PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

www.rsc.org/pccp

ARTICLE TYPE

Density functional calculations of extended, periodic systems using Coulomb corrected Molecular Fractionation with Conjugated Caps method (CC-MFCC)

Levin Brinkmann,[*a] **Eugene Heifets,**[b] **and Lev Kantorovich**[a]

A fragmentation scheme based upon the *Molecular Fractionation with Conjugated Caps* (MFCC) method and derived previously [Journal of Chemical Physics, v. **130**, No. 144104 (2009)] within the remit of density functional theory (DFT) based on local and semi-local functionals, enables one to perform order-N high-quality DFT calculations on extended systems (e.g. collections of organic molecules) via considering its smaller fragments. Here we discuss in detail a considerably improved method which broadens its applicability to a wider class of extended systems: (i) when each individual fragment is considered, the surrounding part of the entire system is not ignored anymore; instead, it is represented by point charges; (ii) the method is generalised to a system of any complexity enabling studying periodic and porous systems in real space; (iii) an appropriate *Coulomb correction* term is derived where clear distinction is made between charge densities of the same cap regions appearing in different fragments. Consequently, our correction term turns out to differ substantially from that derived e.g. by Li *et al.* [Journal of Chemical Physics A, v. **111**, No. 11, pp. 2193 (2007)]. We also discuss a possiblity for the point charges surrounding each fragment to update self-consistently following the calculations of every individual fragment. We examine here a new implementation of our method and its application to a Metal Organic Framework system. Specifically, we consider the structure of MOF-16 and adsorption of Hydrogen molecules in its pores. Possible ways of improving precision and to further widen up applicability of the method are also discussed.

1 Introduction

Understanding of various biological processes requires detailed knowledge of the properties of relevant macromolecules such as e.g RNA, DNA, proteins. Similarly, detailed knowledge is necessary for various porous crystals like zeolite and metal organic frameworks (MOF) which are being increasingly considered for applications in nanotechnology[1]. In both these cases accurate numerical description of the electronic structure of these complex materials from first principles is highly desirable. Unfortunately most of these systems are too large to be calculated by conventional quantum mechanical (QM) methods. Computational schemes for electronic structure calculations scaling linearly with the system size, the so-called order-N methods[2–8], have a great promise of extending our computational capabilities to these systems. In particular, there have been a number of partitioning schemes proposed in which the calculation of the whole extended system is replaced by a set of smaller first principle calculations

performed on its various parts, which have been recently reviewed[9]. These methods naturally scale linearly with the system size even though each individual calculation on a fragment may scale nonlinearly with the number of atoms in the fragment. A classification of such methods has been done[9] based on whether the total energy is obtained in a separate step after gathering the electron densities of the various parts or directly from the energies of the different fragments. Notable examples of the first class of methods are the divide-and-conquer[3,10] and the adjustable density matrix assembler (ADMA)[11] methods. Here we concentrate on the second class of methods, i.e. the so-called one-step[9] or energy based[12] methods.

There is a wide variety of the one-step methods[9], which could e.g. be additionally subdivided by whether QM calculations are performed on (i) monomers, (ii) unions of two or more fragments or (iii) intersecting monomers[13]. In the *fragment molecular orbital* (FMO) method[14,15], which is an example of the second group (ii), unions of two fragments are calculated and in this way the corresponding two-body interaction is fully recovered. Later this method was extended to account for three-[16] and four-body[17] interactions of the many-body expansion (MBE). Usually, to retain the linear scaling,

[*] E-mail: lbrinkm@uni-goettingen.de
[a] Department of Physics, King's College London, London, WC2R 2LS, United Kingdom.
[b] Max Planck Institute for Solid State Research, Heisenbergstr. 1, D-70569 Stuttgart, Germany.

many-body interactions of each fragments with only adjacent fragments are calculated at the QM level.

Efforts have been made within the context of different fragmentation approaches to improve the electron density obtained in different parts of the original system by enlarging the corresponding fragments so that surrounding parts of the fragments are also attached[18–23]. Various one-step methods based on intersecting monomers, and thus belonging to the third group (iii), go back to the main ideas of *molecular fractionation with conjugate caps* (MFCC) method[18,24], in which the macromolecule under consideration is divided into overlapping fragments capped by terminating groups of atoms; then the interaction energy of the system (of e.g. a molecule interacting with a surface) is calculated as a simple sum of the corresponding contributions due to each fragment, minus the contributions due to the overlapping parts of the fragments (i.e. caps). Concepts of the second (ii) and the third (iii) types of fragmentation were mixed when two-body interactions were also added to methods based on overlapping fragments either as a correction[20,25,26] or to additionally account for non-bonded interactions[27–33]. The relation between the two approaches however has been illuminated in more detail recently when an MFCC-like fragmentation scheme has been obtained using the MBE approach applied to auxiliary fragments[26]; a correction term has also been suggested. Other MBE schemes have also been proposed based on overlapping fragments[13,34], which reduce to conventional MBE methods in the case of non-overlapping fragments. Note that MFCC within this framework can be considered as the one-body limit of the MBE with overlapping fragments.

A considerable effort has been spent in the automatisation of the fragmentation schemes for macromolecules[17,35,36]. Collins and Deev[27] developed a method in which fragments are built of small atomic groups consisting of monomers, dimers, trimers, etc. With the so-called higher level of fragmentation, larger fragments are generated and thus convergence with respect to fragment size can be tested. Although the fragmentation scheme appears different, the expression for the total energy does follow the original MFCC formulation.

To include the effect of long-range Coulomb interaction, the electrostatic environment was added when solving for each fragments electronic structure[12,32,33,37]. This approach was shown to improve the accuracy of the fragmentation methods, when Li *et al.*[12,36] presented their *generalised energy-based fragmentation* (GEBF) method. Here MFCC was also reformulated in a more general way in order, amongst other generalisations, to allow for caps themselves to intersect.

Inspired by the original formulation of MFCC[38], we recently proposed a fragmentation method[39] based on a slightly different fragmentation scheme. In our method[39], using rather general assumptions on the system partitioning and the fact that the total energy of a system of interest can be well rep-

resented within the density functional theory (DFT) using one of the existing local or semi-local density functionals, such as LDA or of various GGA flavours, not only the appropriate expression for the total energy was rigorously *derived,* sufficient conditions which are to be satisfied for the fragmentation method to work were also clearly formulated.

In line with the original ideas of MFCC, in our earlier work[39] the role of the caps was to mimic the effect of adjacent fragments[18]; correspondingly, the caps were treated as artificial. In that work for the first time it was explicitly differentiated between an artificial density within the caps and the true or exact electron density within the core regions of fragments which allowed to eliminate exactly the contribution of the caps into the total energy and forces. This led to a new type of correction compensating for the artificial electrostatic energy due to the caps. While this type of correction was implicitly accounted for in early implementations of MFCC[24], the correction due to this artificial interaction got lost in later incarnations of the method (e.g. when embedding fragments into the Coulomb field of their surrounding[12]) or was argued to be negligible[20] (which of course may well depend on the system in question).

Moreover, in our earlier work[39] no account was made of the charge distribution outside the fragments. In this work we generalise this method whereby each fragment is considered in the Coulomb field of surrounding fragments. Correspondingly, we present here an MFCC inspired fragmentation method explicitly correcting for the artificial contributions of the Coulomb interactions between the caps, the central core and surrounding fragments. The resulting expression for the Coulomb correction, derived from the exact DFT expression of the total energy, turned out to differ substantially from the ones published elsewhere[12,33], which is precisely due to the distinction, mentioned above, between an artificial electron density of the caps and the true density in the core region of the fragments which is accounted for explicitly in this work. Consequently, our method is named Coulomb corrected MFCC (CC-MFCC) method. In our present implementation and in line with Refs.[12,33,37,38,40] the Coulomb field due to the charge distribution around each fragment is represented by point charges. Similarly to the earlier work[12,19], these could be obtained in an iterative way using the Mulliken charges on the atoms of the corresponding surrounding parts of the system. In this work we present a correction due to the artificial Coulomb interaction of the caps including an embedding of the fragments into a Coulomb field. To our knowledge, this type of correction within the MFCC fragmentation calculation is proposed for the first time.

In our first publication on this method[39] calculations performed on a number of pairs of simple molecules bound to each other by hydrogen bonds showed a great promise for the method, as it was found that high precision can actually be

achieved in the total energy, geometry (bond lengths and angles) and atomic charges as compared to the reference calculation within the same DFT flavour performed on the whole system (no fragmentation). In this earlier work the derivation of the method was limited to the specific case of an organic dimer[39]. In this paper a more general formulation of our fragmentation scheme applicable to a system of arbitrary complexity including periodic (crystalline) systems is proposed, and an application of our CC-MFCC method to a rather complex metal-organic framework (MOF) system is presented. These highly porous crystals generally consist of metal organic cores linked by organic chains and are of interest in a wide range of fields[1] including hydrogen storage[41,42], electrode design[43], catalyst[44] and as vehicles for drug delivery[45]. In many of these applications a detailed understanding of the electronic structure of the framework is desired e.g for optimising binding of small molecules including hydrogen[41]. We note that dense crystals[31] and liquids[40] were simulated by fragmentation methods in the spirit of GEBF, while adsorption[46] in zeolites and diffusion[47] in mesoporous silica were studied using the FMO method. By considering a metal organic framework here we present - to our knowledge - a new application of the MFCC method. Fragmentation methods may be specifically suitable for porous crystals as their functional groups are usually well separated and a partition scheme may be easily developed. The system we calculated is known as MOF-16 and was first synthesised by O. M. Yaghi[48]. It is a relatively large system making a fragmentation feasible. On the other hand the unit cell of MOF-16 is still small enough to be easily accessed by common approaches not based on any fragmentation, and hence would provide an excellent reference. As such this system serves well as a convenient illustration of the advantages of our fragmentation method. Also the fragmentation for this system could be challenging as it contains benzol rings which are known for their delocalised electrons.

The plan of the paper is as follows. In the next section we shall consider in detail a simple one-dimensional system to illustrate the main aspects of our method; specifically, we shall comment on the necessary implications related to system periodicity. Then a general description of the proposed CC-MFCC method will follow which is valid for a periodic system of any dimension. Subsequently, self-consistency and implementation of the method are discussed and preconditions on the systems electronic structure for an effective partition are indicated. Then the application to the test system MOF-16 is presented along with a detailed discussion of the point charge environment and of other computational details. A short discussion of our results with possible further development of our method is given in the last Section. Atomic units are used throughout this paper if not stated otherwise.

2 Theory

2.1 One-dimensional system

Consider an infinite periodic 1D system, e.g. a polymer or a chain of molecules placed "head-to-tail" to each other. The system can be broken into individual regions as is schematically represented by a single black bar and the vertical lines in Fig. 1(a). Note that the boundaries between regions in the current implementation of the method, in which the concept of atomic charges on atoms is used, should pass somewhere between atoms; their explicit definition is not required.

A common assumption, which is more or less explicitly stated within most fragmentation methods, is that if one can represent the surrounding of a region sufficiently well, then the correct electron density (i.e. close to the density obtained by considering the whole system without partitioning it) within that region can be obtained. In line with the MFCC method each region is terminated by caps consisting of parts of the actual system surrounding the region. In a way, the region of interest is simply extended in size and hence will incorporate parts of its neighbourhood; in our terminology it will then be called a fragment. Fragments are objects which are individually calculated within the fragmentation scheme.

One can assume that increasing the fragment size and hence that of the caps yields the density within the original region to converge to the correct density; necessary preconditions will be discussed in further detail in the context of DFT later on. The assumption stated trivially holds in the limit of considering caps that incorporate the whole remaining system. However, care is needed here as large caps would increase the overhead and prevent one from gaining a proper speedup compared to a calculation of the whole system.

We address this issue by the following design. Caps contain a part of the system surrounding the given region, but are terminated by link atoms. These link atoms are assumed to occupy positions of the actual atoms of the system, but do not correspond to them chemically. The only purpose of the link atoms is to terminate the caps in such a way as to minimise the distortion of the electron orbitals of the fragment and hence drastically improve the convergence of the electron density in the regions with respect to their caps size. To further improve on both the electron density of the caps as well as of the region, the fragment may be placed into the Coulomb field of the surrounding regions.

