

# eCSE10-10: Enhancing long-range dispersion interaction functionality of CASTEP

Peter J. P. Byrne<sup>1</sup>, Dominik B. Jochym<sup>2</sup>, Dawn L. Geatches<sup>3</sup>, Matt I. J. Probert<sup>1</sup>, Phil J. Hasnip<sup>1</sup>, Ian Rosbottom<sup>4</sup>, and Anthony M. Reilly<sup>5</sup>

<sup>1</sup>University of York, Department of Physics, York, YO10 5DD, UK

<sup>2</sup>Rutherford Appleton Laboratories, Chilton, Didcot, Oxfordshire, OX11 0QX, UK

<sup>3</sup>Department of Scientific Computing, STFC Daresbury Laboratory, Warrington, WA4 4AD, UK

<sup>4</sup>Centre for the Digital Design of Drug Products, School of Chemical and Process Engineering,  
University of Leeds, UK

<sup>5</sup>School of Chemical Sciences, Dublin City University, Glasnevin, Dublin 9, Ireland

August 1, 2018

## 1 Introduction

CASTEP[1] is a parallelised plane-wave density-functional-theory(DFT) code for first principles materials modelling. Although in DFT, the kinetic and electrostatic(Hartree) terms have simple forms in terms of either the density or single Kohn-Sham orbitals, the exchange and correlation (xc) terms are more difficult to calculate but in theory, they should be functionals of the electron density.

Van der Waals interactions are inherently long range correlations between atoms. They cause a generally attractive force when spontaneous dipoles form around atoms and cause a sympathetic response in nearby molecules. Although these interactions are complicated, it can be shown that the leading order term is of order  $r^{-6}$ , with higher order even terms, where  $r$  is the distance between the atoms.

The simplest xc functional, the local density approximation(LDA) is a function of just the electron density and is entirely local. Unfortunately, as the density away from a molecular cluster drops to zero exponentially, the xc energy also follows the same pattern which means that local xc functionals cannot capture long-ranged Van der Waals interactions(VdW). On the other hand, the short range correlations are reasonably well represented by local xc functionals.

Attempts can be made to improve upon the LDA by including first and 2nd order derivatives of the density to make them semi-local, but these tend to have numerical stability issues in regions of very low density and again fail to accurately describe VdW interactions.

Although there is an exact form for the exchange energy in terms of the Kohn-Sham orbitals, the effect required is a correlative one and solving for the exact correlation

would be as hard as solving the many-body Schrodinger equation and would dwarf the effort in solving the DFT equations.

In practice, an atomic-site based interaction is added as a post-hoc correction. The simplest schemes use pairwise interactions(e.g. D2[2], TS[3], OBS[4], JCHS[5]), but a few schemes have attempted higher order corrections. In the last few years, the many-body dispersion method[6] has been developed that includes many-order interactions to improve the modelling of dispersion in condensed systems.

## 2 Project Goals

In the eCSE application, there were several main goals presented which will be used as the main section headings as they cover a range of improvements in different areas.

The initial goal was to rewrite the old SEDC module[7] which relied heavily on the use of module level variables and was quite difficult to follow. We combined this with the initial work from a previous eCSE project to implement the Ewald summation method and this was completed. This has already been merged with the main CASTEP codebase and will be a key feature in the next CASTEP release which is due in November.

The 2nd goal was to increase the modularity of the code and in general make it easier to implement new functionality. This was demonstrated by adding the D3/D4 library from Grimme *et al*[8]. The inclusion of this functionality is controlled by a compile flag in the CASTEP Makefile and can be activated in the academic release of CASTEP.

The 3rd goal was to include the effects of variations in the Hirshfeld derivatives in some of the schemes but the effect of these has been found to be small and instead the work focused on improvements to the efficiency of the pairwise schemes. This has been completed and merged into the main CASTEP codebase.

