Partitioning the vibrational spectrum: Fingerprinting defects in solids

Danny E.P. Vanpoucke[a,b,c,*]

[a] Maastricht University, Aachen-Maastricht Institute for Biobased Materials (AMIBM), Brightlands Chemelot Campus, Urmonderbaan 22, 6167RD Geleen, The Netherlands
[b] Institute for Materials Research (IMO), Hasselt University, 3590 Diepenbeek, Belgium
[c] IMOMEC, IMEC vzw, 3590 Diepenbeek, Belgium

ABSTRACT

Vibrational spectroscopy techniques are some of the most-used tools for materials characterization. Their simulation is therefore of significant interest, but commonly performed using low cost approximate computational methods, such as force-fields. Highly accurate quantum-mechanical methods, on the other hand are generally only used in the context of molecules or small unit cell solids. For extended solid systems, such as defects, the computational cost of plane wave based quantum mechanical simulations remains prohibitive for routine calculations. In this work, we present a computational scheme for isolating the vibrational spectrum of a defect in a solid. By quantifying the defect character of the atom-projected vibrational spectra, the contributing atoms are identified and the strength of their contribution determined. This method could be used to systematically improve phonon fragment calculations. More interestingly, using the atom-projected vibrational spectra of the defect atoms directly, it is possible to obtain a well-converged defect spectrum at lower computational cost, which also incorporates the host-lattice interactions. Using diamond as the host material, four point-defect test cases, each presenting a distinctly different vibrational behaviour, are considered: a heavy substitutional dopant (Eu), two intrinsic point-defects (neutral vacancy and split interstitial), and the negatively charged N-vacancy center. The heavy dopant and split interstitial present localized modes at low and high frequencies, respectively, showing little overlap with the host spectrum. In contrast, the neutral vacancy and the N-vacancy center show a broad contribution to the upper spectral range of the host spectrum, making them challenging to extract. Independent of the vibrational behaviour, the main atoms contributing to the defect spectrum can be clearly identified. Recombination of their atom-projected spectra results in the isolated spectrum of the point-defect.

1. Introduction

Vibrational spectroscopy is a widely used tool for the structural characterization of solids [1–9]. Modeling these experimental spectra at the quantum mechanical level starts from the calculated vibrational spectrum of the system, with the appropriate intensities for the individual spectral modes determined depending on the target experimental technique (e.g., Infra-Red, Raman,…) [10,2]. Since atomic scale calculations allow for full control over the underlying atomic structure, they also provide a direct link to the features of the resulting vibrational spectrum. As such, they give access to an invaluable source of structural information and fundamental understanding; e.g., about the impact of (point-) defects [11–16]. However, such calculations for point-defect systems at the quantum mechanical level face a significant limitation: the calculation of an accurate and converged vibrational spectrum

requires the use of large cells, which is computationally very demanding [17,18,9]. Furthermore, in the case of point-defects, the main interest goes to the modifications of the host spectrum due to the defect. It is therefore of interest not to spend computational resources on the reconstruction of the host spectrum, but to limit the calculations to the contributions of the point-defect itself.

In contrast to solid state modeling, where the phonon spectra are generally only considered for small unit cell systems [19,20,12,21,16,22–24], several approaches have been developed to deal with large systems within the context of (bio-) molecular structure investigations [25]. Examples are the selective calculation of specific normal modes [26,18], the partial optimization of the molecular geometry [27], and the reduction of the Hessian by assuming rigid subsystems [28–30]. Others have presented a fragment approach, in which a large system is decomposed into fragments for which high quality