Generally each region is to be capped at its both ends (left and right) by parts of its neighbouring regions. To clarify our notations, let us consider region J having two neighbouring regions I and K, see Fig. 1 (a). The region J will then have two caps. The cap J_I terminates region J from the left by a cap taken from the neighbouring region I; similarly the cap J_K terminates region J from the right by a cap extracted from re-
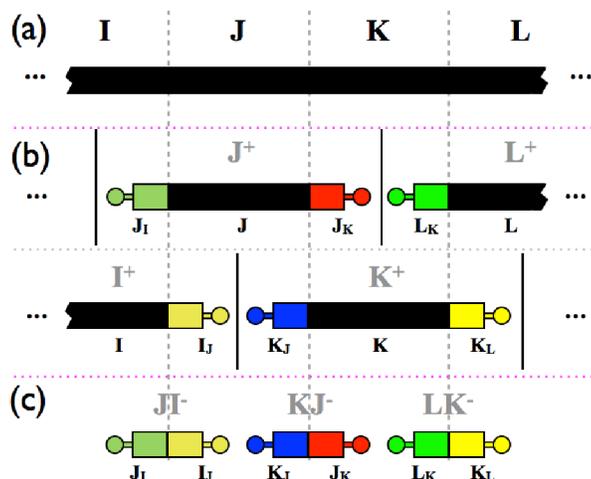
Fig. 1 Schematic representation of a 1D system: (a) the original system which is broken down into separate regions ..., I, J, K, L, ..., the division is shown by vertical broken lines; (b) "positive" fragments J^+ and K^+ (as well as partially shown I^+ and L^+), consisting of the corresponding "black" regions of the original system and terminated at both sides with caps; the latter consist of a part of the original system (shown by a coloured rectangular) and a terminating link atom shown as a small circle of the same colour. For instance, the fragment $J^+ = [J_I, J, J_K]$ consists of the region J terminated from the left by the cap J_I containing a part of region I and by the cap J_K from the right containing a part of region K (hence the subscripts in the caps notations); (c) "negative" fragments, consisting of the two caps associated with the common border, e.g. the caps K_J and J_K make up a negative fragment $KJ^- = [K_J, J_K]$.

gion K. The caps are named by the region they terminate with the subscript indicating the region they are actually part of. A region of the original system together with its corresponding caps will be called a "positive" fragment and denoted by the regions name with the attached plus sign, e.g. $J^+ = [J_I, J, J_K]$ which is shown in Fig. 1 (b). Similarly we can define the "positive" fragment $K^+ = [K_J, K, K_L]$, as shown in Fig. 1 (b). We also define 'negative' fragments containing two caps of neighbouring regions at a common border, e.g. the caps of the common border of J and K form the negative fragment $KJ^- = [J_K, K_J]$ (see Fig. 1 (c)). The terminology of "positive" and "negative" fragments is convenient, as it corresponds to the sign with which the respective fragments energy enters the final energy, as it will described in detail later on. Note that the notation for the negative fragment is symmetric in both letters used, i.e. $XY^- = YX^-$. Note also that these negative fragments correspond to the conjugated caps in early work on MFCC[18]. For clarity we reserve the term *system* $\mathscr{S}$ for the unpartitioned original system, while parts of the system are called *regions*. The term *fragment* should only be used for the

"molecules" calculated individually, namely the positive and negative fragments described before.

The scheme presented so far can be used for various system types differing in the number of chunks and the type of the boundary conditions, i.e. whether the system is considered as periodic or not. A relatively simple example of a periodical system is an infinite copolymer consisting of two different monomers, which are periodically repeated, see e.g. J and K in Fig. 1. One can map such a system to the scheme in Fig. 1 by identifying $L \equiv \tilde{J}$ and $I \equiv \bar{K}$ as translated copies of J and K, respectively. Here the overbar indicates a translation to the left while the tilde indicates a translation to the right. The positive fragments are similar to the two described before, namely $J^+ = [J_{\bar{K}}, J, J_K]$ and $K^+ = [K_J, K, K_{\tilde{J}}]$. Identifying $L_K \equiv \tilde{J}_K$ with J_I and $K_{\tilde{J}}$ with I_J there are two distinct negative fragments $KJ^- = [K_J, J_K]$ and $LK^- = [L_K, K_L] \equiv [\tilde{J}_K, K_{\tilde{J}}] = \tilde{J}K^-$. The negative fragment $IJ^- = [I_J, J_I]$ in Fig. 1 is a translated copy of the latter one. Note that exclusively parts of the two regions J and K are used for the caps; however, in some cases atomic positions have to be shifted by a single translation as indicated by the overbar/tilde. The system of a copolymer presented here will be generalised into three dimensions in section 4.1.

A simple finite system is the one of a dimer. A fragmentation method of such a system suggested earlier in Ref.[39] can in fact be recast into a form which is equivalent to the scheme we present in this paper. Indeed, consider a 1D collection of two molecules A and B connected head-to-tail with each other as shown in Fig. 2(a). If we now divide each molecule into three regions A_i and B_i ($i = 1, 2, 3$) as shown in Fig. 2(c), then the fragments introduced with the current method and shown in Fig. 2(b) would correspond exactly to the fragments introduced previously with the notations shown on the right there. One can see that the old fragment A now corresponds to the new fragment $I^+ = [I; I_J] \equiv [A_3; A_2; A_1] = A$; similarly, $K^+ = [K_J; K] \equiv [B_1; B_2; B_3] = B$, $J^+ = [J_I; J; J_K] \equiv [A_2^\bullet; A_1; B_1; B_2^\bullet] = C$, $JI^- = [J_I; I_J] \equiv [A_2^\bullet; A_1] = C_A$ and $KJ^- = [K_J; J_K] \equiv [B_1; B_2^\bullet] = C_B$, where the fat dot near the symbol of the region in the old notations indicates the fact the a link atom is attached to this particular region; this is left out in the current notations as each cap contains a link atom by default. Here regions A_1 and B_1 of the previous method represent the "caps" I_J and K_J of the current one; these caps did not have a link atom attached to them in our earlier version of the method[39]; however, as it was already mentioned in[39], this could have been done. If it was, as shown by the dashed lines in Fig. 2(b), then the partition into fragments would exactly correspond to the current scheme with a fragmentation into three parts. This example is also useful in that it clearly shows that the caps are actually constructed out of the regions of the original system (e.g. $J_I \equiv A_2^\bullet$ and $I_J \equiv A_1$).

Hence, one can apply the fragmentation scheme to both periodic and finite systems. Two important observations can be
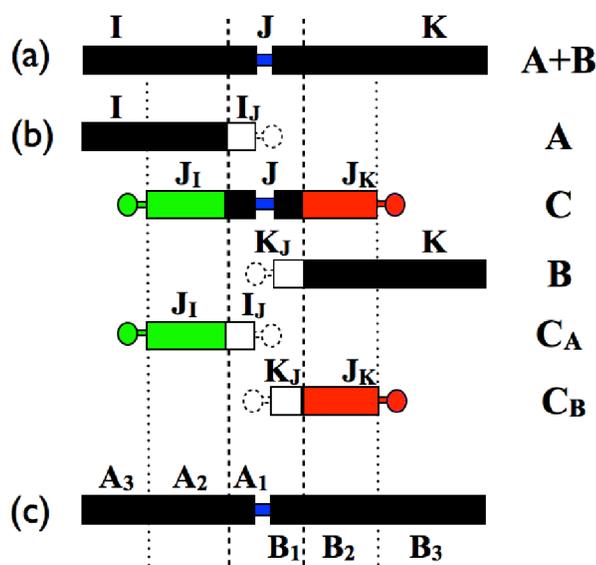
Fig. 2 Comparison of the current and the previous[39] fragmentation schemes for the case of two molecules A (on the left) and B (on the right) bound to each other as shown in (a); (b) five fragments as proposed in the current scheme for a partition into three regions; we stress that no link atoms are used in the previous scheme[39] to terminate the right and the left ends of the molecules A and B; as noted before[39] this could have been done and hence it is shown here with the dashed line to ease the comparison; on the right in (b), previous[39] notations for the fragments are also given; (c) a detailed division of the system $A + B$ of the two molecules into 6 regions used before[39]. It is clearly seen from this detailed division that the "caps" used in (b) in fact represent regions of the actual molecules.

made irrespective of the topology of the system. Firstly, if all "black" (internal) parts of the "positive" fragments are combined, they will make up exactly the whole original system. Secondly, each cap in "positive" fragments has one and only one exact counterpart in the "negative" ones; in the case of a periodic system one has to consider only a finite number of "positive" and "negative" fragments which are periodically repeated.

Next, we introduce the necessary notations for the electron densities. Care is needed here since the same regions may appear in different fragments at the same time which, when calculated, would yield different densities. We shall call $\rho_X^Y(\mathbf{r})$ the density of a region X obtained by considering fragment Y. We also use $\rho_X^0(\mathbf{r})$ as the "exact" density of the same region which is obtained by considering the whole system in a single calculation; this density serves as the reference density which we can compare with. If region J within fragment J^+ is caped appropriately, then in line with the discussion at the beginning of this section we can assume that the electron density in region J is approximately the same as the exact one:

$$\rho_J^{J^+}(\mathbf{r}) \approx \rho_J^0(\mathbf{r}), \qquad (1)$$

while the density in the caps, e.g. the density $\rho_{J_K}^{J^+}(\mathbf{r})$ of region J_K of fragment J^+, would certainly differ from the exact one, $\rho_{J_K}^0(\mathbf{r})$. Now, consider the cap J_K that terminates two different fragments J^+ and KJ^-. Since K_J, which is the second cap in the fragment KJ^-, is a part of the region J in the fragment J^+, the densities in the cap J_K would be approximately the same in both fragments J^+ and KJ^-, no matter in which of the two fragments they are considered, i.e. $\rho_{J_K}^{J^+}(\mathbf{r}) \simeq \rho_{J_K}^{KJ^-}(\mathbf{r})$; this will be true provided the region K_J in the fragment KJ^- represents well the region J at the common border of J and K. A single superscript 1 will be used for the densities of the caps, no matter if they were calculated within a positive or a negative fragment:

$$\rho_{J_K}^{J^+}(\mathbf{r}) \approx \rho_{J_K}^{KJ^-}(\mathbf{r}) \equiv \rho_{J_K}^1(\mathbf{r}). \qquad (2)$$

Summarising, only two types of densities exist in our approach: those in the caps, calculated either in "positive" or "negative" fragments, e.g., $\rho_{J_K}^1(\mathbf{r})$, and those in the internal parts of the "positive" fragments, $\rho_J^0(\mathbf{r})$. In the following, if not indicated otherwise, an index J will refer to the region J with the actual density which has the superscript zero, while e.g. J_K will refer to the cap region J_K with auxiliary density having the superscript one.

2.2 Derivation of the energy expression

The idea of the fragmentation method is to avoid the calculation of the whole system and consider instead a collection of smaller fragments which energies, when combined appropriately, give a very good approximation to the exact energy (and hence forces) for the entire system. To construct the appropriate energy expression for the whole system via total energies of the appropriate fragments, we note following Ref.[39] that within any of the local or semi-local density functionals of the Density Functional Theory (DFT) the total energy of an arbitrary system for the given position of atomic cores consists of essentially two types of terms:

$$
\begin{aligned}
E[\rho(r)] &= \int \varepsilon_{loc}[\rho(r)]\, dr + \frac{1}{2} \int \int \frac{dr\, dr'}{|r - r'|} \rho_{tot}(r)\rho_{tot}(r') \\
&= E_{loc} + E_{Coul} \qquad (3)
\end{aligned}
$$

The first term, E_{loc}, has a *local* character and is *volume-additive* since it is represented via a single integral over the volume of the entire system, with the local energy density

$$\varepsilon_{loc}[\rho(r)] = \left(-\frac{1}{2}\Delta_r + \widehat{V}_{ps}'\right) \rho(r, r')\,|_{r' \to r} + \rho(r)\varepsilon_{xc}[\rho(r)] \qquad (4)$$

containing the kinetic energy, interaction of electrons with *chargeless* atomic pseudo-cores (i.e. $\widehat{V}'_{ps}$ differs from the usually used atomic pseudopotentials $\widehat{V}_{ps}$ in that the Coulomb (long-range) part of the pseudopotential is removed), and finally the exchange and correlation energies (the last term). The second term in the right hand side of Eq. (3), E_{Coul}, represents the total Coulomb interaction within the system. Note that the full interaction of the electrons and the atomic cores as well as the mutual interaction of atomic cores are both included since $\rho_{tot}(r) = -\rho(r) + \rho_c(r)$ represents the total charge density which incorporates the charge density $\rho_c(r)$ of the atomic cores and the electron density $\rho(r)$. The Coulomb term, E_{Coul}, is obviously not volume-additive as it is given by a double volume integral.

If we now split the entire volume V available to the system *completely arbitrarily* into non-overlapping regions, $V = V_1 \cup V_2 \cup \ldots$, then it becomes apparent that the total energy

$$E_{tot} = \sum_J E_J^{int}[\rho] + \frac{1}{2} \sum_{I,J\,(I \neq J)} C_{I,J}[\rho] \tag{5}$$

can be considered as a sum of internal energies,

$$E_I^{int}[\rho] = \int_{V_I} \varepsilon_{loc}[\rho(\mathbf{r})]\, d\mathbf{r} + \frac{1}{2} \int_{V_I} d\mathbf{r}_1 \int_{V_I} d\mathbf{r}_2 \frac{\rho_{tot}(\mathbf{r}_1)\rho_{tot}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{6}$$

depending only on the electron density in the given local region V_I (the integration is performed only within V_I), and a term due to Coulomb interaction between *different* regions

$$C_{I,J}[\rho] = \int_{V_I} d\mathbf{r}_1 \int_{V_J} d\mathbf{r}_2 \frac{\rho_{tot}(\mathbf{r}_1)\rho_{tot}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \ , \ I \neq J \tag{7}$$

Note that for convenience the internal energy above in Eq. (6) includes the intra-Coulomb interaction energy within the volume V_I. As the integration is performed in separate regions I and J and according to condition (1), the charge densities of the individual regions used in Eqs. (6) and (7) can be taken from the calculations on the corresponding positive fragments, e.g.:

$$C_{I,J} = \int_{V_I} d\mathbf{r}_1 \int_{V_J} d\mathbf{r}_2 \frac{\rho_{tot,I}^{I^+}(\mathbf{r}_1)\rho_{tot,J}^{J^+}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \ , \ I \neq J \tag{8}$$

where the total charge $\rho_{tot,J}^{J^+}(\mathbf{r}_1) = -\rho_J^{J^+}(\mathbf{r}) + \rho_{c,J}(\mathbf{r})$ is composed of the electron, $\rho_J^{J^+}(\mathbf{r})$, and core, $\rho_{c,J}(\mathbf{r})$, densities of the region J.