The final goals were centred around improving the efficiency of the many-body dispersion method and to implement analytic forces. Initially, we had thought to optimise the parallel matrix diagonalise, but a recent development of a reciprocal-space version of the MBD algorithm was formulated and this was implemented instead because the expected parallel gains were much larger. This has been integrated into the next CASTEP release.

## 3 Rewrite of the SEDC module

### 3.1 Introduction

In practice, one tends to add a corrective atom based interaction term of the form

$$E_{\text{vdW}} = \sum_{a,b} -f_{\text{damp}}(r_{ab}, r_{ab}^0) \frac{c_{ab}^6}{r_{ab}^6}, \quad (1)$$

where  $a$  and  $b$  are atomic indexes,  $c_{ab}^6$  and  $r_{ab}^0$  are parameterised interaction coefficients which depends on atomic species or site,  $r_{ab}$  is the distance from atom  $a$  to  $b$  and  $f_{\text{damp}}$  is a damping function which is 0 at short range and 1 at long range.

The purpose of the damping function is to only add the correction at long range and not correct the short range correlation which is already included in standard xc functionals. The exact range that determines what is a long range correlation is dependent on the correction scheme and the xc-functional in use and in general is determined by some sort of fitting procedure which is why these schemes are referred to as semi-empirical dispersion correction (SEDC).

CASTEP already had the ability to calculate SEDC with a choice of 4 schemes [2, 3, 4, 5], but it was quite limited. The code was not parallelised and relied heavily on the use of module level variables. The initial goal of the project was to rewrite this code in a more modern and maintainable fashion while adding parallelism.

It had previously been found that for some high pressure systems, the large numbers of atoms and the long range interactions made the calculation of the SEDC very slow. A previous eCSE had shown promising results in implementing an Ewald method to alleviate this issue but had not been completed and integrated into CASTEP. It was decided that rather than re-implement the direct real space summation, we would finish the Ewald method as the basis for the pairwise correction methods.

### 3.2 Implementation

The Ewald method gives us a form for calculating the summation of a periodic  $r^{-6}$  potential, but the formulation of the SEDC schemes has a damping function at short range, but we can rewrite this as,

$$E_{\text{VdW}} = \sum_{a,b} -c_{ab}^6 r_{ab}^{-6} + (1 - f_{\text{damp}}) c_{ab}^6 r_{ab}^{-6} \quad (2)$$

The first term can be evaluated directly using the Ewald summation, while the second term is evaluated directly in real space. In contrast to the previous formulation, this term is now short ranged and as the damping length scale tends to be on the order of 5–10 Å, it is very fast to evaluate.

The resulting form of the VdW energy is,

$$E_{\text{VdW}} = E_{\text{real}} + E_{\text{recip}} + E_{\text{self}} + E_{\text{damp}}, \quad (3)$$

where  $E_{\text{real}}$  is the real term excluding self interactions and  $E_{\text{recip}}$  is the reciprocal space interaction which includes an average self-interaction. The  $E_{\text{self}}$  term removes these spurious self interactions and the final term  $E_{\text{damp}}$  removes the short range interactions. These summations are also easy to parallelise as the results are simply summed and so it can be easily parallelised across either atom index and this parallelisation was implemented.

Derivatives of the interaction energy with respect to atomic positions gives the forces on each atom within the periodic system. These are calculated analytically and cost approximately  $3\times$  the computational effort of the SEDC energy calculation as the only contributions are from the real, recip and damping terms.

In the explicit real-space formulation, the stress can be found from the virial theorem for the forces. This can only be used when the explicit force between a pair of atoms,  $F_{ab}$  can be calculated. This is not possible from the Ewald summation. Instead the virial theorem is used for the real-part while the reciprocal contribution is calculated explicitly as it has a simple form.

### 3.3 Parallelisation

To reduce the duplication of code while also allowing parallelism, a general parallel converger was written. This takes a function which calculates the energy, force and stress contributions for a particular atom pair and performs the summations across both atomic indexes in parallel. The scheme synchronises once per loop to check convergence and then performs an MPIREDUCE at the end to get the total energy, forces and stresses. This means that the parallelisation is implemented for both real, reciprocal and damping functions in one routine and this can be tested for correctness.

