Pseudo-fragment approach for extended systems derived from linear-scaling DFT

Laura E Ratcliff and Luigi Genovese

1 Department of Materials, Imperial College London, London SW7 2AZ, United Kingdom
2 Argonne Leadership Computing Facility, Argonne National Laboratory, IL 60439, United States of America
3 University Grenoble Alpes, CEA, INAC-MEM, L_Sim, F-38000, Grenoble, France

E-mail: laura.ratcliff08@imperial.ac.uk

Received 17 January 2019, revised 29 March 2019
Accepted for publication 5 April 2019
Published 30 April 2019

Abstract
We present a computational approach which is tailored for reducing the complexity of the description of extended systems at the density functional theory level. We define a recipe for generating a set of localized basis functions which are optimized either for the accurate description of pristine, bulk-like Wannier functions, or for the in situ treatment of deformations induced by defective constituents such as boundaries or impurities. Our method enables one to identify the regions of an extended system which require dedicated optimization of the Kohn–Sham degrees of freedom, and provides the user with a reliable estimation of the errors—if any—induced by the locality of the approach. Such a method facilitates on the one hand an effective reduction of the computational degrees of freedom needed to simulate systems at the nanoscale, while in turn providing a description that can be straightforwardly put in relation to effective models, like tight binding Hamiltonians. We present our methodology with SiC nanotube-like cages as a test bed. Nonetheless, the wavelet-based method employed in this paper makes possible calculation of systems with different dimensionalities, including slabs and fully periodic systems.

Keywords: linear scaling density functional theory, fragment approach, nanotubes

(Some figures may appear in colour only in the online journal)

1. Introduction

Density-functional theory (DFT) [1, 2] is arguably the most popular method for electronic structure calculations, and is routinely applied to a range of materials and properties for systems containing up to a few hundred atoms. In recent years, the development of linear scaling (LS) approaches [3, 4], which overcome the computational limits of traditional cubic scaling implementations, has enabled the treatment of thousands of atoms from first principles. Such LS methods open up the possibility not just of treating larger systems, but also new types of materials and calculations [5].

One popular route to LS involves the creation of a minimal set of localized support functions (SFs), which are expressed in an underlying systematic basis set and optimized in situ to reflect their chemical environment. Importantly, this approach, which has been adopted in ONETEP [6], CONQUEST [7] and BtoDFT [8, 9], preserves the same high level of accuracy as standard cubic scaling approaches employing systematic basis sets, such as plane waves or wavelets. This brings accurate DFT calculations of very large systems into reach, although the prefactor associated with the SF optimization can be high.

In tandem with the development of LS-DFT, various fragment and embedding methods have also been developed to reduce the computational cost of treating large systems using DFT, in some cases in conjunction with other quantum mechanical methods. Such approaches have in common the division of a target system into two or more subsystems and are generally applied in one of two ways. In the case of fragment methods, this is primarily as a ‘divide and conquer’ approach to reducing the cost of treating a system at a single level of theory, along the same lines as Yang’s LS-DFT approach of the same name [10]. In other words, all fragments are treated on equal footing, e.g. [11–14]. Embedding methods, on the other
hand, are generally used as a means of implementing a multi-

scale approach, wherein one fragment (the active region) is
treated using a higher level of theory, while the remainder of
the system (the environment) is treated with a computa-
tionally less expensive method, e.g. [15–23]. The active region
is typically treated using a higher order quantum chemistry
method, hybrid functional DFT and/or a larger basis set, while
the environment is treated with for example semi-local DFT
and/or a smaller basis set.

The accuracy of an embedding calculation depends on both
the choice of method for treating the interactions between
subsystems and the way in which the system is divided. For
molecular systems, the fragments are usually the constituent
molecules of a supramolecular system, or a larger molecule
and surrounding solvent molecules. Along these lines, we
and others have recently implemented a molecular fragment
approach [24] in the wavelet-based BigDFT code [25]. This
approach takes advantage of similarity between molecules
with similar local environments, by using SFs generated for
isolated ‘template’ fragments, in this case molecules, and
using them as a fixed basis for a large system of many such
molecules. The fragmentation is thus at the level of the basis,
with the fragments fully interacting in the final calculation.
The approach is well suited to materials which consist of
clearly separated fragments. We have therefore recently for-
malized such a fragmentation concept [26], showing that a
suitable inspection of density matrix related quantities of a
full LS calculation may be used to determine whether or not
a fragment can be seen as a separable ‘moiety’ of the whole
system.

In extended systems, however, there is often no clear
fragmentation scheme, so that the assignment of an atom to
a particular subsystem is somewhat arbitrary. Furthermore,
depending on the method used to account for the interactions,
both the accuracy and computational cost can be highly sensi-
tive to the chosen fragmentation. Nonetheless, to our knowl-
edge no scheme exists to define a sensible fragmentation a
priori. As such, a means of predicting whether or not a given
fragmentation choice is reasonable would be highly useful.
To this end, in the following we explore the advantages of using
optimized SFs from LS-DFT both to set up a fragment-like
approach and as a means of gaining insight into the suitability
of a given fragmentation of an extended system.

Given that the SFs in LS-DFT are optimized to reflect their
local environment, if two SFs associated with different atoms
are the same, this implies that their environments are also the
same. For an infinitely repeating system this should be the
case, to within some level of numerical noise coming from
the representation in a systematic but nonetheless incomplete
basis. However, where there is a defect or deformation, SFs
close to the defect would differ significantly from those far
away. If one could quantify the similarity between two SFs,
it would thus be possible to assess the physical extent of the
defect, i.e. the distance over which it has a significant effect on
the SFs and thus the electronic structure.

To show how such a quantity might be used in practice,
we introduce in this work a pseudo-fragment (PFrag) method
based on LS-DFT which is able to treat extended systems
containing some degree of repetition, e.g. a periodic system
comprising many repeat units. Such an approach might easily
be defined if, instead of taking an isolated template frag-
ment, one generates the SF basis by embedding the PFrag
in a representative environment. As with the molecular fragment
approach, this can result in considerable computational sav-
ings. More interestingly it also provides the opportunity to
explore in detail the errors introduced by different levels of
approximation.