We can now write down the energy expression for the original system shown in Fig. 1(a) using the fragments densities:

$$E_{tot} \approx \sum_J E_J^{int} + \frac{1}{2} \sum_{I,J\,(I \neq J)} C_{J,I} \tag{9}$$

Here E_J^{int} is the internal energy of region J obtained from a calculation of the fragment J^+. Similarly, one can write down the total energies of every individual fragment.

We are now ready to derive the general expression for the total energy of the partitioned system via energies and densities of its fragments. However, it is convenient at this stage to introduce some additional notations which would allow us later on to generalise the formalism to systems of arbitrary structure and dimensionality. Consider a region J of an arbitrary partitioned system $\mathscr{S}$ having N_J neighbouring regions. A positive fragment will now have not just two caps (on the left and right) as in the 1D system considered so far, but N_J of them, one at the border of each of its next neighbours. They will still be denoted as J_K with the index K indicating the neighbouring region the cap is pointing to. The set of all such nearest neighbours of region J will be denoted by calligraphic $\mathscr{J}$. Consequently the region J together with its set of caps $\{J_K; \forall K \in \mathscr{J}\}$ will make up the positive fragment J^+. Similarly the N_J negative fragments associated with J are $\{JK^- = [J_K, K_J]; \forall K \in \mathscr{J}\}$. Note that these notations do not imply any kind of structure or dimensionality.

Aiming to express the total energy as a sum of energies of the different fragments and some correction term, we need to write down the total energies of every fragment and then compare with the total energy of the whole system (9). When considering each fragment, we will be assuming that the charge density surrounding each fragment, $\rho^0(\mathbf{r})$, provides an external potential acting on electrons and cores of the fragment. Here we take the actual charge density from each region surrounding the fragment in question, and hence the superscript 0 in the density. Similarly to the total energy of the whole system (9), we write down the total energy of a "positive" fragment J^+ in detail as

$$E_{J^+} = E_J^{int} + \sum_{K \in \mathscr{J}} E_{J_K}^{int} + C_{J^+} \tag{10}$$

where the first two terms represent the internal energies of the regions and its caps (which are additive), whereas the third one is the Coulomb interaction energy:

$$\begin{aligned} C_{J^+} = & \sum_{K \in \mathscr{J}} \left(C_{J_K,J} + \frac{1}{2} \sum_{L \in \mathscr{J} \setminus K} C_{J_K,J_L} \right) \\ & + \sum_{K \in \mathscr{J}} \left(C_{J,K \setminus J_K^0} + \sum_{L \in \mathscr{J}} C_{J_L,K \setminus J_K^0} \right) \\ & + \sum_{I \in \mathscr{S} \setminus \mathscr{J}} \left(C_{J,I} + \sum_{L \in \mathscr{J}} C_{J_L,I} \right) \end{aligned} \tag{11}$$

The first term here is the mutual Coulomb interaction of the internal region J and all its caps within the fragment J^+ (note that intra-Coulomb interactions within each region are already

included in the internal energies). The second term describes the Coulomb interaction between the fragment J^+ and the remainder of the neighbouring regions K obtained by removing the caps J_K^0 from them. Here J_K^0 is composed from all atoms in region K which belong to the cap J_K *including* the actual atom of K, which is replaced by a link atom in J_K. Finally, the last term in Eq. (11) is the Coulomb interaction between J^+ and all other regions around it excluding the neighbouring ones. The individual Coulomb interaction terms C_{XY} are of similar form to the ones used in the total energy (8). For the internal regions J, K,... the densities used corresponds to the "exact" electron density according to the assumption stated by Eq. (1). In contrast for the caps the density of the associated fragment is to be used, which corresponds to the assumption (2) for the auxiliary density. For instance, $C_{J,I}$ describes the Coulomb interaction between regions J and I calculated using "exact" charge densities $\rho_J^0(\mathbf{r})$ and $\rho_I^0(\mathbf{r})$ of the system (i.e. the corresponding charge densities of the atomic cores are also included), while e.g. $C_{J_K,I}$ describes the Coulomb interaction between regions J_K and I based on charge densities $\rho_{J_K}^1(\mathbf{r})$ and $\rho_I^0(\mathbf{r})$, and so on. Note that when calculating the interaction energy of either the region J or one of its caps with the adjacent regions $K \in \mathscr{J}$, we have only to include atoms of region K which are not contained in region J_K, hence the discrimination between neighbouring regions and the remaining ones in the second and third terms. In particular, the actual atom of region K which is replaced by a link atom belonging to J_K is also to be removed. All atoms in K apart from those that belong to J_K are referred to as belonging to the subregion $K \backslash J_K^0$ in the second term in Eq. (11) which is obtained by "subtracting" the set of atoms J_K (together with the exact electron density associated with the corresponding sub-region, see Fig. 3) from the set K. As the charge density associated with the removed atoms is assumed to be "exact", the superscript 0 is consistent with the notation used here for the cap J_K^0.

While the internal interaction represented by the first term in Eq. (11) is independent of the electrostatic environment, the latter two terms do depend on it. In particular, the contribution in these terms will depend on the implementation of the environment; for instance, if the fragment is calculated without Coulomb field due to surrounding regions, these two terms are dropped; in our implementation the Coulomb interaction is only considered within a cutoff radius.

The equation (11) above can be simplified if we notice that the Coulomb interaction terms $C_{J,K \backslash J_K^0}$ and $C_{J_L,K \backslash J_K^0}$ under the second sum can be written slightly differently. Consider, for instance, the first one, $C_{J,K \backslash J_K^0}$. It corresponds to the Coulomb interaction of region J and the rest of region K obtained by removing from it all atoms of the cap J_K including any link atoms associated with it, see Fig. 3. As was explained above, the removed atoms correspond to region J_K^0, and hence the
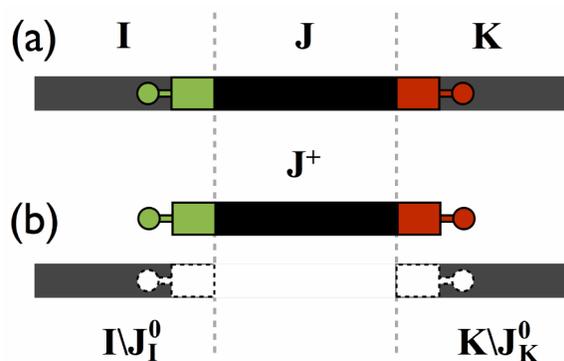
Fig. 3 To the calculation of the Coulomb field acting on the fragment J^+ from the rest of the adjacent regions I and K. (a) The fragment J^+ contains caps J_K and J_I which mostly consist of atoms of the neighbouring regions K and I, respectively (indicated in red and green). (b) Separation of the fragment J^+ and its surrounding (sub-regions $I \backslash J_I^0$ and $K \backslash J_K^0$, indicated in grey) shown explicitly. Note that when calculating the Coulomb field acting on the fragment J^+ from the rest of regions I and K the atoms and the corresponding "exact" electron densities due to the caps J_K and J_I of the fragment J^+ (shown using broken lines) are to be removed. As the density to be removed is the "exact" one, the superscript 0 is used in the notations of the sub-regions.

notation $K \backslash J_K^0$ for the atoms in K which are left after the removal. Now, we have found that the simplification is possible if we formally add and subtract the interaction C_{J,J_K^0} of J with the removed atoms of K, yielding:

$$C_{J,K \backslash J_K^0} = \left[C_{J,K \backslash J_K^0} + C_{J,J_K^0} \right] - C_{J,J_K^0} = C_{J,K} - C_{J,J_K^0} \qquad (12)$$

Note that now the first term, $C_{J,K}$, in the right hand side contains the Coulomb interaction of J with the whole region K. Since, following approximation (1), the "exact" density of K is used in $C_{J,K}$, the "exact" density of its part J_K^0 in C_{J,J_K^0} is also to be used. Similarly, the interaction $C_{J_L,K \backslash J_K^0}$ between the cap J_L and the region $K \backslash J_K^0$ is calculated as a difference of two terms:

$$C_{J_L,K \backslash J_K^0} = C_{J_L,K} - C_{J_L,J_K^0} \qquad (13)$$

Note that in the case of $K = L$ in (13) there is a Coulomb interaction between two different densities within the same region: while J_K refers to the density $\rho_{J_K}^1(\mathbf{r})$ obtained using either of the fragments J^+ or JK^-, J_K^0 refers to the density $\rho_{J_K}^0(\mathbf{r})$ of the same region. The latter density is obtained while considering the fragment K^+ which contains all atoms of region J_K (including the corresponding link atom(s)), see Fig. 1(b).

Using the above two equations, the Coulomb energy (11)

can be conveniently rearranged:

$$C_{J^+} = \sum_{I \in \mathscr{S} \backslash J} C_{I,J} + \sum_{K \in \mathscr{J}} \left(\sum_{I \in \mathscr{S} \backslash J} C_{J_K,I} - C_{J,J_K^0} \right.$$

$$\left. - \sum_{L \in \mathscr{J}} C_{J_K,J_L^0} + \frac{1}{2} \sum_{L \in \mathscr{J} \backslash K} C_{J_K,J_L} \right) \qquad (14)$$

Applying the same type of reasoning to a negative fragment, one obtains:

$$E_{JK^-} = \left(E_{K_J}^{int} + E_{J_K}^{int} \right) + C_{JK^-} \qquad (15)$$

where the corresponding Coulomb energy is:

$$C_{JK^-} = \sum_{L \in \mathscr{S}} (C_{J_K,L} + C_{K_J,L}) + C_{J_K,K_J} - C_{J_K,J_K^0}$$

$$- C_{K_J,K_J^0} - C_{J_K,K_J^0} - C_{K_J,J_K^0} \qquad (16)$$

Note that for the internal energy and the individual Coulomb terms it is not important whether a cap is situated in a positive or in a negative fragment, as the corresponding charge density is identical in both cases by the assumption (2).

It is easy to see now that in order to match the sum of local energies in Eq. (9), one has to sum up the local energies of the regions, "black" coloured in our 1D example (Fig. 1), which are only available from the energies of the "positive" fragments; at the same time, if we simply sum up these energies, we shall also acquire a sum of all local energies of the caps. However, the sum of the caps local energies is contained exactly in the sum of total energies of the "negative" fragments. Therefore, by subtracting the sum of total energies of all "negative" fragments from the sum of the energies of all "positive" ones (and hence their names), one shall recover exactly the required sum of the local energies of the "black" regions. This is due to the fact that (i) the local energies are *volume-additive*; (ii) "black" regions assemble exactly into the original system and (iii) every cap in the "positive" fragment has its single counterpart in one of the "negative" ones, as was noticed above.

At the same time, since the Coulomb energy is not *volume-additive*, there will be a correction term to the energy which is needed to avoid double counting of some interaction energies and to compensate for the interaction with the auxiliary charge densities of the caps:

$$E_{tot} = \sum_{J} \left[E_{J^+} - \frac{1}{2} \sum_{K \in \mathscr{J}} E_{JK^-} \right] + \Delta E_{Coul} \qquad (17)$$

Here the factor of one half takes care of the fact that while summing over all positive fragments J^+ the same negative fragment due to neighbouring region K will appear twice as JK^- (when $K \in \mathscr{J}$) and KJ^- (when $J \in \mathscr{K}$).

To obtain the energy correction, we take the total Coulomb interaction energy of the whole system as in Eq. (9), then subtract from it the Coulomb energies associated with all positive fragments, Eq. (14), and add those of the negative ones, Eq. (16):

$$\Delta E_{Coul} = \frac{1}{2} \sum_{J,J' (J \neq J')} C_{J,J'}^{00} - \sum_{J} \left[C_{J^+} - \frac{1}{2} \sum_{K \in \mathscr{J}} C_{JK^-} \right] \qquad (18)$$

This gives after some straightforward algebra:

$$\Delta E_{Coul} = \sum_{J} \left[-\frac{1}{2} \sum_{I,I \neq J} C_{J,I} + \sum_{K \in \mathscr{J}} \left[C_{J,J_K^0} + \frac{1}{2} C_{J_K,K_J} \right. \right.$$

$$\left. \left. - C_{J_K,K_J^0} - \sum_{L \in J \backslash K} \left(\frac{1}{2} C_{J_K,J_L} - C_{J_K,J_L^0} \right) \right] \right] \qquad (19)$$

Interestingly, the correction does not have anymore any terms containing Coulomb interaction within the same spacial regions, they all have cancelled out in the final expression.

The forces $\{\mathbf{F}_A\}$ on atoms A are obtained by differentiating the total energy with respect to atomic positions $\mathbf{r}_A$:

$$\mathbf{F}_A = -\frac{\partial}{\partial \mathbf{r}_A} E_{tot} = \sum_{J} \left[\mathbf{F}_A^{J^+} - \frac{1}{2} \sum_{K \in \mathscr{J}} \mathbf{F}_A^{JK^-} \right] - \frac{\partial}{\partial \mathbf{r}_A} \Delta E_{Coul} \qquad (20)$$

Here $\mathbf{F}_A^M$ corresponds to the force on atom A as obtained by considering the fragment M. These forces are available from the corresponding single-point DFT calculations on the fragment M. Since when considering a particular fragment M it is placed in a Coulomb field of all surrounding atomic charges, the total energy of this fragment would depend on positions $\{\mathbf{R}_A\}$ of *all* atoms of the entire system including even those which are outside the fragment. Therefore, there is a contribution to the force $\mathbf{F}_A^M$ from any fragment M for all atoms, both inside and outside the fragment; the latter corresponds to the forces on the point charges surrounding the fragment as they are normally available in most of the quantum-chemistry codes. For a periodic system forces on the same atom within different unit cells are to be summed up.