### 3.4 Performance

The convergence of the components of the Ewald energy and the direct real-space method are shown in figure 1 for graphene. As can be seen they converge much faster in each case in the Ewald method. Similar effects are seen for the forces (fig. 2) and stress(fig. 3)

Material	DFT Time(s)	Old SEDC(s)	New-Ewald(s)	Parallel-Ewald(s)
Medium Organic	529.62	8.21	7.49	0.29
Large Organic	1130.84	47.16	40.99	1.38
HP Methane	8.06	113.99	3.93	0.12

Table 1: Comparison timings for some systems on Archer. Parallel-Ewald refers to our parallel implementation of the Ewald method. All calculations were performed on a single node with 24-way MPI parallelism. The medium and large organic molecules were 108 and 256 atom systems respectively. HP Methane is high pressure methane. The convergence criterion was that the absolute change in energy per atom from including an additional shell was less than  $1 \times 10^{-8}$  eV.

Timings for some realistic systems are shown in table 1. For both the organic systems, the DFT takes the majority of the computational time. The time to perform the Ewald summation in serial is faster by only a very small margin due to overheads in the Ewald method. When parallelised on 24 cores, these get approximately a  $24\times$  speed up which indicates that there is very little MPI overhead.

The problematic high-pressure methane system shows a marked improvement with the Ewald summation, and the SEDC time which was previously dominating the calculation is now smaller than the DFT time. With parallelisation, this is decreased even further again achieving near perfect parallelisation.

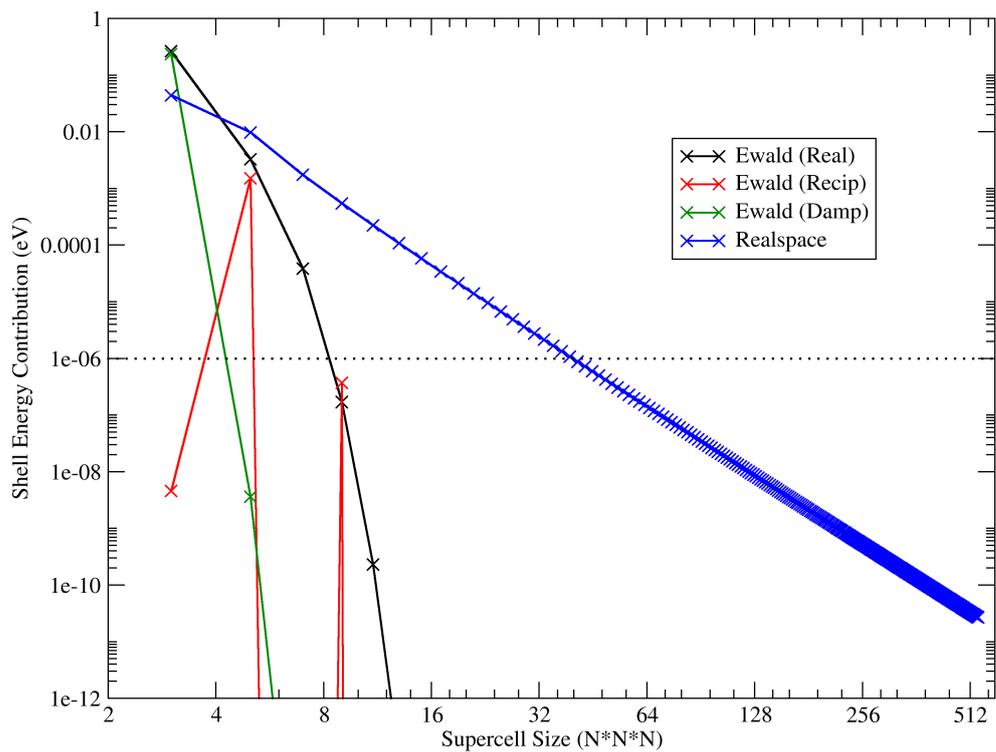


Figure 1: The change to the total energy by increasing the summation by a supercell in each direction for graphene. The dashed line indicates the default CASTEP convergence criterion for total energy.