Using the PFrag approach, we show how the similarity
between SFs located on atoms with similar chemical environ-
ments relates to the accuracy of a given PFrag setup. That is,
the SF similarity provides a means of determining to what
extent a given fragmentation of an extended system is rea-
sonable, i.e. whether or not an atom should best be associ-
ated with an active region where further SF optimization is
needed or an environment region where the template SFs are
already able to accurately represent the electronic structure.
Furthermore, such a measure can also be used to give a reli-
able estimation of the errors induced by the imposed pseudo-
fragmentation. For the sake of conciseness in this work we
focus on the case study of a one-dimensional system, in
this case a nanotube, however we stress that the approach
could easily be adapted to other extended systems. Similarly,
although we have implemented the approach in BigDFT, it
could also be implemented in other LS-DFT codes which
generate optimized SFs.

The layout of this manuscript is as follows. In section 2 we
recap the approach to LS-DFT implemented in BigDFT and
explain the key concepts of the molecular fragment approach.
In section 3 we then present the embedded PFrag approach
as applied to a SiC nanotube (NT). We consider both an in-
initely repeating NT and a finite NT, quantifying the similarity
between SFs on equivalent atoms and demonstrating the appli-
cability of template SFs generated in a short NT for a much
longer NT. In section 4 we then consider the effect of intro-
ducing deformations to the system, first in the form of random
noise in the atomic coordinates, and then by considering finite
nanotubes of varying lengths. Finally, in section 5, we present
a summary and outlook on future applications of our method.

2. Background theory: molecular fragments in
BigDFT

2.1. Linear scaling BigDFT

Thanks to the nearsightedness principle, in the \(\mathcal{O}(N)\)
formalism implemented in the BigDFT code, we assume that the
density matrix of the system \(\hat{F}\) can be defined from a set of
localized SFs \(\phi_\alpha\) as follows:

\[
\hat{F} = \sum_{\alpha, \beta} \phi_\alpha K^{\alpha\beta} \phi_\beta,
\]

with an SF overlap matrix \(S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle\), which can be
chosen to have a unit diagonal and where \(K^{\alpha\beta}\) is the so-called
density kernel. In the LS-BigDFT approach, the SFs are
expressed in a Daubechies wavelet basis and are defined such as
to optimize the target function

\[
\hat{F} = \sum_{\alpha, \beta} |\phi_\alpha\rangle K^{\alpha\beta} |\phi_\beta\rangle.
\]
\( \Omega = \text{tr} \left( \hat{F} \hat{H}_c \right) \)

where the operator \( \hat{H}_c = \hat{H}\text{KS}[\rho] + \hat{V}_c \) is the sum of the density-dependent Kohn–Sham (KS) Hamiltonian plus a confining operator \( \hat{V}_c \) such that

\[
\langle \phi_\alpha | \hat{V}_c | \phi_\beta \rangle = \delta_{\alpha\beta} \langle \phi_\alpha | \hat{V}_c^n | \phi_\alpha \rangle,
\]

\[
V_c^n (\mathbf{r}) = c_\alpha |\mathbf{r} - \mathbf{R}_\alpha|^4,
\]

which has the purpose of keeping the SFs confined in their localization regions, centered in the position \( \mathbf{R}_\alpha \), while reducing the KS band structure energy. Usually \( \mathbf{R}_\alpha \) coincides with the position \( \mathbf{R}_\alpha \) of the atom \( \alpha \) where \( \phi_\alpha \) is initially centered at the beginning of the SCF optimization procedure. To some extent this enables one to associate \( \alpha \) to a particular atom \( \alpha \). The coefficient \( c_\alpha \) is dynamically adjusted during the basis set optimization procedure. For a molecular calculation, we obtain in this way a minimal set of molecular orbitals that, by construction, exactly represent the occupied KS orbitals, and are thus referred to as e.g. s-like SFs. For a given SF basis, the density kernel is obtained choosing a set of molecular orbitals that, by construction, exactly represent the occupied KS orbitals, and are thus referred to as e.g. s-like SFs. 

2.2. Molecular fragment approach

When simulating a system that is made from a collection of various molecules it would appear reasonable to associate to each of these a ‘fragment’ whose electronic structure might be represented by a subset of the SFs of the entire system. Such an approach has been already presented in [24, 27] and further formalized in [26], where a recipe has been provided to identify such fragments out of the density matrix of the full system. The use of such a fragment approach substantially reduces the computational cost, and has been applied to the calculation of electronic charge transport parameters in a disordered supramolecular morphology [27].

The molecular fragment approach in BigDFT is conceptually straightforward, and can be summarized in three steps:

(i) **template calculation**: generate the SF basis \( \{ \phi_\alpha^T \} \) for each unique fragment type \( \tilde{\gamma} \) by performing a full LS calculation for an isolated fragment, optimizing both the SFs and density kernel.

(ii) **SF replication**: take the SFs from the isolated template calculations and replicate them for the full system of interest.

(iii) **full calculation**: perform a LS calculation for the full system using the replicated SFs as a fixed basis, optimizing only the density kernel.

Both the first and last steps use the standard machinery of a LS calculation. The second step can be easily automated, however, it is complicated by the need to account for a suitable set of rototranslation transformations. Namely, given the range of potential orientations and positions of fragments in a particular target system, it is essential to have a method which is able to rototranslate the template SFs from their template coordinates to the correct location and orientation in the full system of interest.