properties are calculated. The resulting fragment properties are then recombined as approximation of the original system [31]. Yamamoto *et al.* [32,33] showed this method reproduces spectra of large molecules faithfully as long as "suitable" fragments were selected. They also note that the force field transfer accounted for nearly half of the observed error [32]. Hanson-Heine *et al.* [34] presented a local mode approach which can be used within the context of 2DIR spectroscopy of large systems, where it provides a platform for the parameterization of site frequencies and coupling maps with regard to the geometry of different functional groups. Another proposed fragment strategy is centered on the construction of the Hessian matrix considering only the atoms in the region of interest [35]. This approach efficiently succeeds in reproducing the spectra of interest, requiring only a small number of atoms to be considered. In this method, the selection of the atoms belonging to the fragment is rather *ad hoc*. Furthermore, delocalized modes can not be treated with this approach. More specifically, within this approach the remainder of the system is kept frozen which may in some cases lead to unphysical behaviour if a normal mode is not tightly localized. A similar partial Hessian approach is the so-called Mobile Block Hessian approximation [29,30]. In this approach, the remainder of the system is also considered, but to reduce the computational cost, the atoms outside the fragment of interest are grouped in rigid blocks, which have no internal degrees of freedom, only 6 external degrees of freedom. In contrast to these fragment approaches, Teodoro *et al.* [18] consider the full system using a computationally cheap approximate method to obtain the initial full spectrum, after which only the normal-modes of experimental interest are selected and re-evaluated with a more accurate method. Within the context of solids, some authors have presented approaches in which the real space force constants matrix of a defect super cell is embedded in a much larger force constant matrix of the host material [11,9]. Such approaches have been used to investigate disorder and dopant incorporation [20,13].

In this work, we approach the problem of creating a defect-spectrum from a different angle (*cf.*, Fig. 1). Starting from a comparison to the vibrational spectrum of the host material, we present a method to identify the atoms contributing to the vibrational spectrum of a defect. The overlap of the atom-projected spectrum with the reference host spectrum is presented as a suitable quantitative measure. A small defect cell is shown to suffice to clearly identify the defect atoms, which are limited in number. In addition, the defect spectrum obtained through the combination of the atom-projected spectrum of the defect atoms is well-converged even using a small defect system cell. The resulting defect spectrum is continuous in nature, due to the incorporation of defect-host interactions. Four diamond-based test cases are considered, for which the individual spectra of the point-defects are determined.

Within the context of the fragment approximations mentioned earlier, this method could resolve the *ad hoc* nature of atom selection.
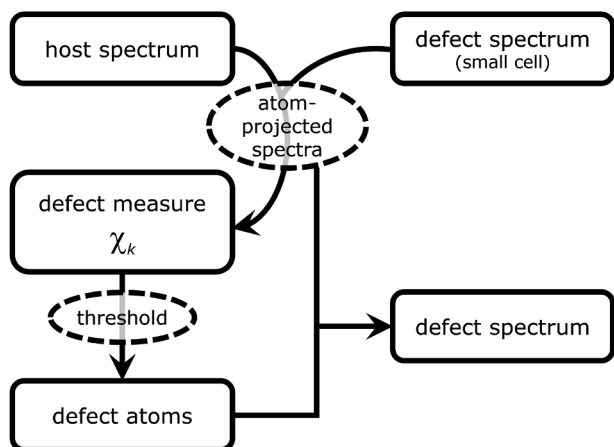

Fig. 1. Flowchart of the method used to create defect spectra.

Furthermore, within the context of the partial Hessian approximations, having a quantitative measure for the defect nature of an atom, would allow for more targeted selection of Hessian sub-blocks. In both cases, small supercells can be used to identify specific defect-atoms, and partial Hessian calculations on large supercells to obtain the spectrum of interest, reducing the computational cost of obtaining an accurate quantum mechanical vibrational spectrum in a periodic solid.

2. Computational methods

First-principles calculations are performed within the Density Functional Theory (DFT) framework using the VASP package [36]. The kinetic energy cutoff of the plane wave basis set is set to 600 eV to obtain well converged forces, while the exchange correlation functional as proposed by Perdew, Burke and Ernzerhof (PBE) is used to describe the valence electron interactions [37]. The point-defects are imbedded in a 64-atom conventional supercell, with the first Brillouin zone sampled by a $5 \times 5 \times 5$ Monkorst-Pack grid. The negatively charged NV center is explicitly modeled using a charged cell. Note that for practical algorithmic stability, the VASP package counters this charge with an opposing uniformly charged field. Although Coulomb interactions require the use of very large cells to obtain converged defect energies, we will see that for our purposes, the 64 atom cell already provides a clear qualitative picture. Further details on the computational settings used are presented elsewhere [38,15].

3. Harmonic phonon spectrum of solids

In the following, we use the compact notation for the dynamical matrix of Born and Huang [39], adding superscripts to distinguish between different practical situations. Vectors are indicated in bold, and full matrices are indicated with a bar.