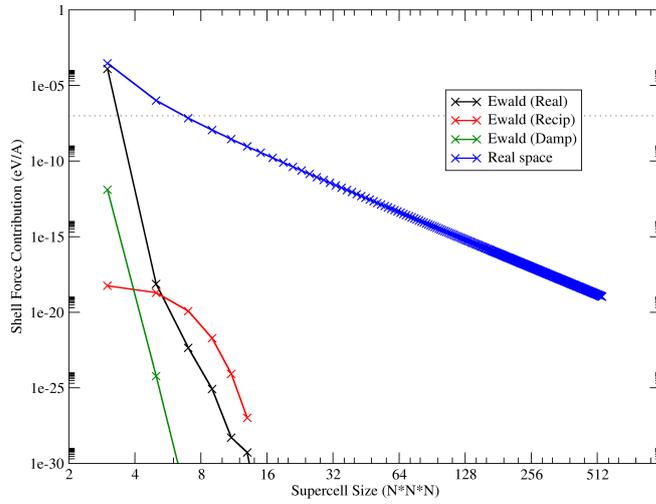


Figure 2: The change to the magnitude of the force on an atom by increasing the summation by a supercell in each direction for graphene. The dashed line indicates the default CASTEP convergence criterion for forces.

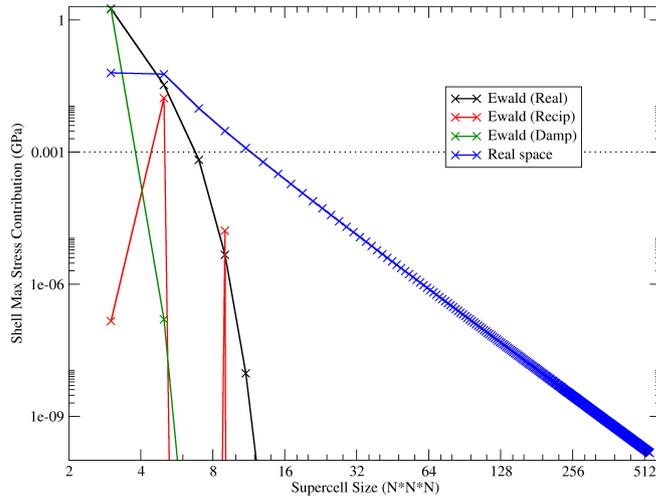


Figure 3: The change to the total energy by increasing the summation by a supercell in each direction for graphene. The dashed line indicates the default CASTEP convergence criterion for total stress.

## 4 D3/D4 implementation

The increase in modularity of the code was demonstrated by adding an interface to the D3/D4 library of Grimme *et al*[8]. This is a higher order SEDC method which includes 3-body terms as well as an  $r^{-8}$  interaction term. The scheme also takes the coordination number of an atom into account when determining the  $c^6$  coefficients. This increase in fitting coefficients increases the accuracy of the correction but also pushes it further from being *ab initio*.

As this is an optional module and the Grimme library is licenced under the GPL, it has to be optionally linked to CASTEP. Two new wrapper modules were added, `dftd_d3.f90` and `dftd_d3_dummy.f90`, which both have the same interface. The dummy module just raises an error when called whereas the real module calls the corresponding library interface. This avoids the use of optional compile statements in high level modules. The choice of which of these libraries to use is controlled by a flag in the Makefile.

## 5 Many-body Dispersion

Although many schemes just include pairwise interactions, there are cases where the corrections can be improved by including 3- and 4- body effects. In general, we would expect to include all possible coupled interactions. This is particularly important for interactions between molecules and surfaces.