2.2.1. Rototranslation procedure.

For the case where there is no internal deformation of a given fragment, i.e. the transformation from template to system fragment is a rigid rotation, one can easily find the appropriate transformation. As such, we therefore apply the rotation matrix \( R^{T \rightarrow S} \) which minimizes the following cost function \( J \):

\[
J \left( R^{T \rightarrow S} \right) = \frac{1}{2} \sum_{a=1}^{N} | | R^T_a - \sum_{b=1}^{N} R^{T \rightarrow S}_{ab} R^S_b | |^2,
\]

where \( N \) is the number of atoms in the fragment and \( R^T_\alpha \) are the coordinates of the template (system) fragment. The identification of the optimal transformation from such a cost function is a well known problem [28, 29], and may easily be found using a simple singular value decomposition based approach [30]. Such a transformation may then be combined with the translation \( T^{T \rightarrow S} \) which is straightforwardly defined from the respective centre of mass shifts. Having found the appropriate rototranslation, we then use a wavelet-based interpolation scheme to perform the identified transformation, as explained in [24]. The computational cost of this interpolation step is kept low, as is the error introduced by this rototranslation.

To summarize, let us assume that a system is made of a collection of fragments \( \{ \phi^{(T)}_\alpha \} \), with the index \( \ell \) labelling the fragment type and the index \( i \) running over all the fragments of the same type. A set of fragment templates \( \{ \phi^{(T)}_\alpha \} \) can then be chosen and the associated SFs \( \{ \phi^{(T)}_\alpha \} \) extracted as per step (i) above. The rototranslation procedure guarantees that the calculation of step (iii) employs the basis set defined as \( \{ R^{T \rightarrow S}_{\alpha} \phi^{(T)}_\alpha \} \) where the rototranslation \( R^{T \rightarrow S}_{\alpha} \) connects the template fragment of type \( \ell \) with the \( i \)th moiety of the same type. Such a rototranslation minimizes by construction the cost function \( J \) of equation (5).

The approach described above is performed at the level of the SFs rather than the density operator, so that the fragments are still fully able to interact with each other via the full system density without the need to account for overlapping fragments. Nonetheless, the approach is designed for the creation of a molecular basis, which relies on two key assumptions. First, that the fragments are separable, i.e. that the density matrix can be reasonably approximated by a ‘block-like’ matrix, each block being associated to a fragment, and that the interaction between neighbouring fragments is weak enough to not significantly impact on the form of the SFs with respect to the template basis set. Second, that the fragments are close to rigid, i.e. that the internal deformations between different fragment instances remain small. For the latter, the cost function of equation (5) may be used to quantify the level
Figure 1. Panel (a) pristine SiCNT treated with periodic boundary conditions, comprising 14 NT units (392 atoms). Si (C) are atoms depicted in green (grey). Visualization (panel (b)) of select optimized \( s \) and \( p_z \) type SFs and corresponding \( S_{\alpha\beta} \) onsite values (panel (c)). Onsite overlap values have been calculated between each SF and that of the equivalent type (i.e. between \( s \) and \( s \)-type functions) and the equivalent atom in the centre of the NT (i.e. between atom \( i \) of fragment \( j \) and atom \( i \) of the central fragment), where the central region is shaded in grey. The values in the central region determine the error associated with the numerical noise of the reformatting procedure. For large values of \( J \), the internal deformations are too high and the fragment approximation breaks down.

3. Embedded fragment approach: case study of a SiC nanotube

The approach described so far is well-suited for molecular systems, or more precisely for systems where the fundamental constituents have molecular character. The key question is therefore whether it is possible to define an equivalent approach for extended systems, wherein a template basis is generated from a small representative system. In this section we focus on responding to this question.

We take the case study of a (quasi-)one dimensional system, namely a SiC nanotube. We begin with a pristine SiCNT in periodic boundary conditions, for which the structure is depicted in figure 1(a). We first assess the extent to which the SFs follow our expectations of similarity along the length of the NT (i.e. equivalent SFs for equivalent environments), and then demonstrate how to take advantage of such repetition in practice. In order to quantify the error resulting from the strict localization of the SFs, we also perform initial calculations using the cubic scaling approach of BigDFT, wherein the KS wavefunctions are calculated directly without any additional localization constraints. To this end, we take a long NT containing 14 repeat units, comprising 392 atoms in total; this is both short enough to be accessible to the cubic scaling approach of BigDFT and long enough to avoid the need for \( k \)-point sampling. Computational details are given in appendix A.

3.1. Assessing basis similarity: onsite overlap matrix

Before considering how to employ a fragment-like approach for the SiCNT, we first pose some questions on the nature of the optimized SFs. In this case we do not have the notion of ‘moiety’, as the system cannot be partitioned into distinctly separable fragments. However we expect that for a pristine SiCNT, the SFs in equivalent locations along the length of the NT should be quasi-identical, with only small differences due to numerical noise arising from the egg-box effect. Select optimized SFs resulting from a standard LS calculation are depicted in figure 1(b), where it can be seen that the selected SFs are indeed qualitatively the same. However, it would be useful to have a more reliable measure than visual inspection, and furthermore to have a quantitative measure of the degree of similarity.

To this end, we define a quantity which we refer to as the onsite overlap matrix, \( S_{\alpha\beta} \). This is defined as

\[
S_{\alpha\beta}^{\text{ onsite}} \equiv \langle \mathcal{R} \mathcal{T}^{R_\alpha \rightarrow R}_\alpha \phi_\alpha^{(t)} | \mathcal{R} \mathcal{T}^{R_\beta \rightarrow R}_\beta \phi_\beta^{(t)} \rangle, \tag{6}
\]

i.e. as the overlap between two SFs (which may or may not originally be centred on the same atom) if they were to be centred on the same site of centre \( R \). Where \( S_{\alpha\beta}^{\text{ onsite}} = 1 \), this implies SFs \( \alpha \) and \( \beta \) are identical, presumably due to having both the same character (\( s \), \( p_z \), etc) and the same local environment, irrespective of whether there is a spatial overlap. The onsite overlap is thus independent of the physical location of the SFs and is instead purely a measure of the degree of similarity between two SFs, and correspondingly the chemical environment of the atoms with which they are associated.