Concerning the last term in the expression for the force (20) which is due to the Coulomb correction, the corresponding contribution to the force $-\frac{\partial}{\partial \mathbf{r}_A} \Delta E_{Coul}$ can be calculated analytically by differentiating the expression in Eq. (19). This is facilitated by the fact that the Coulomb interaction in the correction is modelled by point charges on atoms.

The formulae written above are general in a sense that their validity is independent of the topology of the system. It is a simple exercise to consider now a system formed by a periodic repetition of a set of several (irreducible) fragments. In this case only energies of irreducible fragments should be calculated as the energies of periodically repeated fragments

are identical. Consequently, the total energy becomes proportional to the number of unit cells, and hence one can calculate the total energy per cell. As an example, let us now apply Eqs. (17) and (19) to the periodic system of a copolymer introduced above in Section 2.1. A unit cell $\mathscr{U}$ of this system consist of two regions namely J and K. The energy per unit cell for this system is given by

$$E_{tot} = [E_{J^+} + E_{K^+} - E_{KJ^-} - E_{\tilde{J}K^-}] + \Delta E_{Coul} \quad (21)$$

where a rather lengthy Coulomb correction has the following form:

$$
\begin{aligned}
\Delta E_{Coul} &= -\frac{1}{2}C_{\mathscr{U},\mathscr{S}\setminus\mathscr{U}} + C_{J,J_K^0} + C_{J,J_{\tilde{K}}^0} + C_{K,K_J^0} + C_{K,K_{\tilde{J}}^0} \\
&\quad + C_{J_K,K_J} + C_{J_{\tilde{K}},\bar{K}_J} - C_{J_K,K_J^0} - C_{K_J,J_K^0} - C_{J_{\tilde{K}},\bar{K}^0_J} \\
&\quad - C_{\bar{K}_J,J_{\tilde{K}}^0} - C_{J_K,J_{\tilde{K}}} - C_{K_J,K_{\tilde{J}}} + C_{J_K,J_{\tilde{K}}^0} + C_{J_{\tilde{K}},J_K^0} \\
&\quad + C_{K_J,K_{\tilde{J}}^0} + C_{K_{\tilde{J}},K_J^0} \quad (22)
\end{aligned}
$$

The first term is the Coulomb interaction between the unit cell and the remaining infinite system calculated using the "exact" density, while the others denote interaction between individual parts similarly to Eq. (19). Note that the Coulomb correction inevitably contains caps which are situated in neighbouring unit cells. The corresponding electron density needed for the Coulomb correction and for calculating the Madelung field acting on fragments in the unit cell has to be taken from the central unit cell due to periodicity of the density. The periodic system formulated is a rather simplified example. In section 4.1 our model will be extended to accommodate more complex systems like MOFs.

Formulae (17), (19) and (20) represent the main result of this Section. They generalise the result obtained previously [39] in several directions: (i) a general formula is derived for division of an arbitrary system, however complex; (ii) the system can be either periodic or not; (iii) each fragment experiences the Coulomb potential from the rest of the system which allows treating a wider class of systems. In fact it can easily be shown that in the case of a partition into three regions and without external potential acting on each fragment, expression (19) reduces exactly to the formula derived earlier [39].

To finish the theory section, we shall briefly discuss a possible relationship between orbitals localisation and our method. Although this point is not trivial and requires a proper consideration which goes beyond the scope of this paper, we thought that a brief discussion of it may be beneficial to the reader. For our method to be valid, it is required by Eq. (1) in particular, that the electronic density in the core of each positive fragment be close to the exact density within the same region of space as calculated without fragmentation. This condition can be satisfied, at least approximately, if the core part of the fragment electron density can be adequately represented by

some localised orbitals obtained from the Kohn-Sham orbitals of the *entire system,* i.e. when no fragmentation is made. The two sets of orbitals are related by some unitary transformation (see, e.g. [49–51] and references therein). Other one-electron orbitals of the entire system which come out of the localisation procedure would only have quickly decaying tails in that region. In that case, if the positive fragment is chosen to be large enough to include the main part of the localised orbitals, one may assume that very similar orbitals can also be obtained from consideration of the fragment itself. Then the density in its core where these orbitals are mostly localised would be correctly given by the fragment. Note that different localisation criteria exist, however, in many cases very similar localisation of orbitals is obtained [49–51].

Likewise, we also require, as stated by Eq. (2), that the density within the caps in the negative fragments be close to the density of the same caps when calculated by considering the corresponding positive fragments. Again, a sufficient condition for this to happen is to choose the caps in such a way as to ensure that localised orbitals representing the caps in both the negative and positive fragments are very close to each other.

Based on the above considerations, the following guidelines for the construction of different fragments from a partition of a system into regions can be suggested. A minimum size positive fragment has to contain all localised orbitals of the entire system having a considerable overlap with the corresponding core region of that fragment. This implies that all localised orbitals which have a considerable overlap with two adjacent regions have to be covered by the corresponding negative fragment. Although we do not follow this line of thought in this paper, relying mostly on our intuition and experience in choosing the regions and fragments, we believe that the procedure outlined above and based on localised orbitals may be a useful avenue to be pursued in the future.

3 Self-consistency and implementation

In MFCC the energies E_{J^+}, E_{K^+}, E_{JK^-}, etc. entering the final energy expression in Eq. (17) are obtained in individual DFT calculations on the fragments J^+, K^+, JK^-, etc. Each of these fragments is considered in an "external" Coulomb field. This external field is represented by point charges on the atomic positions of surrounding regions. There will be detailed discussion on the geometry of these point charges in Section 4.3. Since the charges on atoms are not known *a priori*, these have to be calculated via a self-consistency procedure. The self-consistency with respect to atomic charges should be performed for each geometry before calculating the total energy and atomic forces. Alternativey, one may determine atomic charges by calculating isolated positive fragments at the beginning of the calculation and then keeping them unchanged during geometry optimisation. Although the latter procedure

is indeed approximate, the former one looks more appropriate as it takes account of the fact that the charges would change if atoms move. However, as will be discussed below, in the current implementation this method leads to some noise in atomic forces.

We implemented our method in a manager code similarly to our previous approach[39], but with the additional ability to handle surrounding point charges. The code constructs the positive and negative fragments for a given partition of the entire system, calls a quantum chemistry code to perform calculations on each individual subsystem and finally calculates the forces on atoms and the total energy. The DFT calculations of each fragment are performed using the quantum chemistry code GAMESS-UK[52]. Note that the GAMESS-UK performs only a single-point calculation on each individual fragment, yielding the total energy, atomic charges and the forces on atoms when necessary (see below), geometry optimisation is performed by our manager code.
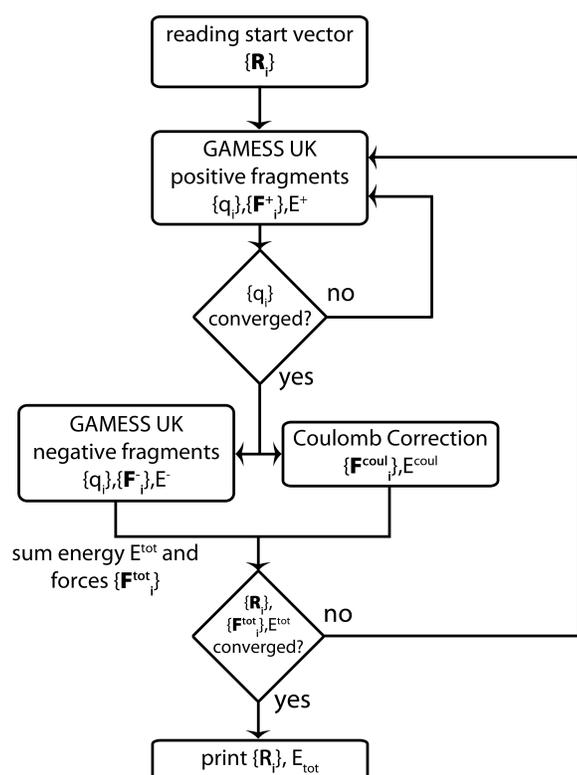
Fig. 4 A flow chart of the calculation.

The flow chart of the full procedure is shown in Fig. 4. The calculation starts with a set of particular atomic positions $\{\mathbf{R}_i\}$ of all atoms of the entire system. Initially all atomic charges are set to zero and hence no Coulomb field is considered within the first calculation of the "positive" fragments.

Once all "positive" fragments have been computed individually, atomic charges $\{q_i^0\}$ are obtained from atoms of the original region within each fragment. While the total charge of each fragment is zero by a constraint within GAMESS-UK, this does not hold for parts of the fragments, i.e. of the internal regions which make up the full system. To avoid artificial effects caused by a charged system the charges on atoms are slightly modified prior to further calculation so that the total charge of the unit cell be zero. The required shift of the atomic charges is typically of the order of $0.001e$. Subsequently all "positive" fragments are calculated again, this time in the Coulomb field of the surrounding system, and this is continued until convergence with respect to the charges on all atoms of the system is reached (the smaller loop in Fig. 4), i.e. when charges on atoms change by no more than $0.01e$. At the next step all "negative" fragments are considered in the Coulomb field of all atoms surrounding them using the charges obtained at the previous calculation. Note that "negative" fragments are considered only once at each geometry.

After this set of calculations is finished, the charges on atoms of the caps $\{q_i^1\}$ are available, the total energy is computed in accordance with Eqs. (17) and (19), while the forces on atoms $\{\mathbf{F}_i\}$ are obtained from Eq. (20). All atoms are moved into a new geometry $\{\mathbf{R}_i\}$ using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, and the calculation is returned back to the inner loop to obtain the new charges on atoms, see Fig. 4. This time the charges obtained in the previous geometry offer an initial guess and form the external field for the first set of calculations. Normally we find that the change in the atomic charges is so small that the inner loop (the convergence with respect to atomic charges) needs to be performed just once.

It is important to note that the forces as implemented here lack one contribution. The forces on atoms of each fragment are calculated according to the Hellman-Feynman theorem and by an analytic differentiation of the correction term. However there is a peculiar contribution to atomic forces (somewhat similar in spirit to the Pulay forces[53]), which is due to the fact that the point charges surrounding the fragments are updated for each geometry and thus these charges implicitly depend on the atomic positions[13]. In our implementation of the fragmentation CC-MFCC scheme this contribution to the atomic forces is not included. As was pointed out recently in Ref.[13] it is not only being neglected in a wide range of methods, it is even rarely discussed. Calculations on the particular system to be discussed in the forthcoming Section seem to imply that there are good reasons to believe that this missing contribution to atomic forces is not significant (but noticable), at least for the MOF system we consider. Note however that this contribution may be larger for other types of systems and hence in the future no effort should be spared in working out the corresponding missing contribution to the forces due to

self-consistency of atomic charges.

4 Application

4.1 Fragmentation of MOF-16
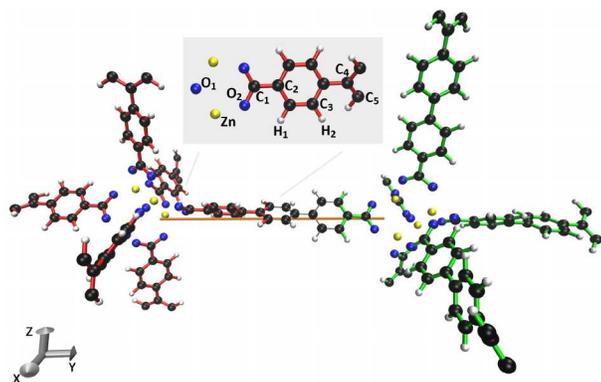
Fig. 5 Two unit cells of MOF-16 shown in two ways: either as a symmetric unit cell using red bonds between atoms (at the left half of the figure), or as used in fragmentation via green bonds (at the right half of the figure). The vertex clusters consist of 4 zinc atoms (yellow) and 13 oxygens (blue) and are connected with each other by three benzol rings (carbons: black; hydrogens: white). Inset: a subset of the irreducible set of atoms within the unit cell shown in greater detail with some atoms labeled. The electrostatic potential presented in Fig. 8 was calculated along the orange line. The experimentally found lattice constant is 21.4903 Å [48].