Many-body dispersion[6] is a scheme to attempt to include these higher order effects to all orders by constructing an interaction matrix and diagonalising it to get the coupled VdW modes that would form. In general, this matrix is of size  $3N \times 3N$  where  $N$  is the number of atoms in an isolated system. To use this for a periodic system, we treat the system as a cluster and include neighbouring unit cells up to some interaction range to approximate the VdW energy for the central unit cell. The number of atoms is approximately  $m^3N$  where  $m$  is a factor large enough that the cluster size is larger than the interaction range. The problem with this method is that matrix diagonalisation is an expensive operation —  $O(m^9N^3)$  and this dominates the calculation time for small systems.

### 5.1 Theory

*Bučko et al*[9] have shown that the method can be formulated in a periodic way. This breaks the single large problem into a series of decoupled matrix diagonalisations that only extend across the unit cell, but at sampled points across the first Brillouin zone which contribute to the total energy. This allows the use of much of the internal CASTEP machinery for constructing and symmetrising k-point meshes and also fits nicely into the internal parallelisation scheme. To allow the same range of interactions, we only require  $m^3$  matrix diagonalisations of  $3N \times 3N$  atom matrices so the computational effort scales as  $O(m^3N^3)$ . This massively improves performance for small systems (where  $m$  is large and  $N$  is small) and allows us to go to higher interaction ranges as shown in figure 4.

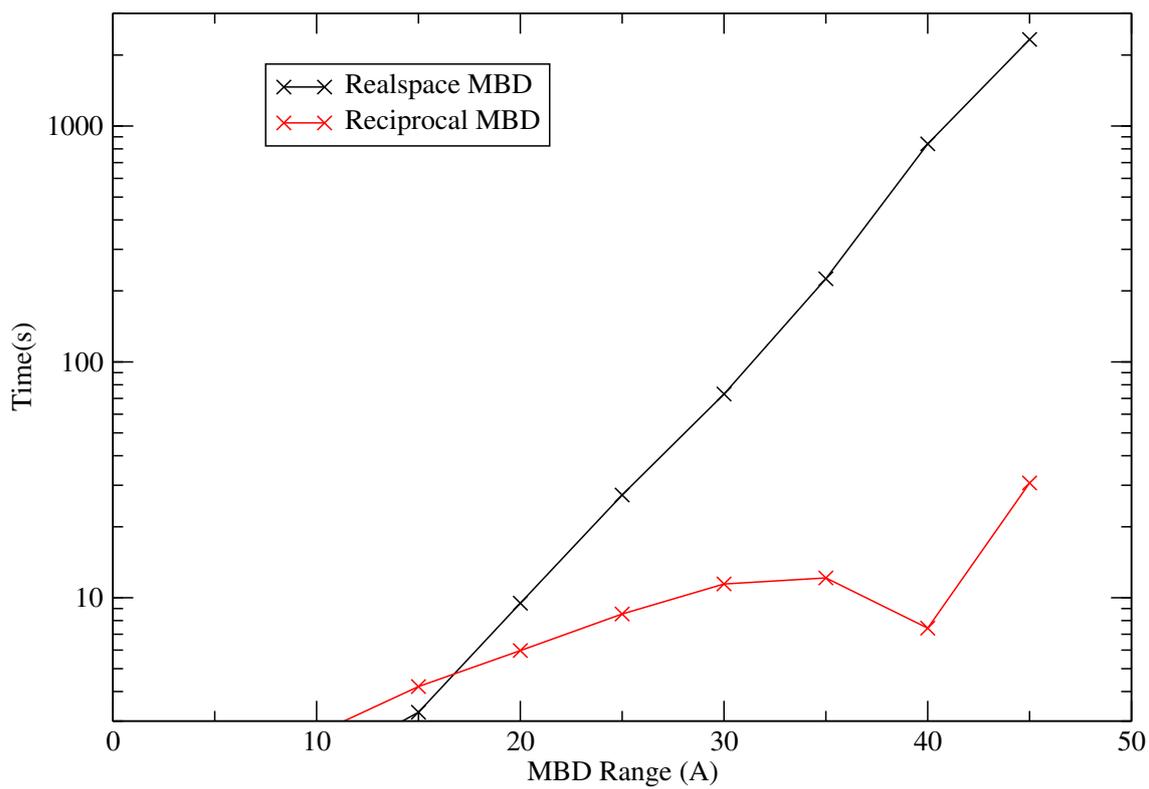


Figure 4: Time scaling of real and reciprocal space based MBD methods with interaction range.