We calculate \( S_{\alpha\beta}^{\text{ onsite}} \) by first choosing a localization region associated with an arbitrary atom of position \( R \), then using the interpolation scheme discussed above to reformat all SFs into this region. The rotations—if needed—should be chosen such as to properly align the local environment of the atoms associated to the indices \( \alpha \) and \( \beta \). Following this, the full onsite overlap matrix can be calculated using the same machinery used to calculate the standard (spatial) SF overlap matrix. This process introduces some numerical noise, whose size depends on the employed grid spacing. We therefore do not expect to see exact agreement between two SFs in equivalent environments, however this discrepancy should be much smaller than the signal in which we are interested. Such an indicator should therefore help in answering the question of what extent the presence of a defect (e.g. a boundary or an impurity) at a given distance might result in differences between SFs.
s density matrix.

the fundamental constituents of the system

In some sense, it may be ascribed to a

procedure of

before the SFs can no longer be considered to be equivalent.

the values remain small throughout, in line with our obser-

vations that the SFs in different rings of the NT are indeed

in grey) should be zero, since they correspond to the overlap

in figure 1(c). The values in the central region (highlighted

Figure 2. Panel (a) schematic illustrating the PFrag setup for the

periodic SiCNT of figure 1, with one PFrag type generated in

a six unit periodic NT template. Si atoms are depicted as larger

spheres. Panel (b) DoS for the PFrag and LS approaches. Panel (c)
electron densities (ρ) averaged over the xy-plane (left) and their

corresponding errors (∆ρ) relative to the LS reference (right).

Returning to the specific example of the SiCNT, we have
taken a ring of atoms in the centre of the NT as a reference,

and calculated onsite overlap matrix elements between SFs

of the same type centred on equivalent atoms along the NT,

t.i.e. with the same x and y coordinates but different z coordi-

nates. The various elements of the system are therefore simply

translated with respect to each other. For symmetry reasons,

we expect that such a choice of the reference SFs would pro-

vide a unit value of $S_{\text{onsite}}$. The results of 1 − $S_{\text{onsite}}$ are plotted

in figure 1(c). The values in the central region (highlighted

in grey) should be zero, since they correspond to the overlap

between an SF and itself; the values in this region therefore

give an estimate of the noise level associated with the wavelet

grid spacing. Although there is some variation along the NT,

the values remain small throughout, in line with our obser-

vations that the SFs in different rings of the NT are indeed

equivalent. Such a measure therefore helps us in quantifying

to what extent the value of $S_{\text{onsite}}$ might deviate from unity

before the SFs can no longer be considered to be equivalent.

In some sense, it may be ascribed to a calibration procedure of

the fundamental constituents of the system’s density matrix.

3.2. (Pseudo) Fragments in a periodic system

Given the above, one might imagine that, rather than opti-
mizing the SFs for the entire NT, one could take the SFs from

a single slice, and replicate them across the NT. Assuming

that the fluctuations in the onsite overlap matrix are not

meaningful (i.e. purely noise), this approach should have

little impact on the accuracy of the calculation. We note that

although there are other possible divisions of a SiCNT into

fragments, we choose to take the simple definition of one

complete (covalently bonded) ring of the NT (28 atoms) for

convenience. Unlike in a molecular system, such a fragment

is not in any sense separable from the full system, so we use

the term pseudo-fragment (PFrag) to distinguish it from its

molecular counterpart.

The question becomes, how does one perform the equiva-
lent of a molecular template calculation, i.e. how can one gen-
erate the SFs for a PFrag which are appropriate for an extended

system? It is clear that optimizing the SFs for an isolated ring

would be a poor approach. Instead, following the principles

of nearsightedness, we propose to perform a template calcula-
tion wherein the PFrag is embedded in a representative

environment. Given a sufficient number of nearest neighbours

(i.e. neighbouring slices), this should result in an SF basis

which is of a similar quality as that which has been directly

generated in the full system.

3.3. Pseudo-fragment workflow

The workflow for the SiCNT is thus as depicted in

figure 2(a). The template SFs are generated for a shorter

SiCNT periodic supercell, in this case containing 6 ring

units (168 atoms). The SFs associated with a given ring are

then replicated for the 14 unit calculation, while the SFs

from other five rings are discarded (although they could

also be taken into account by averaging, see appendix B.1).

Aside from the need for an embedded template calculation,

the calculation follows along similar lines as a molecular

fragment calculation, with the SFs remaining fixed for the

full system, requiring only the (self-consistent) optimi-

zation of the density kernel. The only exception is that

building the density kernel from embedded fragments is less

straightforward than in the isolated case. Here and in the

following we choose to build the initial guess for the den-
sity kernel directly from the density kernel of the template
calculation(s), as discussed in appendix B.2.

The total energies and band gaps resulting from the LS and

PFrag setups are given in table 1, where it can be seen that

the PFrag approach results in a negligible error in the total

energy—around 3 meV/atom. The band gap is also in good

agreement. It should be noted that the band gap is signifi-
cantly overestimated compared to the cubic scaling approach,

by more than 1 eV. This is to be expected, since the SF basis

is explicitly optimized to represent the occupied KS states.

In situations were one requires access to one or two unoc-
cupied states, one might explicitly add them into the SF

optimization procedure, in which case it is necessary to use

the direct minimization approach for density kernel optim-

ization [8]. For a larger number of states, it would be neces-
sary to use an approach such as the one used in the onetep

code and optimize a second set of SFs specifically to repre-
sent the unoccupied states [31]. For the purposes of this work,

however, we are interested only in demonstrating the ability to

reproduce the band gap of the optimized SF basis, which we

can see is the case to within 0.05 eV.
The density of states (DoS) for the two methods are depicted in figure 2(b), where the KS energies were obtained from a one-off diagonalization in the SF basis at the end of the LS and PFrag calculations. The curves for the two approaches are indistinguishable at the given level of smearing. Although not depicted here, we note that aside from the aforementioned discrepancy for the unoccupied states, the linear and cubic scaling DoS are also in excellent agreement, demonstrating that the PFrag approach is indeed able to reproduce the correct occupied DoS.

