chem. Phys.* **2004**, *121*, 1214–1227.

(38) Nakao, Y.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 6375–6380.

(39) Nielsen, I. M. B.; Janssen, C. L. *J. Chem. Theory Comput.* **2007**, *3*, 71–79.

(40) Maschio, L. *J. Chem. Theory Comput.* **2011**, *7*, 2818–2830.

(41) Sun, J.-Q.; Bartlett, R. J. *J. Chem. Phys.* **1996**, *104*, 8553–8565.

(42) Marsman, M.; Gruneis, A.; Paier, J.; Kresse, G. *J. Chem. Phys.* **2009**, *130*, 184103.

(43) Gruneis, A.; Marsman, M.; Kresse, G. *J. Chem. Phys.* **2010**, *133*, 074107.

(44) Lippert, G.; Hutter, J.; Parinello, M. *Mol. Phys.* **1997**, *92*, 477–488.

(45) Martinez, T. J.; Carter, E. A. *J. Chem. Phys.* **1994**, *100*, 3631–3638.

(46) Ishimura, K.; Ten-no, S. *Theor. Chem. Acc.* **2011**, *130*, 317–321.

(47) Hirata, S.; Iwata, S. *J. Chem. Phys.* **1998**, *109*, 4147–4155.

(48) The CP2K developers group, CP2K is freely available from: http://www.cp2k.org/, 2012.

(49) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103 – 128.

(50) Aissing, G.; Monkhorst, H. J. *Int. J. Quantum Chem.* **1993**, *48*, 81–89.

(51) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810–2821.

(52) Blochl, P. E. *J. Chem. Phys.* **1995**, *103*, 7422–7428.

(53) Genovese, L.; Deutsch, T.; Neelov, A.; Goedecker, S.; Beylkin, G. *J. Chem. Phys.* **2006**, *125*, 074105.

(54) Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124–140.

(55) Krack, M.; Parrinello, M. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2105–2112.

(56) MPI, Message Passing Interface. `http://www.mpi-forum.org/`.

(57) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(58) Thom H. Dunning, J. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(59) Woon, D. E.; Thom H. Dunning, J. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(60) Guidon, M.; Schiffmann, F.; Hutter, J.; VandeVondele, J. *J. Chem. Phys.* **2008**, *128*, 214104.

(61) Guidon, M.; Hutter, J.; VandeVondele, J. *J. Chem. Theory Comput.* **2009**, *5*, 3010–3021.

(62) Paier, J.; Diaconu, C. V.; Scuseria, G. E.; Guidon, M.; VandeVondele, J.; Hutter, J. *Phys. Rev. B* **2009**, *80*, 174114.

(63) Maschio, L.; Usvyat, D.; Schütz, M.; Civalleri, B. *J. Chem. Phys.* **2010**, *132*, 134706.

(64) Allen, F. H. *Acta Crystallogr., Sect. B: Struct. Sci* **2002**, *58*, 380–388.

(65) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(66) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(67) Maschio, L.; Civalleri, B.; Ugliengo, P.; Gavezzotti, A. *J. Phys. Chem. A* **2011**, *115*, 11179–11186.

(68) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(69) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(70) Trusler, J. P. M. *J. Phys. Chem. Ref. Data* **2011**, *40*, 043105.

(71) Warshel, A.; Lifson, S. *J. Chem. Phys.* **1970**, *53*, 582–594.

(72) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243 – 252.

(73) Shipman, L. L.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1976**, *80*, 52–54.

(74) Keesom, W.; Kohler, J. *Physica* **1934**, *1*, 655 – 658.

(75) Curzon, A. *Physica* **1972**, *59*, 733.

(76) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(77) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(78) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(79) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793–9802.

(80) Distasio JR., R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073–1083.

(81) Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80–85.

(82) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.