In this section the fragmentation of MOF-16 as well as the corresponding energy correction will be presented. We confine ourself to this specific system since it serves as an excellent testbed for illustrating the implementation of the theory and its performance; the partition topology and correspondingly the correction should however be generally applicable to a number of other MOFs and zeolite systems. MOF-16 is made up of a metal organic cluster containing a central oxygen atom surrounded by four zinc atoms and another twelve oxygen atoms as shown in Fig. 5. Each pair of neighbouring clusters is linked by a chain of three benzol rings. Each cluster is coordinated to six neighbouring clusters as illustrated in Fig. 6(a). Therefore the topology of the system is the one of a cubic grid as shown schematically in Fig. 6(b) and will be explained in more detail below. We shall also choose the directions of the Cartesian axes along the linkers. Experimentally [54] determined Bravis lattice is simple cubic. Disorder was found [54] in the particular orientation of the subgroups (i.e. the central cluster, benzol rings) within the network. One orientation can be turned into the other by plane reflections. Note that the space group $Pm\bar{3}m$ was obtained in [54] assuming equal probabilities for all orientations. We chose one particular ori-

entation for the whole system as shown in Fig. 5 breaking the symmetry of $Pm\bar{3}m$.
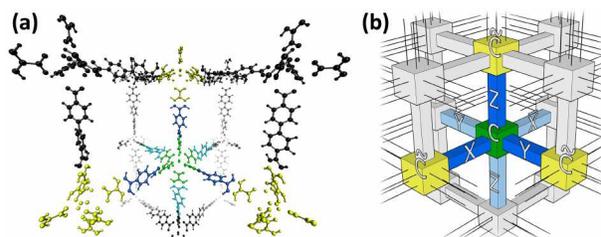
Fig. 6 Partition of the MOF-16 shown using (a) balls and sticks model and (b) schematically. The unit cell is shown consisting of a centre cluster (coloured in green) and three linkers (blue); neighbouring regions are coloured in light blue (linkers) and yellow (centre cluster).

In total there are 113 atoms in the unit cell. The natural symmetric unit cell of MOF-16 following the point group symmetry of the system is represented by red coloured bonds in Fig. 5 and corresponds to a cut through the central benzol ring of the linker. To apply our fragmentation method to this system, it is convenient to use a different unit cell represented by green coloured bonds in Fig. 5. This corresponds to a cut through the first and the third benzol rings. In the following we will be referred to these two unit cells representations as the symmetric and asymmetric ones, respectively.

As shown in Fig. 6(a) these cuts simultaneously divide the unit cell into four regions schematically represented by cuboids in Fig. 6(b). The green coloured centre region C consist of zinc, oxygen and half of a benzol ring on each of the six sides along the Cartesian axes. Blue coloured are the three organic linker regions X,Y,Z named by their direction in space. Together they exactly make up the zero unit cell $U = [C,X,Y,Z]$ indicated by the green bonds in Fig.5.

The light blue coloured linkers $\bar{X},\bar{Y},\bar{Z}$ in Fig. 6 are the three other linkers neighbouring the same centre region C; they belong to different unit cells. An overbar in the names of the linkers is used to indicate that they point in the negative x, y, z directions. Similarly each of the linker regions X,Y,Z borders another centre region, which belong to neighbouring unit cells and are yellow coloured in Fig. 6(b). These three centre regions situated in the positive x, y, z directions, respectively, are denoted $\tilde{C}$. One such region is shown on the right in Fig. 7(a).

Having chosen the regions, we can now construct the irreducible fragments: both positive and negative ones. This is done in the following way. Consider a central region C shown on the left in Fig. 7(a). The benzol ring of each linker to this region was cut into two halves during the partition; therefore, we construct the central fragment C^+ by capping the region C from each of its six sides by the other half of the benzol rings,
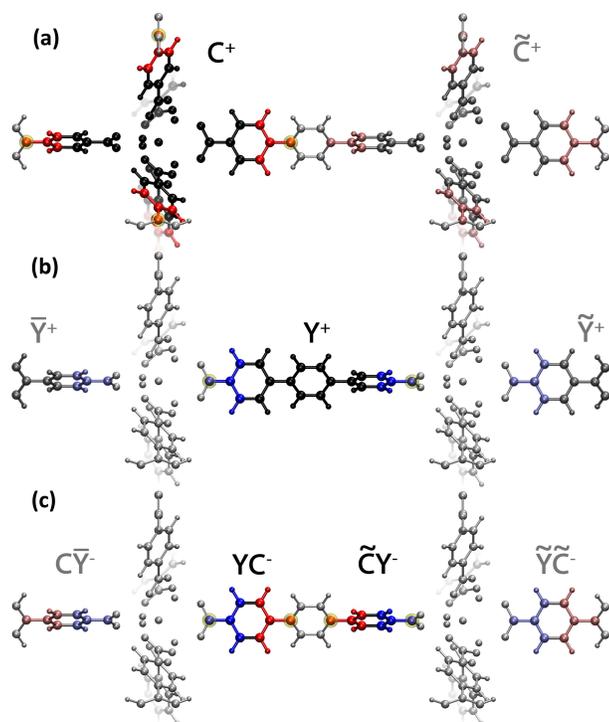
Fig. 7 Fragmentation of MOF-16: (a) centre region C (black) and its caps C_K (red) forming the centre fragment C^+ (where $K = X, Y, Z, \bar{X}, \bar{Y}, \bar{Z}$); (b) linker region Y (black) and its caps Y_C and $Y_{\tilde{C}}$ (blue) forming the linker fragment Y^+; (c) there are two distinct negative fragments YC^- (blue/red) and $\tilde{C}Y^-$ (red/blue). Instances of the same fragments in neighbouring unit cells in the y direction are coloured in with lower opacity.

these are red coloured in Fig. 7(a). Electrons are known to be delocalised over the carbon atoms of a benzol ring. The minimum positive fragment (according to Section 3) thus has to include the three carbon atoms of the benzol ring complementing the ones within the central region. For each benzol ring two additional hydrogens are added at their natural positions (e.g H_2 and its upper counterpart in Fig. 5) and one hydrogen is added replacing a carbon as a link atom (e.g. C_4 in Fig. 5). The link atoms are indicated by a yellow transparent sphere in Fig. 7(a,b). Correspondingly, a positive linker fragment Y^+ is constructed as shown in Fig. 7(b), where the caps (representing half of benzol rings) are shown blue coloured. All caps are terminated by a link atom indicated by a yellow transparent sphere in Fig. 7(a,b). So, this way we build a central fragment C^+ and a link fragment Y^+ which is positioned with respect to the central fragment along the positive direction of the y axis. There will be of course other linker fragments connected to the same central fragment: X^+ and Z^+ which run along positive directions of the other two Cartesian axes, and $\bar{X}^+, \bar{Y}^+, \bar{Z}^+$ which run along their negative directions. As was

noted above, the fragments $\bar{X}^+, \bar{Y}^+, \bar{Z}^+$ correspond however to different unit cells.

It follows from this partition that neighbouring regions share exactly one benzol ring. These, including the corresponding link atoms, make up exactly two negative fragments along each linker. For instance, if we consider the Y linker, then we arrive at the negative fragments YC^- and $\tilde{C}Y^-$ which are shown in Fig. 7(c). Note that like in the schematic presented earlier in Fig. 1, combining the black regions of all positive fragments one obtains the full system, cf. Fig. 7(a,b). Also each cap in the positive fragments has exactly one counterpart within the negative fragments. In Fig. 7 only regions in the y direction are shown for clarity and also to allow for a direct comparison with the 1D scheme presented in Section 2.1.

Similar to the case of a periodic 1D copolymer of Section 2.1, we also have fragments which are translational copies of the irreducible ones belonging to a single unit cell U. In a fragmentation scheme the topology is important rather then the system dimension. The difference with the 1D system of a copolymer is that in 3D there are three different directions in which periodicity is implied instead of one (hence three different linkers) and the centre region having six neighbours instead of two. To identify various regions and fragments, we extend the notations introduced earlier to our 3D system. Consider fragment C^+ in the zero unit cell. Neighbouring fragments along the positive direction of the Cartesian axes will be denoted with the tilde as $\tilde{C}^+$, while equivalent neighbouring to C^+ fragments along the negative directions of the axes with the bar, as $\bar{C}^+$. Correspondingly, caps of regions will be denoted by the region symbol with the subscript indicating the neighbouring region which has the cap as its part. For instance, the two caps applied to region Y and shown as blue in Fig. 7(b) are denoted Y_C (the left one) and $Y_{\tilde{C}}$ (the right one). Unlike the case of the one dimensional system of Section 2.1, in the 3D case there are three different directions (x, y, z) in which a region can be translated corresponding to the three linkers orientations; however, there should be no confusion as it should be clear from the context in each case which linker and the corresponding direction are involved.

Note that our system has a rather high symmetry; in particular, the fragments X^+, Y^+ and Z^+ are related by rotations. However, the symmetry was not exploited here to leave the scheme as general as possible.

The energy per unit cell and the corresponding Coulomb correction can be derived in the same spirit as for the 1D periodic system. Similarly to Eqs. (21) and (22), the energy per unit cell is now given by

$$E_{tot} = E_{C^+} + \sum_{K \in \{X, Y, Z\}} [E_{K^+} - E_{KC^-} - E_{\tilde{C}K^-}] + \Delta E_{Coul} \quad (23)$$

where C^+ is a central fragment (e.g. the one in the zero unit cell U) and the sum is run over its three linkers belonging to the same unit cell. Correspondingly, the energy correction (19) is:

$$
\begin{aligned}
\Delta E_{Coul} = & -\frac{1}{2}\left[C_{U,S/U} + \sum_{J \in \mathscr{U}} \sum_{K \in \mathscr{U}/J} C_{J,K} \right] \\
& + \sum_{K \in \mathscr{U}} \left[C_{C,C_K^0} + C_{C_K,K_C} - C_{C_K,K_C^0} - C_{K_C,C_K^0} \right. \\
& \left. - \sum_{L \in \mathscr{U}/K} \left(\frac{1}{2} C_{C_K,C_L} - C_{C_K,C_L^0} \right) \right] + \sum_{J \in \{X,Y,Z\}} \left[C_{J,J_C^0} \right. \\
& \left. + C_{J,J_{\bar{C}}^0} - C_{J_C,J_{\bar{C}}} + C_{J_C,J_{\bar{C}}^0} + C_{J_{\bar{C}},C_J^0} \right]
\end{aligned} \tag{24}
$$

where $C_{U,S/U}$ describes the Coulomb interaction between the zero unit cell U and the rest of the system S/U (here S is used for the whole system), the sum over J runs over all regions within U, $K \in U/J$ means all regions in U excluding J, and so on. Therefore, here the first term corresponds to the Coulomb interaction of each fragment with its surrounding; the second term contain summations over all regions K and L ($L \neq K$) within one (zero) unit cell and the third term contain summation exclusively over the three linkers.

It is important to note at this point that the particular partition used is very similar to O.M. Yaghi's scheme of a straight forward design of MOFs[54]. O.M. Yaghi showed that there is a wide range of different MOF which can be built out of a limited number of building blocks, e.g. the centre clusters and the linker. Our fragmentation method follows the underlying idea of such building blocks and can easily be adopted to match different topologies.

4.2 Computational details

In our implementation of the CC-MFCC method link atoms replace three-coordinated C atoms and are situated at their respective positions. Therefore, the force on such atoms comes from both contributions, when they are real atoms in some fragments and link atoms in another. In our calculations we either used hydrogen atoms or one-electron H-like pseudoatoms as link atoms. The latter was found[39] to be appropriate for our method, but were replaced in some calculations by hydrogen for comparison and performance reasons. In all our calculations described below we used the PBE density functionals and the basis set 6-31G[55] for all atoms apart from the H-like pseudo atoms, which were set up according to Ref.[56]. No periodic boundary conditions were applied when calculating individual fragments. The core electrons for all species were treated explicitly. The geometry optimisation was run until the forces on atoms were smaller than 0.1 eV/Å.

To verify results of our calculations within the proposed MFCC and CC-MFCC techniques against standard periodic DFT calculations we employed the CRYSTAL code[57]. This code is capable of performing calculations using the same Gaussian-type basis set and the same density functional and therefore allows for the direct verification of our method. In our CRYSTAL calculations we employed the same 6-31G basis set and the PBE density functional. We used a unit cell centred at the oxygen atom located in the centre of the metal cluster (the "symmetrical" unit cell in the left half of Fig. 5). The sums of Coulomb and exchange integrals were truncated following overlap thresholds of 10^{-8}, 10^{-8}, 10^{-8}, 10^{-8} and 10^{-16}, respectively (see[57]). Integration over **k**-points in the reciprocal space was performed using a $4 \times 4 \times 4$ Monkhorst-Pack grid[58]. The self-consistent procedure for electron density was performed until changes in the total energy per unit cell became less than 10^{-7} Hartree. We used internal redundant coordinates for geometry optimisation. Both the unit cell and the positions of atoms were optimised. The geometry optimisation was considered completed when simultaneously four thresholds were reached: the largest absolute (la) and root mean square (rms) values of atomic forces became smaller than $4.5 \cdot 10^{-4}$ and $3.0 \cdot 10^{-4}$ a. u. and la and rms values of atomic displacements became smaller than $1.8 \cdot 10^{-3}$ and $1.2 \cdot 10^{-3}$ a. u., respectively.

4.3 Geometry of the Point Charge Cloud

For the calculation of the different fragments the Coulomb field of the remaining regions is represented by point charges at respective positions of atoms. For an infinite system, this Coulomb field can be e.g. calculated using the Ewald method[59], however, due to the size of the system we believe that this method is not very efficient. Instead, a finite summation of the atoms surrounding the given region in question was implemented using a number of methods. In the following two different methods for defining the cutoff for selecting charges around regions are presented. To evaluate the convergence with respect to the cutoff, different selections of point charges around the fragment Y were made, and in each case the Coulomb potential was calculated along the straight line close to the y axis as shown by the orange line in Fig. 5. The results for the potential are presented in Fig. 8, while the force on a particular atom of the same linker calculated by the GAMESS-UK using the corresponding point charge cloud is depicted in Fig. 9.