Finally, the average electronic densities along the axis of the SiCNT are plotted in figure 2(c). As with the DoS, the average density of the PFrag approach is indistinguishable from the LS calculation. Looking also at the average differences along the NT with respect to the reference LS density, it is clear that the error for the PFrag approach is very small compared to the absolute values. For reference, the average density difference between the linear and cubic scaling calculations is of the order of $10^{-4}$. In short, by all of the above measures, one can conclude that the PFrag approach is indeed able to accurately reproduce the electronic structure of the full SiCNT.

Once the PFrag basis is generated, such an approach provides in principle a set of pre-optimized basis functions, whose utility is twofold: on one hand, they offer a basis set that might be employed as-is for the simulation of large pristine systems, enabling considerable computational savings compared to optimizing the SFs for the full system, as well as a set of fixed degrees of freedom on which ‘second principles’ Hamiltonians (like for example tight binding models) can be designed. On the other hand, more interestingly, the pre-optimized basis set equipped with the calculation of $S_{\text{ onsite}}$ provide a gauge to measure the actual extension of defective regions, i.e. the region for which the pristine SFs cannot be used to obtain high levels of precision. We will see in the following an example of this procedure.

### 3.4. Edge effects in a finite SiC nanotube

We now consider what happens in the case of a finite SiCNT, i.e. what effect do the edges of the NT for a given surface reconstruction have on the SFs? We again take a 14 unit SiCNT, this time with the edge termination from [32], as depicted in figure 3(a). We expect that SFs associated with atoms near the centre of the SiCNT should be similar, while those near the edges should differ. As with the periodic NT, we start by calculating $S_{\text{ onsite}}$, which is plotted in figure 3(c), with selected SFs depicted in figure 3(b). As expected, the differences between edge SFs and central SFs are significant, while the values of the $S_{\text{ onsite}}$ vary smoothly as one moves along the NT. Indeed, those in the centre, beyond around 10 Å from the edge, show the same level of consistency as the SFs of the periodic calculation, confirming that one can define a bulk-like central region.

The simplest option for performing a PFrag calculation would be to take the SFs from the periodic template, as depicted in figure 4(a). However, while this might represent a reasonable approximation for the edges of the NT, this would clearly be unsatisfactory for the edges where the SFs are very different. A better approach would therefore consist of defining several PFrag types, including edge, bulk and, if needed, a (yet to be determined) number of intermediate types. Following such intuition, confirmed by our values of $S_{\text{ onsite}}$, one could define a series of setups, depicted in figures 4(b)–(d), wherein the SFs are generated for a SiCNT which has an increasing number of intermediate PFrag types and is therefore increasingly long. We note that, as with the full system, we relaxed the geometries of the template structures for all PFrag setups using the parameters described in appendix A.

In the following we test each of the proposed PFrag setups, again comparing with the results of the full LS approach. Although it is possible to account for reflections using the fragment approach implemented in BigDFT [24], we choose to avoid doing so in order to simplify the calculation setup and analysis. Therefore, taking the setup of figure 4(b) as an example, this corresponds to two separate edge PFrag types and two separate bulk PFrag types. We therefore refer to this setup as 4 PFrags from now on.

Given the values of $S_{\text{ onsite}}$, we expect that SFs centred on atoms beyond around 10 Å from the edge should be far enough away to be treated as bulk like. Therefore, we expect
the setup of figure 4(d) to be sufficient to reproduce the correct electronic structure of the LS reference calculation. In order to determine whether or not this is indeed the case, we also take one additional setup, i.e. 10 Pfrag types in total. Finally, we also consider an example of combining PFrag types from the periodic and finite setups, by replacing the central Pfrag SFs with those taken from the full SiCNT, i.e. for the 10 Pfrag case this is the average across all SFs which are in the central bulk-like region. This is compared to the corresponding $\Delta E$. We also include the equivalent values for the periodic SiCNT on the same plot. By fitting a linear function to the data, we show in figure 3(d) that the average $S_{\text{onsite}}$ at a given distance from the edge indeed provides a reliable indication of whether or not a given region may be treated as bulk-like or if additional Pfrag types are required.

4. Pseudo-fragments and lattice distortions: the cost function $J$ for error estimation

Thus far we have demonstrated the utility of the onsite overlap matrix for predicting whether a particular Pfrag setup is likely to be appropriate for a given target system, for example by determining the lengthscale over which a perturbation might affect the form of the SFs. In this section we now consider the related question of the suitability of a Pfrag setup wherein the geometry of the template PFrag types does not match that of the target system, i.e. the value of the cost function $J$ is non-zero.

As with the molecular fragment approach, the embedded Pfrag method allows for small deviations in geometry between template and system PFrag types, however if these differences become large then additional types of Pfrag may be required. For systems with high levels of disorder and therefore very many Pfrag types (i.e. little repetition) such an approach might therefore not be appropriate. Stated otherwise, the $S_{\text{onsite}}$ indicator might be useful only when there is not too much geometrical difference—quantified by the cost function $J$—between the template PFrag types and the target system. It would therefore be useful to predict whether or not a Pfrag setup is valid based on how much noise is present in the system. That is, for a given value of the cost function $J$, what is the magnitude of the error introduced?

In order to answer this question, we first take the case of the pristine SiCNT and investigate the effect of adding random noise to the system. We then return to the fixed length
4.1. Adding noise: effect of imperfect templates

To better understand if the cost function $J$ might be used as a quality indicator, we add increasing levels of random noise to the atomic positions of the 14 unit periodic SiCNT of figure 1(a). For each level of noise, we compare the total energy using fully optimized SFs with that using the PFrag SFs from the pristine NT. For small levels of noise the PFrag setup should remain reasonable, while for high levels of noise the errors are expected to be significant.