For the purpose of comparison, we have kept the same interaction range as in the original MBD implementation ( $25\text{\AA}$ ) unless explicitly stated, but there is certainly room to improve this in the future.

The original MBD implementation calculated forces by finite displacement which added an extra factor of  $3N$  to the overall scaling. By calculating the forces analytically, many of the interaction contributions are 0 and so do not need to be calculated and the computational effort is reduced to  $3\times$  that of the VdW energy. Unfortunately this is not true for the stress and the computational effort to calculate the stress is approximately the same in either case. Due to this, we focused our effort on implementing the analytic forces in an efficient way and did not implement analytic stresses. Instead, the stress is implemented by a finite displacement scheme with a total cost of approximately  $12\times$  the cost of an individual energy evaluation.

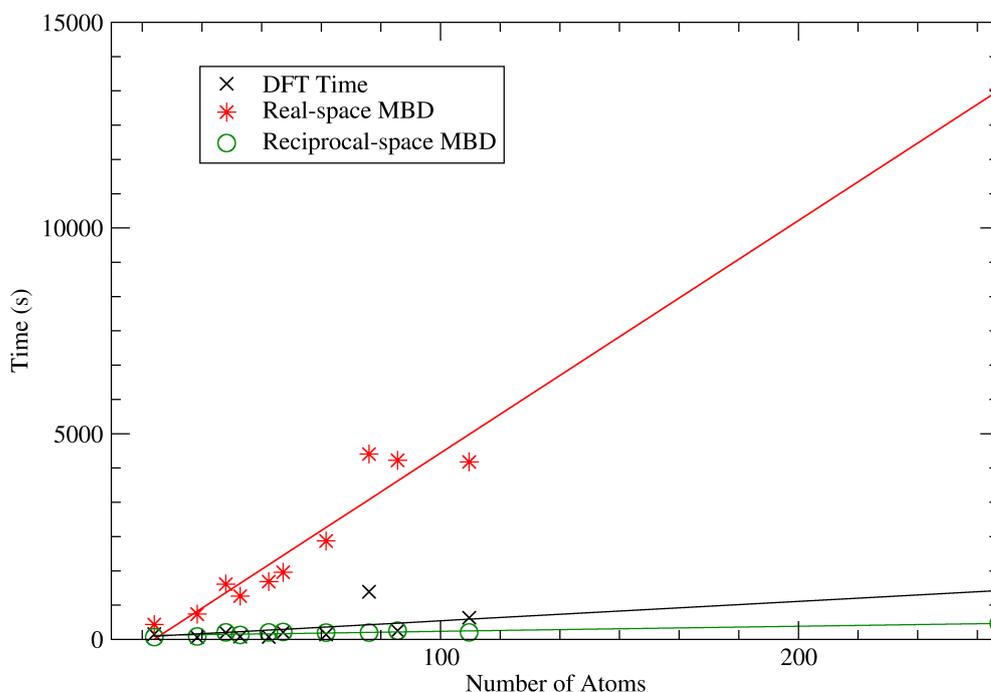


Figure 5: Comparison timings for a range of organic molecules using the many-body dispersion scheme. Lines are also fitted to show the linear nature of the scaling within this regime and as a guide to the eye. These calculations were performed in 24-way MPI on a single node.

As can be seen in figure 5, the direct MBD method is slower than DFT for our

system sizes and scales poorly with the number of atoms. For these systems, although the number of atoms within  $25\text{\AA}$  remains approximately constant, the calculation of forces still scales as  $N$  and so the time still increases rapidly from already being the dominating factor at small cell sizes. The cost is approximately 52s per extra atom. For the reciprocal space method, the cost of an extra atom is on the order of 1s.