3.1. Construction of the atom-projected phonon DOS

Most modern quantum mechanical and quantum chemistry packages provide access to the vibrational spectrum of a system at the center of the first Brillouin zone, as a means to assess the quality of the obtained atomic structure [40]. This vibrational spectrum at the Γ-point, can be obtained by the diagonalisation of the mass-weighted Hessian matrix, also called dynamical matrix:

$$D_{\alpha,\beta}^{mol}\left(k,l,\Gamma\right) = \frac{1}{\sqrt{m_k m_l}}\Phi_{\alpha,\beta}\left(k,l\right) \tag{1}$$

with

$$\Phi_{\alpha,\beta}\left(k,l\right) = \left(\frac{\partial^2 E}{\partial x_\alpha(k)\partial x_\beta(l)}\right)_0. \tag{2}$$

In these matrix elements, E represents the total energy of the system, k and l are indices for the atoms of the system with m_k and m_l their respective atomic mass. The cartesian directions are indicated by $\alpha, \beta = x, y, z$. In practical calculations, the individual matrix elements $\Phi_{\alpha,\beta}(k,l)$ can be determined either as the second derivative of the total energy with regard to the displacements $x_\alpha(k)$ and $x_\beta(l)$ of atom k and l, respectively, or as the derivative of the forces acting on atom k by the displacement of atom l: [41]

$$\Phi_{\alpha,\beta}\left(k,l\right) = \frac{\partial^2 E}{\partial x_\alpha(k)\partial x_\beta(l)} = -\frac{\partial F_\alpha(k)}{\partial x_\beta(l)}. \tag{3}$$

For molecular systems, or clusters, a dynamical matrix as presented in Eq. (1) would provide a complete picture. However, in the case of a periodic solid there are two complications that need to be dealt with: (1) the infinite nature of a theoretical crystal and (2) the finite size of the first Brillouin zone.

For a periodic crystal, all relevant physics is contained in a single unit cell, reducing the number of atoms to consider from infinity to a (small) finite number. On the other hand, to obtain the phonon density of states of a periodic solid, one needs to integrate the spectrum over the full Brillouin zone (similar as for the calculation of the electronic density of states). As such, one infinity is replaced by another, albeit a more manageable one, since the practical integration over the Brillouin zone (BZ) is performed over a discretized grid of $\mathbf{q}$-points. The vibrational spectrum at each $\mathbf{q}$-point is determined through the diagonalisation of the dynamical matrix: [39]

$$D_{\alpha,\beta}^{BZ}\left(k, l, \mathbf{q}\right) = \frac{1}{\sqrt{m_k m_l}}\Phi_{\alpha,\beta}\left(k, l\right)\exp i\mathbf{q}\left(\mathbf{r_k} - \mathbf{r_l}\right), \quad (4)$$

with $\mathbf{r_k}$ and $\mathbf{r_l}$ the real space position vectors of atoms k and l, respectively.

Furthermore, because interatomic interactions are infinitely ranged, the dynamical matrix needs to incorporate interactions with other unit cells as well. Indexing the unit cells, with $R = 1$ being the reference unit cell (UC), the general form of the dynamical matrix can be written as:

$$D_{\alpha,\beta}^{BZ,UC}(k, l, \mathbf{q}) = \sum_{R=1}^{\infty} \frac{1}{\sqrt{m_k m_l}}\Phi_{\alpha,\beta}(k, l)\exp i\mathbf{q}(\mathbf{r_{k,1}} - \mathbf{r_{l,R}}), \quad (5)$$

with $k, l \in [1, n]$ and n the number of atoms in the unit cell. For practical purposes, R can be truncated to a finite number of unit cells, R_{max}, as the contributions to the dynamical matrix of unit cells farther away becomes vanishingly small [39,42]. The convergence of the vibrational band structure and density of states (DOS), as function of R_{max}, is shown in Fig. 2. Note that diamond has a rather small primitive unit cell. For large unit cell systems, such as for example metal-organic frameworks [43], a converged spectrum may be obtained already at the unit cell level (i.e., $R_{max} = 1$).

Furthermore, for supercell calculations [44], which are generally used to obtain vibrational spectra from quantum-mechanical calculations [45], the supercell dynamical matrix, Eq. (4), and the unit cell dynamical matrix, Eq. (5), are related through symmetry. Both give rise to the same phonon DOS. However, as matrix diagonalization scales approximately as $O(n^3)$, Eq. (5) is much more efficient for larger supercells. This is of interest when using a dense sampling of the BZ.