For systems which contain more than one PFrag type, the cost function $J$ represents a useful quantity for describing the departure from rigid roto-translations. In this case, each PFrag instance will have its own $J$ thanks to the added noise; in the following we take the average value, $J_{AV}$. The level of noise for a given geometry may also be characterized by the average absolute difference between the noisy and pristine geometries, $\Delta R$, which in turn should be proportional to $J_{AV}$. In the middle panel of figure 6 we plot $\Delta R$ versus $J_{AV}$, showing that this is indeed the case. In the left hand panel we show the error in the energy associated with the PFrag approach—as expected this increases for higher levels of noise. For $\Delta R \approx 0.01 \text{ Å}$, the error is of the order of 50 meV/atom.

More interestingly, one can also assess the correlation between $J_{AV}$ and $\Delta E$. As shown in the right hand panel of figure 6, these two quantities are well correlated within the range of values considered, being well described by a linear fit. One could therefore imagine applying a correction to the PFrag energy by predicting the error based on the linear fit. If one applies such a correction, as shown in figure 6, this results in an error of less than 5 meV/atom for each case, even for relatively large levels of noise. In other words, the remaining error is purely due to that of the interpolation when roto-translating fragments.

For systems with little noise ($J \lesssim 0.005 \text{ Å}^2$) one can therefore apply the PFrag approach without inducing significant additional errors due to non-rigid PFrags. For larger values of $J$, one can estimate fairly well the induced error. Clearly, this estimation might only be provided \textit{a posteriori}, i.e. if the
Finally we note that in cases where the cost function is large, the Pfrag setup might still offer a better initial guess for the SFs. In the case of the noisiest geometry considered, using such a guess results in a factor of two reduction in the computational cost compared to using an initial guess based on atomic orbitals. In comparison, directly using the Pfrag approach is more than 17 times quicker than the full LS approach for the same calculation.

4.2. Effect of non-zero cost function for finite SiC nanotube

We now return to the finite SiCNT in figure 3(a). As mentioned in section 3.2, the geometries were relaxed for all NTs, both the templates and the 14 unit target system. As such, we expect to see non-negligible values of $J$, particularly for very short template systems, where the differences in geometry with respect to the longer NT will be significant. As we can see from table 1, this is indeed the case, with a large value of $J_{\text{av}}$ for the 1 Pfrag case, and decreasing values for increasing numbers of PFrags.

Given the above considerations, we might estimate the error associated with the deviation in geometry between the template and target system PFrags, by using the linear fit between $J_{\text{av}}$ and $\Delta E$ for the noisy system. In other terms, we assume that the error coming from the non-zero cost function is equivalent to that due to random distortions in the geometry. The values for the predicted error for the finite SiCNT are given in table 1. For large values of $J$ the errors are significantly overestimated, which is unsurprising given this is well outside the regime used for the fit. For smaller values of $J$ the errors agree relatively well with the actual errors. Importantly, comparing the 8 and 10 Pfrag calculations the further improvement in the energy can indeed be attributed to the decrease in the cost function. Such a qualitative behaviour is also present in the mixed periodic/finite Pfrag setup.

4.3. Finite SiC nanotubes of increasing length

Finally, we also consider SiCNTs of increasing length to better understand the transferability of a given set of PFrags to similar systems of diverse sizes. It was previously suggested [32] that the energy $E_{\text{NT}}$ of a SiCNT of a given diameter with $n$ units can be described by

$$E_{\text{NT}}(n) = (n - 2n_{\text{edge}}) E_{\text{bulk}} + 2E_{\text{edge}},$$

where $n_{\text{edge}}$ is the number of units which can be considered to form the ‘edge’ region at one end of the NT (i.e. the termination), $E_{\text{edge}}$ is the corresponding energy and $E_{\text{bulk}}$ is the energy of a bulk-like unit. A value of $n_{\text{edge}} = 3$ was previously used, which agrees well with the above conclusions concerning both the distance from the edge of the NT beyond which the SFs revert to the equivalent bulk-like form, and the length beyond which the deviations in atomic coordinates (indicated by the cost function $J$) are small. In the following we vary the length of the finite SiCNT, in order to determine how well this model is obeyed. For each length of SiCNT, we relax the structures, as described in appendix A. We compare the results of fitting
the above model for full LS calculations, and for Pfrag calculations using the 8 and 10 Pfrag setups.

The results are plotted in figure 7. For each scenario we have plotted both the actual energies obtained and the curve resulting from fitting the data to equation (7). It can be seen that for each setup the energies fit the model very well, confirming the above observations. Furthermore, the agreement between the Pfrag and LS setups is excellent, with only very small (less than 7 meV/atom) differences appearing for the longer SiCNTs. Referring to the centre panel of figure 7, it can be seen that this is due to larger values of $J$. That is, the difference in geometry with respect to the template SiCNT increases with length until the point where relaxing the geometry has no further effects. While the value of $J$, and correspondingly the associated errors, remain small, the cost function and the error in energy are again well correlated, as can be seen in the right panel of figure 7. It is therefore clear that $J$ can be seen as a general purpose indicator of the suitability of a given Pfrag setup (i.e. the approximate induced errors) for cases where there are distortions in geometry, whether due to random noise as discussed above, or due to more systematic differences resulting from structural relaxations.

For the above calculations the Pfrag approach is around eight times faster than the full LS approach. The smaller speedup compared to the periodic NT calculations is due to the density taking longer to converge for the finite NT in the Pfrag approach. Nonetheless, the computational savings are significant.

5. Conclusion

The treatment of large systems in a density functional theory framework is a task which poses a number of challenges. The most important is the reduction of the number of degrees of freedom, within a paradigm with controllable accuracy. The LS approaches such as those adopted in onetep, Conquest and BtoDFT represent a first fundamental step in this direction: the SFs are adapted to the chemical environment surrounding each region. Such adaptivity allows the reduction of the degrees of freedom, with essentially no cost in terms of accuracy, as the occupied KS orbitals are well represented by design within such an approach.