Two models for choosing the point charge cloud were tried. In the simple sphere method all point charges within a certain cutoff radius from the centre of the fragment in question are included in the point charge cloud. In this case it is impossible to guarantee that the charge cloud is neutral as not all charges from each unit cell are included around the bor-
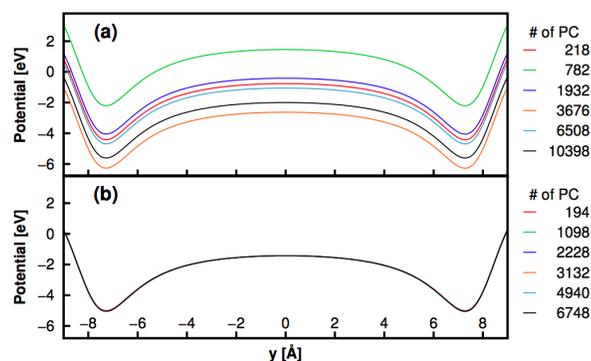
Fig. 8 The Coulomb potential along the linker Y shown by the straight orange line in Fig. 5 using (a) the simple sphere and (b) the symmetric unit cell scheme for the point charge cloud. Each curve corresponds to a particular selection of charges; the total number of charges included in each case is shown on the right of the graphs. Note that in (b) different curves are indistinguishable from each other.

Fig. 9 Force on atom H_1 (see Fig. 5) in Y^+ in the y direction for different point charge geometries and sizes as a function of the number of point charges included in the cloud: the red line corresponds to the sphere method, while the green and blue lines correspond to unit cell method using symmetric (green) and asymmetric choices (blue).

der of the sphere; in fact the charge of the cloud depends on the sphere radius. As a consequence, this choice of the point charge cloud results in a poor convergence of both the potential in Fig. 8(a) and of the force on the atom shown in Fig. 9(red line). In the second method the point charge cloud was build of complete unit cells using the distance of their centre to that of the fragment as the cutoff distance. It turned out that it is crucial to use the symmetric unit cells for that; these are indicated by red bonds in Fig. 5. The potential curves for different number of point charges shown in Fig. 8(b) all lie on top of each other and the atomic force in Fig. 9 shown by the green line converge very quickly with the number of point charges included in the cloud in this case. For the choice of the asymmetric unit cells (shown by green bonds in Fig. 5) the force in Fig. 8 converges to a different value compared to the other schemes due to a dipole moment associated with this choice of the unit cell.

The asymmetric zero unit cell overlaps with four different symmetric unit cells: the green coloured centre cluster and three neighbouring unit cells having a yellow coloured centre cluster as shown in Fig. 6. In the calculations presented in the following each fragment was first embedded into the point charges of these four symmetric unit cells. Subsequently a necessary number of symmetric unit cells were added to this point charge cloud if their centre was within a certain distance from the centre of the zero unit cell. This scheme ensures that the point charge clouds for all fragments to be equal, which is important for the corresponding terms in Eq. (18) to cancel out exactly. In practice a point charge cloud of about 6000 charges was used.

5 Results

We performed full geometry relaxation calculation of the MOF-16 system using the fragmentation scheme described in Sec. 4 including the optimisation of the unit cell. The fragmentation calculation is compared with that performed using CRYSTAL code where the whole system was considered without any fragmentation and using an identical basis set. Calculations using hydrogens with its standard basis as link atoms turned out to perform better compared to H-like pseudo atoms with respect to charge and geometry agreement with the results of CRYSTAL (data not shown). Correspondingly simple hydrogens were used as link atoms in all the calculations presented in the following. We shall start our discussion by looking at the charges on atoms.

5.1 Charges

The scheme described in Section 3 contains an internal SCF loop where the charges on the atoms are obtained iteratively; correspondingly, it is relevant to discuss the convergence of these charges first. To illustrate their convergence we describe a single point calculation at the geometry obtained by CRYSTAL after full geometry optimisation was performed. We find that normally charges on atoms are established after the first iteration, only a very small change is observed at the following iterations: after the second iteration the charges were typically changed by no more than $0.05e$, while after three iterations they change by $0.01e$ at most and the convergence of the atomic charges is assumed.

In Table 1 we show the results for two types of calculations: for the Mulliken charges obtained within the scheme in which point charges outside each fragments were taken into account, and (in brackets) from the calculation in which this has not been done. Note that in the latter case no SCF loop is run to converge the charges. At the same time, the charges

Atom	q_{CRY}	$q_{Frag}^{0} - q_{CRY}$	$q_{Frag}^{1(pos)} - q_{CRY}$	$q_{Frag}^{1(pos)} - q_{Frag}^{1(neg)}$
Zn	0.84	0.05 (0.04)	-	-
O_1	−0.88	−0.05 (−0.06)	-	-
O_2	−0.54	−0.01 (−0.02)	-	-
C_1	0.51	0.01 (−0.00)	−0.09 (−0.43)	−0.00 (−0.00)
C_2	0.06	−0.04 (−0.05)	−0.17 (−0.11)	0.02 (0.02)
C_3	−0.15	0.03 (0.03)	−0.05 (−0.04)	−0.02 (−0.03)
C_4	0.03	0.01 (0.00)	0.20 (0.06)	0.07 (0.01)
C_5	−0.15	−0.02 (−0.02)	-	-
H_1	0.18	0.00 (0.00)	−0.06 (−0.04)	0.00 (−0.00)
H_2	0.15	−0.04 (−0.01)	−0.01 (0.00)	0.06 (0.00)

Table 1 Mulliken charges on selected atoms (see Fig. 5) after convergence was reached. The results of the calculation without point charges are shown in brackets. Second column: charges, q_{CRY}, as calculated by CRYSTAL and used as a reference. In the third column charges within positive fragments as obtained within the fragmentation scheme, q_{Frag}^{0}, are compared with CRYSTAL values. Fourth column: charges on caps atoms as calculated in positive fragments are compared with the CRYSTAL charges. Fifth column: charges on atoms in the caps within the positive and the negative fragments are compared.

are compared with those obtained by CRYSTAL which are listed in the second column and used as a reference. One can see that the deviation of the charges computed with the fragmentation scheme from the ones calculated by CRYSTAL is typically less then $0.04e$ with the the atoms in the metal complex Zn and O_1 differing most. There is no big difference in the atomic charges calculated with and without point charges surrounding the fragments in this case. We see that either fragmentation scheme (with and without point charges outside the fragments) results in reasonable agreement with the atomic charges obtained by CRYSTAL when no fragmentation was done.

An assumption expressed in Eq. (2) and used in the derivation of the Coulomb correction (19) is that the charge density of the caps in the positive fragments is close to the one of the negative fragments. The difference shown in the fifth column of Table 1 illustrates this point. Deviations of the Mulliken charges on the selected atoms of the caps are typically smaller then $0.03e$ with the exception of H_2 and C_4 which have a difference of $0.06e$ and $0.07e$, respectively. For other atoms in the caps not shown here the deviations are much smaller. Therefore, we can conclude that the current fragmentation scheme is adequate to satisfy the condition of Eq. (2). Furthermore our correction term (19) discriminate between the electron density of the actual system (superscript 0) and the one of the caps (superscript 1). The non negligible differences shown in the fourth column of Table 1 support this to be well-founded. The latter also shows that for the CC-MFCC

calculation with point charges surrounding the caps the maximum deviation in atomic charges (and hence, in their electron density) as compared to the actual ones is reduced from $0.43e$ without point charges to $0.20e$ including them.

5.2 Geometry relaxation

Geometry relaxation of the fragmented system was performed by our manager code following the CC-MFCC approach as described in Sec. 3 including point charges mapping the environment of fragments and the corresponding Coulomb correction term. A relaxation calculation following the MFCC method in which no point charges were used surrounding the fragments and without any energy correction was also performed. We used the experimental structure by O.M. Yaghi[48] as initial geometry in our calculations. Hydrogen atoms were added to this structure at guessed positions, following usual C-H distances in organic molecules. Within 30 iterations changes in energy went down below 0.03 eV and the largest atomic force down to 0.1 $eV/Å$ in both cases. Further convergence was difficult for the CC-MFCC calculation due to noise in the atomic forces.

Relaxation of the lattice vectors is not straightforward within the fragmentation method, as it requires construction of a stress tensor from the calculations performed on different fragments (which are considered as "molecules"). However, in the case of the MOF-16 system with the simple cubic lattice the optimal lattice constant can be obtained trivially by relaxing the atomic geometry for cubic unit cells of different sizes (with no other restrictions applied) and then interpolating the energy. In Fig. 10 the corresponding calculations are presented. Several values of the lattice constant were considered, parabolas were fitted to the corresponding energies and the corresponding minima for the two fragmentation schemes (MFCC and CC-MFCC) and the one of the CRYSTAL calculation were obtained as indicated by vertical lines. Note that because of the efficiency considerations the CRYSTAL calculations were performed using the space group $P23$, which is a subgroup of the group $Pm\bar{3}m$ still corresponding to the simple cubic lattice and preserving the required tetrahedral symmetry of the central clusters.

A relaxation of the lattice vectors by CRYSTAL resulted in a lattice constant of 21.79 Å; this is 1.7% larger then the experimental value of 21.49 Å[48]. The energy corresponding to the fully relaxed CRYSTAL calculation is depicted by the filled black square in Fig. 10. The results of the MFCC and CC-MFCC calculations are shown by red and green circles, respectively. One can see that both values of the lattice constant obtained by either of the fragmentation schemes (MFCC and CC-MFCC) in the minima of the parabolas are within 0.2% of the CRYSTAL value. While MFCC method (without point charges and correction) slightly underestimates the system en-
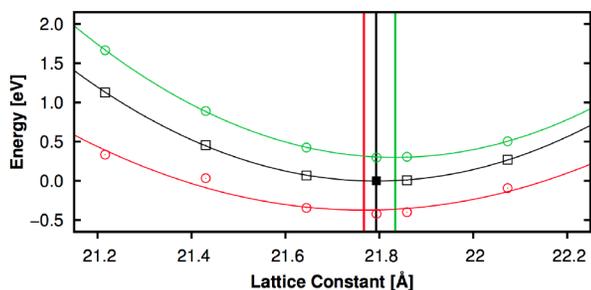
Fig. 10 Relative total energies (per unit cell) of the MOF-16 after geometry relaxation for different values of the lattice constants for the CC-MFCC (with point charges and correction, green circles) and MFCC (without point charges, red circles) methods. A single filled black square corresponds to the CRYSTAL calculations with full geometry relaxation using the fixed space group $P23$ (the lattice constant of 21.79 Å). Open square symbols correspond to CRYSTAL data obtained at a set of fixed lattice constants, still keeping the same space group.

ergy (by 0.42 eV), the CC-MFCC scheme slightly overestimates it by 0.29 eV. The Coulomb correction calculated within the CC-MFCC scheme is relatively small and equal to, 1.22 eV.

All structures obtained using CRYSTAL, MFCC and CC-MFCC differ little from the experimental structure in the size of the unit cell. A detailed comparison with experiment is however hampered by the fact that there is an uncertainty in the orientation of the subgroups of the system mentioned earlier; in fact, the actual structure is disordered as the orientation of the central clusters and of the rings may change across the network. At the same time, what has been calculated here corresponds to a structure in which a certain *single* orientation was chosen. In the MFCC and CC-MFCC calculations the precise orientation of the benzol rings adjacent to the central clusters, see Fig. 5, deviates from measured experimentally (assuming only one of the two possibilities) by $\simeq 10°$ and $\simeq 7°$, respectively. Note also that the relaxed geometries obtained in both simulations correspond at least to the space group $P23$. This was also the group used in our CRYSTAL calculations, for which the described orientation deviated from the experimental one by $8°$. Nevertheless overall the experimental structure is reproduced in all our calculations extremely well.

We shall now compare the relaxed structures obtained using the fragmentation methods with that found by CRYSTAL. We first note that visually the three structures are extremely close: they are virtually indistinguishable if overlaid with each other. To compare the structures quantitatively, bond lengths, angles and dihedral angles of all groups of neighbouring atoms of the periodic system were calculated, and the CC-MFCC and MFCC values were compared with the CRYSTAL ones. For each atom A within the unit cell, the bond length to each

neighbouring atom B was calculated. The angle between A, B and each neighbouring atom C of B (not being A), as well as the dihedral of A, B, C and each neighbouring atom D of C (not being A or B) were also calculated.
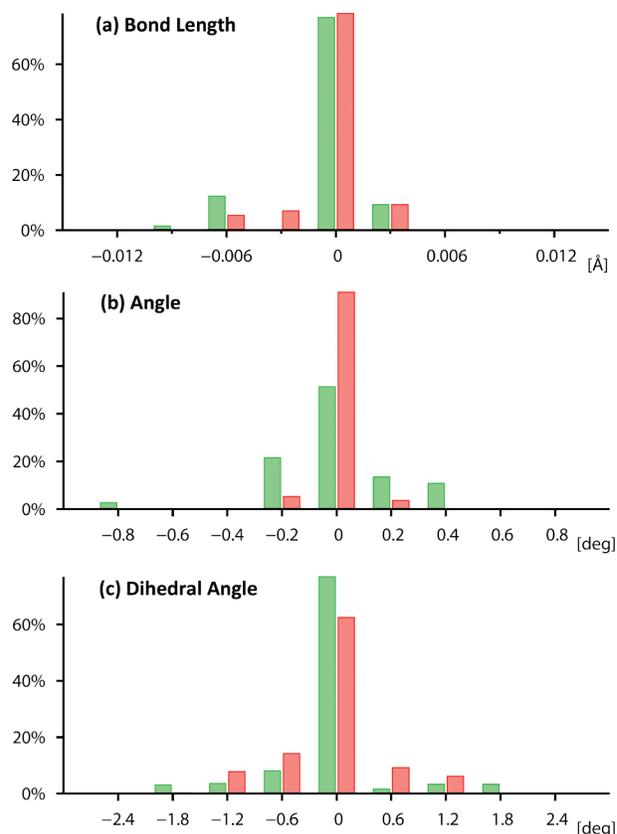
Fig. 11 Histograms showing deviations of bond lengths (a), angles (b) and dihedral angles (c) of all groups of neighbouring atoms (see text for definitions) obtained using MFCC (red bars) and CC-MFCC (green) methods from the values obtained by CRYSTAL.