With the periodic MBD method, the time for a small system is comparable with the DFT time which is mostly due to overheads in setting up constructs for reciprocal MBD. The method scales slightly better than the DFT time across the region of study and for the larger systems, it tends to be approximately half of the DFT time compared with the old scheme which was  $13\times$  larger.

This means that the method is now much less computationally intensive and can be routinely applied to a large variety of systems. The new modular SEDC interface also allows this VdW scheme to be used in finite-displacement phonon calculations. This is of particular interest as many of the pairwise schemes suffer from poor phonon prediction due to the damping functions being very sharp.

Further work in the future will be performed to determine if the  $25\text{\AA}$  cutoff is high enough for most materials and whether it should be increased now that it is computationally feasible.

## 6 Conclusion

The rewrite and parallelisation of the semi-empirical pairwise dispersion code(**WP1** and **WP2.a**) has been completed. This has been integrated into the main CASTEP codebase. Convergence is now faster for all systems and calculated in parallel, showing almost perfect parallelisation for even small systems. Support for the reference library supporting the D3/D4 method(**WP2.b**) of Grimme *et al* has been implemented and is now in the main CASTEP repository. Due to the licence on this library, it cannot be distributed commercially, but can be included with the academic release.

Finally, a periodic many-body dispersion method(**WP4**) has been implemented and along with analytic forces(**WP3**), shows a massive increase in performance. This makes the method now reasonable to use in a day-to-day calculation and gives high accuracy VdW results. This uses much of the internal parallelisation mechanics in CASTEP and so the maintenance will be relatively low. The time to calculate the MBD correlation energy is decreased by factor of six for 20-atom organic crystal and decreased by a factor of 25 for larger (108–256-atom) systems. The MBD time is now less than the DFT time rather than being the dominant part of the computation. There was an additional task in (**WP5.a**) to compute lattice energies for a range of organic molecular crystals; this work is being undertaken by project partners from the ADDoPT consortium.

## References

- [1] Stewart J. Clark, Matthew D. Segall, Chris J. Pickard, Phil J. Hasnip, Matt I. J. Probert, Keith Refson, and Mike C. Payne. First principles methods using CASTEP.

*Zeitschrift für Kristallographie - Crystalline Materials*, 220(5/6), Jan 2005.

- [2] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, 2006.
- [3] Alexandre Tkatchenko and Matthias Scheffler. Accurate molecular Van der Waals interactions from ground-state electron density and free-atom reference data. *Physical Review Letters*, 102(7), Feb 2009.
- [4] F. Ortman, F. Bechstedt, and W. G. Schmidt. Semiempirical Van der Waals correction to the density functional description of solids and molecular structures. *Physical Review B*, 73(20), May 2006.
- [5] Petr Jurečka, Jiří Černý, Pavel Hobza, and Dennis R. Salahub. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with *ab initio* quantum mechanics calculations. *Journal of Computational Chemistry*, 28(2):555–569, Jan 2007.
- [6] Alberto Ambrosetti, Anthony M. Reilly, Robert A. DiStasio, and Alexandre Tkatchenko. Long-range correlation energy calculated from coupled atomic response functions. *The Journal of Chemical Physics*, 140(18):18A508, May 2014.
- [7] Erik R. McNellis, Jörg Meyer, and Karsten Reuter. Azobenzene at coinage metal surfaces: Role of dispersive Van der Waals interactions. *Physical Review B*, 80(20), Nov 2009.
- [8] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-d) for the 94 elements H-Pu. *The Journal of Chemical Physics*, 132(15):154104, Apr 2010.
- [9] Tomáš Bučko, Sébastien Lebègue, Tim Gould, and János G Ángyán. Many-body dispersion corrections for periodic systems: an efficient reciprocal space implementation. *Journal of Physics: Condensed Matter*, 28(4):045201, Jan 2016.