This suggests that for large systems the degrees of freedom may be reduced further, by identifying the various regions which might be expressed with the same set of SFs, in other terms those regions which have a similar local chemical environment. This is important not only in view of computational savings, but also to gain further insights into the
systems’ constituents. The reduction of the complexity of the
description is very important in this context: performing a
set of production quantum mechanical simulations with an
approach which is unnecessarily costly would provide a study
of poor quality, as the simulation scheme would entangle
interactions of different length scales and couplings.

A framework for identifying and exploiting regions
with similar chemical environments has already proven
very useful in the context of molecular systems, where the
search for systems’ constituents naturally leads to fragment-
tation. We have provided in this paper a similar approach for
extended systems. In such cases the concept of a separable
fragment is not meaningful. However, we have provided
indicators which help to distinguish regions of an extended
system which may be treated with identical degrees of
freedom from regions which require a different set of SFs
for an accurate simulation.

We have shown how a support function based Pfrag
approach may be applied to a SiC nanotube, demonstrating
the applicability of such indicators in assessing the suitability
of a given calculation setup. This Pfrag approach permits the
exploitation of support function similarity in a whole new
range of materials, and since there are no restrictions on cut-
ting covalent bonds between pseudo-fragments, the user is
free to define them according to their preference. In this work
we focus on a quasi-one-dimensional system, however our
approach might easily be applied to extended systems in two
or three dimensions, for example for the treatment of defects.
Work in this direction is ongoing. Both the setup and anal-
ysis of the presented calculations (including the generation
of input files) has been done via the use of a Jupyter Notebook,
with the aim of easing both reproducibility of the results and
the adoption by interested users. We also note that in cases
where a Pfrag approach is not directly applicable, e.g. when
there is significant distortion in the geometry of the system (as
indicated by a cost function), one might still use it to generate
a better input guess for the SFs. This can significantly reduce
the number of iterations required to further optimize the SFs.

The set of methods and indicators defined in this paper
might also be considered as a first step towards the control
of the setup of more complex approaches, such as the treat-
ment of defects with various levels of theory. Currently we
keep the SFs fixed, however this might be extended in future
to allow the SFs to be further optimized in an active region
and remain fixed in an environment region, i.e. to define an
embedding method at the level of the basis. Long ranged in-
teractions, which are not associated to a modification of the SFs,
might therefore be treated with a different level of theory than
short-ranged ones. In such cases, the indicators presented in
this work could be used to predict the size of the active region,
thereby enabling an a priori control of approaches that so far
have primarily relied on the physical intuition of the user.

Acknowledgments

LER acknowledges an EPSRC Early Career Research
Fellowship (EP/P033253/1) and the Thomas Young Centre
under grant number TYC-101. This research used resources
of the Argonne Leadership Computing Facility at Argonne
National Laboratory, which is supported by the Office of
Science of the US Department of Energy under contract
DE-AC02-06CH11357. We are grateful to the UK Materials
and Molecular Modelling Hub for computational resources,
which is partially funded by EPSRC (EP/P020194/1). This
work also used the ARCHER UK National Supercomputing
Service (www.archer.ac.uk). LER thanks Damien Caliste
for assistance with visualizing the SFs using v_sim.

Calculation input and output files, and the Jupyter Notebook
which has been used both to prepare the inputs and extract the
data presented above, including the majority of the figures, are
available at https://gitlab.com/luigigenovese/pfrag-sicnt.

Appendix A. Computational details

Calculations employed a grid spacing of 0.18 Å, corresponding
to an accuracy of around 1 meV/atom. Four SFs were used for
each atom, with localization radii of 4.5 Å, which were selected
to give an error of 10 meV/atom relative to the cubic scaling
reference. For the simulations in periodic boundary conditions,
we used cell dimensions of 31.8 Å in the x and y directions
to reduce spurious interactions between periodic images. For
the LS calculations the FOE method was used to optimize the
density kernel. We used HGH pseudopotentials [33] and the
PBE exchange correlation functional [34]. All calculations
were performed at the Γ-point only. Structures were relaxed
using the cubic scaling approach using relatively loose conve-
gerence criteria, since the emphasis is on understanding the
lengthscale over which an edge termination affects the elec-
tronic structure, rather than precisely converging the energies.
For the 14 unit structures and shorter template calculations the
convergence criterion was a maximum force component below
0.03 eV Å⁻¹, while for the longer finite SiCNTs a value of
0.1 eV Å⁻¹ was used. For the DoS calculations, Gaussian
smearing of 0.1 eV has been applied and the curves have been
shifted so that the HOMO is at zero.

Appendix B. Implementation details

B.1. Multiple instances of a pseudo-fragment

It should be noted that, unlike for an isolated template calcu-
lation, an embedded template calculation might include more
than one instance of a given Pfrag type. For example, in the
case of the periodic SiCNT template calculation depicted in
figure 2(a) there are six instances of a single Pfrag type. In
this case, the first instance of each Pfrag type is the one for
which the SFs and associated quantities are written to disk and
thus reused for the full calculation, irrespective of whether
this is ‘typical’. For a template calculation where all instances
of each Pfrag type are truly identical, this point is irrelevant.
However in practice such PFrags will not always be identical,
e.g. due to numerical noise for example coming from the egg-
box effect or slightly differing chemical environments.
In principle, one could reduce the impact of arbitrarily selecting a single instance of a PFragment by performing an average over all instances, however the cost of performing this (including communicating the SFs) is non-negligible, and so it is instead left to the user to ensure the PFragment definitions are sensible, i.e. that the differences between different instances of the same PFragment are small. To aid with this process, one could again refer to the onsite overlap matrix—if, for example, in a template calculation, the onsite overlap between SFs on equivalent atoms on different instances of the same PFragment is much less than one, one can infer that the PFrags are distinct and an additional PFragment type should be defined.