In Fig. 11 (a) deviations of the bonds lengths obtained using CRYSTAL and both fragmentation methods are shown. Typically the bonds differ by less then 0.003 Å. Also the bonds involving links atoms (carbons replaced by hydrogens) in some fragments are close, e.g. the $C_1 - C_2$ bond length (Fig. 5) obtained by the CC-MFCC method is practically identical, within the given precision, to the CRYSTAL value. This is remarkable as a $C - H$ bond has a typical length of 1.09 Å compared to the $C_1 - C_2$ bond having a length of 1.48 Å. The negative fragments seem to work well to prevent the shortening of these bonds. In Fig. 11 (b) and (c) similar histograms for the angles and the dihedrals are depicted, respectively. Typical derivations of the angles are smaller then $0.5°$ and of the dihedrals smaller then $1°$. In this sense optimised geometries obtained by both fragmentation schemes, MFCC and

CC-MFCC, show very small deviations from the one obtained by CRYSTAL. For the three properties presented the MFCC method performs slightly better then the CC-MFCC method. The maximum deviations of a bond length / angle / dihedral is 0.005Å / 0.21° / 1.5° for MFCC compared to 0.008Å / 0.82° / 1.7° for CC-MFCC. The largest deviations obtained with the latter method can be found for atoms which are in close proximity to the point charges cloud in certain fragments, e.g. the first and the third benzol rings which also make up the caps and the negative fragments. We want to stress that differences between both geometries are still extremely small.

Concluding, the lattice constant and atomic positions obtained using either of the two incarnations of the fragmentation method (MFCC and CC-MFCC) come out extremely close to the values obtained using the CRYSTAL code when no fragments were considered. The total energies (per cell) in all cases are also very close with the differences of less than 0.0002% from each other. This test calculation shows without doubt a high precision one can achieve by employing the fragmentation technique.

5.3 Adsorption of a hydrogen molecule

In this subsection we shall describe our preliminary results of an application of the MFCC method to an important problem of hydrogen storage[41]. Mueller *et al.*[60] considered several positions of the hydrogen molecule inside the skeleton of the MOF-5 system, and their relative adsorption (formation) energies were calculated to be in the order of tens of meV. To our knowledge, no such calculations have been performed to date for the MOF-16 system with much longer ligands. The magnitude of the hydrogen formation (adsorption) energies must not be too small and too large at the same time to enable both storage and release of the H_2 molecules from the MOF.

Since the results for the MOF-16 were found to be very close with both MFCC and CC-MFCC methods, we have used the former method in the calculations described here. The calculations were performed in the following way. We considered several positions of the molecule with respect to the MOF vertex which were found having well separated adsorption energies in Ref.[60]. In each case full geometry relaxation was performed using the fixed unit cell lattice constant of 21.794Å as found using the MFCC method for the MOF-16 itself. Then, the adsorption energy was calculated via

$$\Delta E_{ads} = E\left(MOF + H_2\right) - E\left(MOF\right) - E\left(H_2\right) + E_{BSSE} \ ,$$

where $E\left(MOF + H_2\right)$ is the total (per cell) energy of the MOF-16 with the hydrogen molecule, $E(MOF)$ and $E\left(H_2\right)$ are the total DFT energies of the MOF-16 (per cell) and the hydrogen molecules relaxed individual, while E_{BSSE} is the basis set superposition error (BSSE) which was calculated using

Position[60]	Coordinates	MOF-16		MOF-5[60]
		ΔE_{ads}	E_{BSSE}	ΔE_{ads}
I-A	(3.17, 3.17, -3.21) (3.60, 3.61, -3.64)	-15.8	20.0	-21.7
II-A	(2.98, 2.98, 2.98) (2.55, 2.55, 2.54)	-3.6	17.5	-9.5

Table 2 Binding energies (in meV) of a hydrogen molecule to the MOF-16 system at several positions (the first column) next to the central fragment as calculated using the MFCC method. Coordinates (in) of the two hydrogens are given relative to the central oxygen. For comparison, the results of plane wave calculations[60] at similar positions for the MOF-5 system are also given. We also show the BSSE energies in each case.

the counterpoise correction method[61]. In this calculation energy differences are to be considered between identical systems (MOF-16 and the H_2 molecule) calculated with different basis sets. Since in all geometries we studied the hydrogen molecule was next to the central fragment, the BSSE calculations were considerably simplified by considering only the central fragment instead of the whole MOF-16.

Coordinates of the relaxed hydrogen molecule shown in Table 2 were found to be within 0.1 Å of those in the reference positions[60]. The resulting energies together with the corresponding BSSE corrections are shown in Table 2. We see that in all cases the adsorption is energetically favourable. We also note that the BSSE correction is significant. We also show in the same Table for comparison the corresponding results calculated using the plane wave pseudopotential DFT method (using PBE, as in our case) for the MOF-5 system. We see that our calculations give adsorption energies very close to those obtained for MOF-5 in Ref.[60]. Moreover, the order of the adsorption sites for the hydrogen molecule with respect to the binding energies is also preserved. We note that the differences between the MOF-5 and MOF-16 values may not only be due to the fact that the two systems are in fact different (the MOF-16 has longer ligands); the two computational techniques are also rather different. Indeed, in our calculations we used a localised basis set instead of the plane waves, and also we did not use the pseudopotential method necessary for performing plane wave DFT calculations as used by[60]. Still, our calculations demonstrate that the fragmentation approach may be quite useful for not only considering the MOF systems themselves, but also for studying chemical reactions of small molecules with them.

6 Discussion and conclusions

In this paper we provide a theoretical foundation of the MFCC method within local and semilocal DFT as applied to periodic

systems. Within this scheme a single *ab initio* local basis set DFT calculation on a composite system is replaced by a set of individual calculations on subsystems capped by parts of adjacent subsystems and treated as "molecules". "Negative" fragments are considered to compensate for artificial contributions of the caps. Provided that the caps map the environment of every subsystem sufficiently well and long range Coulomb interactions are accounted for or small, we demonstrate that one should be able to obtain the total energy, atomic forces and the charge density for local and semilocal DFT flavours with sufficient precision. There is also a correction term derived rigorously within the theory provided which is of purely Coulomb nature.

Our method is capable of calculating the total energy and atomic forces, which is the *most expensive* part of the calculation for almost every system. Thus atomic relaxation and/or molecular dynamics calculation is supported directly. To obtain a (continuous) electron density, in contrast, an additional step is needed. This can be done by collecting the densities from core regions of different fragments similarly to the way it is done in e.g. the ADMA method[11]. Alternatively, a continuous density ρ_{tot} for the entire system can be obtained by summing up electron densities ρ_{J^+} of all positive fragments and subtracting all the densities ρ_{JK^-} of the negative ones:

$$\rho_{tot} = \sum_J \left(\rho_{J^+} - \frac{1}{2} \sum_{K \in \mathscr{J}} \rho_{JK^-} \right) \qquad (25)$$

The same expression has been proposed in[24]. Provided that the conditions (1) and (2) are satisfied, this expression is formally exact; in practice, it should provide us with a reasonable approximation for the density. Either way, knowledge of the density may only be required at the end of the calculations for the analysis. Also, a continuous density is directly available for various parts of the system within the cores of its fragments. This might be sufficient in many cases; since divisions of the system into regions is usually made through non-essential bonds, the most important and interesting chemistry that occurs in the regions is within an easy reach in our method if required. The adsorption of a hydrogen molecule on the MOF-16 system considered above is an example of such a situation since the redistribution of the electron density in the region where the molecule bonds to the MOF is fully provided by our method. Note, however, that the large contribution of the BSSE correction indicates that in this particular case for accurate results a larger basis set would be desireable.

The theory has been tested for one Metal Organic Framework (MOF) system, namely for MOF-16. We calculated the lattice constant, positions of atoms and the Mulliken charges on them, and compared our results with those obtained using the CRYSTAL code when no division into fragments was performed. In both cases all electron calculations were performed

(i.e. we did not use pseudopotentials) and identical basis sets were used. We find that atomic charges are well reproduced as compared with the CRYSTAL calculation, indicating that one obtains a correct electron density using our fragmentation method. Also the total energy comes out very close to the CRYSTAL value with an absolute error being smaller then 0.3 eV which for such a big system corresponds to a relative error of about 0.0002% only. Using a simple quadratic fit the optimal lattice constant for the unit cell was found to be very close to the CRYSTAL value with an error less then 0.2%. It is remarkable that the geometries obtained with our method are almost indistinguishable with those obtained in the CRYSTAL calculation. It can be concluded that the MFCC method may be a feasible choice when it comes to the prediction of structures of porous systems like the MOF investigated here.

We considered two versions of the method: MFCC which neither includes a point charge cloud around each fragment nor the Coulomb correction term, and CC-MFCC which does include both. Both methods are based on an assumption that charge densities within corresponding caps as obtained by considered positive and negative fragments are very close to each other, and our calculations confirm that this is indeed the case. It was also shown that a Coulomb field of point charges can improve the agreement of the charge density for the caps to the actual (CRYSTAL) ones. Although CC-MFCC and MFCC methods give very similar results which are very close (both in the total energy and the atomic geometry) to those obtained by CRYSTAL, a slight improvement for the dihedral angles was found when the CC-MFCC method was used. Overall we did not observed significant effect due to inclusion of the point charge cloud around each fragment in a self-consistent manner and our Coulomb correction term in the calculations.

Considering the contribution to the total energy from the Coulomb correction, the latter is found very small indeed in the MOF-16 calculations comparing with the energy itself. This might be the reason for the MFCC and CC-MFCC methods giving nearly identical results. However, it may not be the case for other systems where division into fragments (or regions) will have to be done in such a way that fragments are charged. Apparently, on average in the case of the MOF-16 the selected regions are nearly neutral. In some applications, however, it might be more convenient allowing the total charge of each fragment to fluctuate. In these cases the Coulomb correction may actually make a substantial difference to the energy and the forces on atoms of the system.

A number of approximations which have been employed in this study for the sake of simplicity deserve further consideration and improvement.

First of all, the Coulomb field was approximated by point charges. This can be well justified for parts of the system being far away from the fragment. However, for the charge

density in close proximity of the caps this can be a rather crude approximation, which can be improved by taking the actual charge density or its multipole expansion. Furthermore, the Coulomb field around the fragments was represented by charges within a finite radius around the fragment only. This may influence the results in spite of the fact that we showed that the electrostatic potential due to the point charge cloud converges well along a specific direction within the cell. Other schemes (e.g. based on the Ewald method) may be more preferable. As the Coulomb field was approximated by point charges, so was the energy correction term. Correspondingly there are the same limitations as the accuracy of the correction term is concerned. While most terms in Eqs. (19) correspond to Coulomb interaction between well separated regions, there are several terms, however, which correspond to the Coulomb interaction between adjacent regions. Thus, an improvement is possible based on using the appropriate electron densities or their multipole expansion instead of point charges. In the current work Mulliken charges were used for obtaining point charges. In comparison, natural charges[62] have been reported to perform better in the context of fragmentation schemes[12,37], however strongly pronounced improvement has only been reported using distributed mulipoles[32]. Using higher multipole moments of the atoms electron density instead of the continuous density may be the method of choice to improve the results without compromising much the efficiency of the calculations, and this can be worth considering for future development.

As pointed out by Richard and Herbert[13] the gradients used in this work (as well as in a number of other earlier publications) are approximate because of a missing contribution due to updated point charges. In the calculations using the CC-MFCC method the convergence of the forces was limited to 0.1 eV/Å. To investigate the magnitude of the observed noise in the forces the forces on three atoms (O_2, H_1 and C_3, see Fig. 5) where calculated numerically from the system energy using a finite difference method and compared with the "analytical" CC-MFCC forces on the same atoms. In these test calculations a geometry which is somewhat off the CC-MFCC relaxed geometry was used. The mean error in the analytic forces within the sample chosen was 0.02 eV/Å, with the largest error found being 0.3 eV/Å. Using the same terms for the energy and the forces, but with fixed charges no force differed by more then 0.02 eV/Å between the two calculations. This indicates that some error in atomic forces within the current implementation of the CC-MFCC method is indeed to be expected due to the updates of the point charges. Note that the error found here is of similar magnitude to the one found in the context of the FMO method, where a maximum deviation in the forces of about 0.5 eV/Å was reported for the $(ALA)^{10}$ system[63].

In our opinion, there are three ways of how one could deal with this issue: (i) use fixed charges; (ii) use biased gradients and (iii) use more sophisticated methods where derivatives of

atomic charges with respect to atomic positions are calculated. Fixed charges for the calculation of periodic systems have been obtained from an idependent calculation beforehand in Ref.[64] using periodic boundary conditions. However, this appriach cannot be considerd as entirely from first-principles. In this work we have chosen the second option, because we expect that updating charges is essential and it is much easier to implement than the third. Of course, this results in a noise in the forces and hence may pose some difficulties in reaching fully converged atomic geometries and system energies. Note, however, that our calculations demonstrated that although a noticable error was found, this effect did not prevent us from converging to the required geometry with an acceptable precision, at least for the systems studied here. Nevertheless, we believe that work should be directed in our fragmentation method towards deriving improved atomic forces where account is taken of the fact that atomic charges are being updated during the iteration procedure, as it was done in the context of XPol[65] and FMO[63] methods.

Although in this work we have used PBE density functional, other density functionals may be used within CC-MFCC approach as well. One has to bear in mind that the total energy (17) as derived above is based on using local (LDA[66]) and semilocal (e.g. PW1[67] or PBE[68]) density functionals. Methods containing intra-atomic corrections, such as DFT+U[69], can be used with the approach developed in the present work because these corrections can be considered as local in Eq. (3). The division of the total energy into local contributions and a contribution from pair interactions between different regions is the key condition for implementation of the developed fragmentation technique. Therefore, if the total energy of a system can be written in a similar way, then the developed technique can be straightforwardly generalised and applied. In principle, the developed method does not offer a sound theoretical ground for the usage of functionals which contain nonlocal exchange contributions, such as e.g. B3LYP (Becke, three-parameter, Lee-Yang-Parr)[70] as well as for van der Waals systems with the nonlocal density functional[71–73]. Still, the method may work reasonably well for these functionals as demonstrated in our earlier work[39]. This would be the case if the nonlocal contributions in the functional are small across fragments boundaries and thus Eq. (17) may still give a good approximation to the total energy.

In our formulation of the method we followed the original ideas of MFCC. However, the novel concept of differentiating between the densities within the caps and in the central core of the fragments may be applicable to a number of more sophisticated fragmentation schemes as well. Correspondingly the correction term derived here should be transferable to other methods based on overlapping fragments, such as e.g. the GEBF method[12], the method due to Collins and Deev[27] based on different level of fragmentation, as well as methods

implementing the many-body expansion of overlapping fragments[13,34].

Within this work the partition of the system was based on chemical intuition. We believe that a rationale for dividing the given system into regions and fragments may be developed based on orbital localisation methods (e.g.[49–51] and references therein). Although this cannot be a practical method as the whole idea of the fragmentation approach is to *avoid* consideration of the entire system, still we believe that this direction of research may be useful in gaining necessary experience in performing fragmentation of various systems and hence is worth pursuing in the future.

Summarising, in this paper we have demonstrated that our fragmentation method can be used for considering complex systems. The simulation of the hydrogen molecule adsorption on the MOF-16 is one possible example of such application. The main advantage of using our method is its potential efficiency: since DFT methods scale nonlinearly with the system size, considering several smaller systems can be more efficient than treating the entire system at once at the same level of theory. Although at present, the partitioning scheme is implemented as a serial code, the scheme can be easily parallelized. Finally, the partitioning method can be of potential use in QM/MM calculations[74–83] as a way of dividing the whole system into quantum and classical parts while partially eliminating the effect of the terminating (link) atoms.

7 Acknowledgment

8 References

References

1 S. Kitagawa, R. Kitaura and S. Noro, *Angew. Chem., Int. Ed.*, 2004, **43**, 2334–2375.

2 S. R. Gadre, R. N. Shirsat and A. C. Limaye, *J. Phys. Chem.*, 1994, **98**, 9165–9169.

3 T.-S. Lee, J. P. Lewis and W. Yang, *Comp. Mater. Sci.*, 1998, **12**, 259–277.

4 S. Goedecker, *Rev. Mod. Phys.*, 1999, **71**, 1085–1123.

5 D. Bowler, T. Miyazaki and M. Gillan, *J. Phys.: Condens. Matter*, 2002, **14**, 2781.

6 J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón and D. Sánchez-Portal, *J. Phys.: Condens. Matter*, 2002, **14**, 2745–2779.

7 C.-K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *J. Chem. Phys.*, 2005, **122**, 084119.

8 D. R. Bowler, J.-L. Fattebert, M. J. Gillan, P. D. Haynes and C.-K. Skylaris, *J. Phys.: Condens. Matter*, 2008, **20**, 290301.

9 M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev*, 2012, **112**, 632–672.

10 W. Yang, *Phys. Rev. Lett.*, 1991, **66**, 1438.

11 T. E. Exner and P. G. Mezey, *J. Comput. Chem.*, 2003, **24**, 1980–1986.

12 W. Li, S. Li and Y. Jiang, *J. Phys. Chem. A*, 2007, **111**, 2193–2199.

13 R. M. Richard and J. M. Herbert, *J. Chem. Phys.*, 2012, **137**, 064113.

14 K. Kitaura, E. Ikeo, T. Asada, T. Nakano and M. Uebayasi, *Chem. Phys. Lett.*, 1999, **313**, 701–706.

15 D. G. Fedorov and K. Kitaura, *J. Chem. Phys.*, 2004, **121**, 2483–2490.

16 D. G. Fedorov and K. Kitaura, *Chem. Phys. Lett.*, 2004, **389**, 129–134.

17 W. Li, T. Fang and S. Li, *J. Chem. Phys.*, 2006, **124**, 154102.

18 D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2003, **119**, 3599–3605.

19 T. E. Exner and P. G. Mezey, *J. Phys. Chem. A*, 2004, **108**, 4301–4309.

20 S. Li, W. Li and T. Fang, *J. Am. Chem. Soc.*, 2005, **127**, 7215–7226.

21 X. Chen, Y. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 2005, **122**, 184105.

22 Y. Mei, D. W. Zhang and J. Z. H. Zhang, *J. Phys. Chem. A*, 2005, **109**, 2–5.

23 W. Li and S. Li, *J. Chem. Phys.*, 2005, **122**, 194109.

24 X. He and J. Z. H. Zhang, *J. Chem. Phys.*, 2004, **122**, 031103.

25 M. A. Addicoat and M. A. Collins, *J. Chem. Phys.*, 2009, **131**, 104103.

26 H.-A. Le, H.-J. Tan, J. F. Ouyang and R. P. A. Bettens, *J. Chem. Theory Comput.*, 2012, **8**, 469–478.

27 V. A. Deev and M. A. Collins, *J. Chem. Phys.*, 2005, **122**, 154102.

28 X. He and J. Z. H. Zhang, *J. Chem. Phys.*, 2006, **124**, 184703.

29 M. A. Collins and V. A. Deev, *J. Chem. Phys.*, 2006, **125**, 104104.

30 R. P. A. Bettens and A. M. Lee, *J. Phys. Chem. A*, 2006, **110**, 8777–8785.

31 H. M. Netzloff and M. A. Collins, *J. Chem. Phys.*, 2007, **127**, 134113.

32 H.-A. Le, A. M. Lee and R. P. A. Bettens, *J. Phys. Chem. A*, 2009, **113**, 10527–10533.

33 X. Wang, J. Liu, J. Z. H. Zhang and X. He, *J. Phys. Chem. A*, 2013, **117**, 7149–7161.

34 N. J. Mayhall and K. Raghavachari, *J. Chem. Theory Comput.*, 2012, **8**, 2669–2675.

35 V. Ganesh, R. K. Dongare, P. Balanarayan and S. R. Gadre, *J. Chem. Phys.*, 2006, **125**, 104109.

36 S. Hua, W. Hua and S. Li, *J. Phys. Chem. A*, 2010, **114**, 8126–8134.

37 N. Jiang, J. Ma and Y. Jiang, *J. Chem. Phys.*, 2006, **124**, 114112.

38 Y. Xiang, D. W. Zhang and J. Z. H. Zhang, *J. Comput. Chem.*, 2004, **25**, 1431–1437.

39 J. He, C. Di Paola and L. Kantorovich, *J. Chem. Phys.*, 2009, **130**, 144104–144104.

40 P. Zhang, D. G. Truhlar and J. Gao, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7821.

41 L. J. Murray, M. Dincă and J. R. Long, *Chem. Soc. Rev.*, 2009, **38**, 1294–1314.

42 N. L. Rosi, J. Eckert, M. Eddaoudi, D. T. Vodak, J. Kim, M. O'Keeffe and O. M. Yaghi, *Science*, 2003, **300**, 1127–1129.

43 B. Liu, X. Zhang, H. Shioyama, T. Mukai, T. Sakai and Q. Xu, *J. Power Sources*, 2010, **195**, 857–861.

44 C.-D. Wu, A. Hu, L. Zhang and W. Lin, *J. Am. Chem. Soc.*, 2005, **127**, 8940–8941.

45 P. Horcajada, C. Serre, G. Maurin, N. Ramsahye, F. Balas, M. Vallet-Regí, M. Sebban, F. Taulelle and G. Férey, *J. Am. Chem. Soc.*, 2008, **130**, 6774–6780.

46 D. G. Fedorov, J. H. Jensen, R. C. Deka and K. Kitaura, *J. Phys. Chem. A*, 2008, **112**, 11808–11816.

47 L. Roskop, D. G. Fedorov and M. S. Gordon, *Mol. Phys.*, 2013, **111**, 1622–1629.

48 M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe and O. M. Yaghi, *Science*, 2002, **295**, 469–472.

49 O. Danyliv and L. Kantorovich, *Phys. Rev. B*, 2004, **70**, 075113.

50 O. Danyliv and L. Kantorovich, *J. Phys.: Condens. Matter*, 2004, **16**, 7233–7246.

51 O. Danyliv, L. Kantorovich and F. Corá, *Phys. Rev. B*, 2007, **76**, 045107.

52 M. F. Guest, I. J. Bush, H. J. J. van Dam, P. Sherwood, J. M. H. Thomas, J. H. van Lenthe, R. W. A. Havenith and J. Kendrick, *Mol. Phys.*, 2005, **103**, 719.

53 P. Pulay, *Mol. Phys.*, 1969, **17**, 197–204.

54 O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature*, 2003, **423**, 705–714.

55 K. L. Schuchardt, B. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045.

56 G. A. DiLabio, M. M. Hurley and P. A. Christiansen, *J. Chem. Phys.*, 2002, **116**, 9578.

57 R. Dovesi, V. Saunders, C. Roetti, R. Orlando, C. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison and I. J. Bush, *Uni. of Torino, Italy*, 2009.

58 H. J. Monkhorst and J. D. Pack, *Phys. Rev. B*, 1976, **13**, 5188–5192.

59 L. Kantorovich, *Quantum Theory of the Solid State: An Introduction*, Kluwer, 2004.

60 T. Mueller and G. Ceder, *J. Phys. Chem. B*, 2005, **109**, 17974–17983.

61 S. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.

62 A. E. Reed, R. B. Weinstock and F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735–746.

63 T. Nagata, D. G. Fedorov and K. Kitaura, *Chem. Phys. Lett.*, 2009, **475**, 124–131.

64 H. Fukunaga, D. G. Fedorov, M. Chiba, K. Nii and K. Kitaura, *J. Phys. Chem. A*, 2008, **112**, 10887–10894.

65 W. Xie, L. Song, D. G. Truhlar and J. Gao, *J. Chem. Phys.*, 2008, **128**, 234108.

66 D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.*, 1980, **45**, 566–569.

67 J. P. Perdew and Y. Wang, *Phys. Rev. B*, 1992, **45**, 13244.

68 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

69 V. I. Anisimov, F. Aryasetiawan and A. I. Liechtenstein, *J. Phys.: Condens. Matter*, 1997, **9**, 767.

70 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.

71 D. C. Langreth, M. Dion, H. Rydberg, E. Schröder, P. Hyldgaard and B. I. Lundqvist, *Int. J. Quant. Chem.*, 2005, **101**, 599–610.

72 T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard and D. C. Langreth, *Phys. Rev. B*, 2007, **76**, 125112.

73 M. Dion, H. Rydberg, E. Schröder, D. C. Langreth and B. I. Lundqvist, *Phys. Rev. Lett.*, 2004, **92**, 246401.

74 H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185.

75 Y. Zhang, H. Lin and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 1378.

76 D. M. Philipp and R. A. Friesner, *J. Comput. Chem.*, 1999, **20**, 1468.

77 D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580.

78 U. Eichler, C. M. Kölmel and J. Sauer, *J. Comput. Chem.*, 1997, **18**, 463.

79 F. Maseras and K. Morokuma, *J. Comp. Chem.*, 1995, **16**, 1170.

80 T. Vreven and K. Morokuma, *J. Comp. Chem.*, 2000, **21**, 1419–1432.

81 R. J. Hall, S. A. Hinde, N. A. Burton and I. H. Hillier, *J. Comp. Chem.*, 2000, **21**, 1433–1441.

82 V. Nasluzov, V. Rivanenkov, A. Gordienko, K. Neyman, U. Birkenheuer and R. N., *J. Chem. Phys.*, 2001, **115**, 8157–8171.

83 Y. Zhang, *Theor. Chem. Acc.*, 2006, **116**, 43.