B.2. Initial guess for the density kernel

For isolated fragment calculations, it is straightforward to build an initial guess for the KS coefficients (or equivalently, the density kernel) from the isolated coefficients (kernel). However, the best procedure is less obvious for the embedded case, since it is not easy to separate the PFragment from its environment. In particular, if one were to use very small PFrags which are strongly interacting with their environment (i.e. connected via covalent bonds), it is not clear how one should extract the KS coefficients associated with a given PFragment—it would be a rather poor approximation to cut the KS coefficients and retain only those associated with SFs belonging to that PFragment. Similarly, taking and occupying only the lowest occupied KS orbitals and rejecting the higher energy occupied orbitals (in order to account for the lower number of electrons in the PFragment compared to the PFragment plus its environment) would result in a significant loss of information.

It is easier to cut the density kernel by retaining only matrix elements between SFs associated to the PFragment, however it still results in a loss of information. In particular the total electronic charge is not correctly preserved. For calculations with relatively large PFrags, this approach nonetheless results in a reasonable initial guess for the density kernel, while the total number of electrons reverts to the correct value after a few self-consistent iterations.

However, for calculations with much smaller PFrags, where the approximation is more severe, an alternative approach might be preferable. In such cases, a more naive, conceptually more straightforward approach would be to generate a ‘diagonal’ density kernel, i.e. giving all SFs an equal weighting and normalizing to give the correct electron number. This has the advantage of not introducing any particular bias, while also straightforwardly preserving the total number of electrons, and in practice has proven to be the most robust of the different options.

A final option is to randomly generate the density kernel (KS coefficients) and purify (orthonormalize). However this approach is rather unstable compared to the alternatives.

**ORCID iDs**

Laura E Ratcliff [https://orcid.org/0000-0002-9760-5465](https://orcid.org/0000-0002-9760-5465)

Luigi Genovese [https://orcid.org/0000-0003-1747-0247](https://orcid.org/0000-0003-1747-0247)

**References**

[1] Hohenberg P and Kohn W 1964 Phys. Rev. 136 B864–71
[2] Kohn W and Sham L J 1965 Phys. Rev. 140 A1133–8
[3] Goedecker S 1999 Rev. Mod. Phys. 71 1085–123
[4] Bowler D R and Miyazaki T 2012 Rep. Prog. Phys. 75 036503
[5] Ratcliff L E, Mohr S, Huils G, Deutsch T, Masella M and Genovese L 2017 WIREs Comput. Mol. Sci. 7 e1290
[6] Skylaris C K, Haynes P D, Mostofi A A and Payne M C 2005 J. Chem. Phys. 122 84119
[7] Bowler D R and Miyazaki T 2010 J. Phys.: Condens. Matter 22 074207
[8] Mohr S, Ratcliff L E, Boulanger P, Genovese L, Caliste D, Deutsch T and Goedecker S 2014 J. Chem. Phys. 140 204110
[9] Mohr S, Ratcliff L E, Genovese L, Caliste D, Boulanger P, Goedecker S and Deutsch T 2015 Phys. Chem. Chem. Phys. 17 31360–70
[10] Yang W 1991 Phys. Rev. Lett. 66 1438–41
[11] Fedorov D G, Jensen J H, Deka R C and Kitaura K 2008 J. Phys. Chem. A 112 11808–16
[12] Fedorov D G, Avramov P V, Jensen J H and Kitaura K 2009 Chem. Phys. Lett. 477 169–75
[13] Fedorov D G and Kitaura K 2017 J. Chem. Phys. 147 104106
[14] Genova A, Ceresoli D, Krishtal A, Andreussi O, DiStasio R A and Pavanello M 2017 Int. J. Quantum Chem. 117 e25401
[15] Manby F R, Stella M, Goodpaster J D and Miller T F 2012 J. Chem. Theory Comput. 8 2564–8
[16] Fornace M E, Lee J, Miyamoto K, Manby F R and Miller T F 2015 J. Chem. Theory Comput. 11 568–80
[17] Cheng J, Libisch F, Yu K, Chen M, Dietrich J M and Carter E A 2017 J. Chem. Theory Comput. 13 1067–80
[18] Cheng J, Yu K, Libisch F, Dietrich J M and Carter E A 2017 J. Chem. Theory Comput. 13 1081–93
[19] Culpitt T, Brorsen K R and Hammes-Schiffer S 2017 J. Chem. Phys. 146 211101
[20] Ding F, Manby F R and Miller T F 2017 J. Chem. Theory Comput. 13 1605–15
[21] Libisch F, Marsman M, Burgdörfer J and Kresse G 2017 J. Chem. Phys. 147 034110
[22] Chulhai D V and Goodpaster J D 2018 J. Chem. Theory Comput. 14 1928–42
[23] Hégely B, Nagy P R and Kállay M 2018 J. Chem. Theory Comput. 14 4600–15
[24] Ratcliff L E, Genovese L, Mohr S and Deutsch T 2015 J. Chem. Phys. 142 234105
[25] Genovese L et al 2008 J. Chem. Phys. 129 014109
[26] Mohr S, Masella M, Ratcliff L E and Genovese L 2017 J. Chem. Theory Comput. 13 4079–88
[27] Ratcliff L E, Grisanti L, Genovese L, Deutsch T, Neumann T, Damilov D, Wenzel W, Beljonne D and Cornil J 2015 J. Chem. Theory Comput. 11 2077–86
[28] Wahba G 1965 SIAM Rev. 7 409
[29] Kabsch W 1978 Acta Crystallogr. A 34 827–8
[30] Markley F L 1989 J. Astronaut. Sci. 38 245–58
[31] Ratcliff L E, Hine N D M and Haynes P D 2011 Phys. Rev. B 84 165131
[32] Pochet P, Genovese L, Caliste D, Rousseau I, Goedecker S and Deutsch T 2010 Phys. Rev. B 82 035431
[33] Hartwigs C, Goedecker S and Hutter J 1998 Phys. Rev. B 58 3641–62
[34] Peredew J P, Burke K and Emzerhof M 1996 Phys. Rev. Lett. 77 3865–8