The dynamical matrix is diagonalised by solving the following eigenvalue problem:

$$\overline{D}^{BZ,UC}(\mathbf{q})\cdot\mathbf{v}(\mathbf{q}, j) = \omega^2(\mathbf{q}, j)\mathbf{v}(\mathbf{q}, j), \quad (6)$$

with $\omega^2(\mathbf{q}, j)$ the jth eigenvalue at wave vector $\mathbf{q}$ and $\mathbf{v}(\mathbf{q}, j)$ the corresponding eigenvector. This eigenvector $\mathbf{v}(\mathbf{q}, j)$ represents the mass-weighted displacement vectors associated with phonon-mode j. From

this it is possible to construct a weighing for each atom:

$$w_k\left(\mathbf{q}, j\right) = \frac{\sum_{\alpha} \left|\mathbf{v}_{\alpha,k}\left(\mathbf{q}, j\right)\right|^2}{\|\mathbf{v}(\mathbf{q}, j)\|^2}, \quad (7)$$

allowing for a partitioning of the phonon DOS. The weighing factors normalize to one as $\sum_k w_k(\mathbf{q}, j) = 1$ [46]. The atom-projected phonon spectrum for atom k at a frequency ν is then calculated as:

$$\omega_k(\nu) = \sum_{j=1}^{3N} \frac{1}{V_{BZ}} \int_{BZ} \omega\left(\mathbf{q}, j\right) w_k\left(\mathbf{q}, j\right)\delta\left(\nu, \omega\left(\mathbf{q}, j\right)\right)d\mathbf{q}, \quad (8)$$

with V_{BZ} the volume of the first Brillouin zone and δ the Dirac delta function. Note that we sum over $3N$ and not $3N - 3$ states. The 3 translation modes, which are zeros at the Γ-point, turn into the 3 non-zero acoustic modes in the remainder of the BZ.

3.2. Differences of spectra

When trying to extract the part of the vibrational spectrum due to a (point-) defect, one may be tempted to take the difference of the full defect system spectrum and a reference spectrum (c.q., the spectrum of the host material). The result will contain clear defect features—such as new peaks outside the host spectrum, and new intense features within the range of the host spectrum. This can provide a reasonable qualitative picture, even though significant noise as well as negative intensities are to be expected. The latter being undesirable during further processing of the spectrum; for example to calculate vibrationally-derived thermal materials properties [47,48].

Correct normalisation with regard to the host spectrum is complicated by the possible difference in number of atoms (e.g., due to an interstitial or vacancy), but also by the question of which atoms belong to the defect (e.g., only the substitutional dopant, or also nearest neighbours?).

To move beyond the qualitative identification of defect-related vibrational states and properties it is necessary to obtain a well-normalised spectrum (i.e., the integrated DOS corresponds to the actual number of states involved in the defect spectrum) as well as an associated listing of defect-contributing atoms. The resulting defect-host partitioning should be independent of the defect system, and provide a quantitative measure for identifying atoms belonging to the defect.

3.3. Isolating the phonon spectrum of a defect

To address the problem of isolating the phonon spectrum of a defect, we start from the inverse problem: Which atoms contribute to the defect spectrum? Although the approach can be extended easily to host materials with multiple inequivalent atomic sites and multiple atomic species, we present the methodology from the perspective of a host material containing only a single atomic species and a single inequivalent atomic position: diamond. In such a system, all atoms are perfectly equivalent, making their contribution to the phonon spectrum identical (i.e., $\frac{1}{n}$ times the total phonon spectrum of a cell containing n atoms). Therefore, the impact of a defect can be identified clearly as the deviation from this reference spectrum.

A straightforward method to quantify this difference is by means of the Root-Mean-Square-Deviation (RMSD) of the two (normalized) spectra:

$$RMSD(k) = \sqrt{\frac{\int_{\nu=0}^{\nu_{max}} (\widetilde{\omega}(\nu) - \widetilde{\omega}_k(\nu))^2 d\nu}{\nu_{max}}}, \quad (9)$$

with ν_{max} the highest frequency of the spectra $\widetilde{\omega}$ and $\widetilde{\omega}_k$, the normalised host and atom spectrum, respectively. In this case, a host atom has
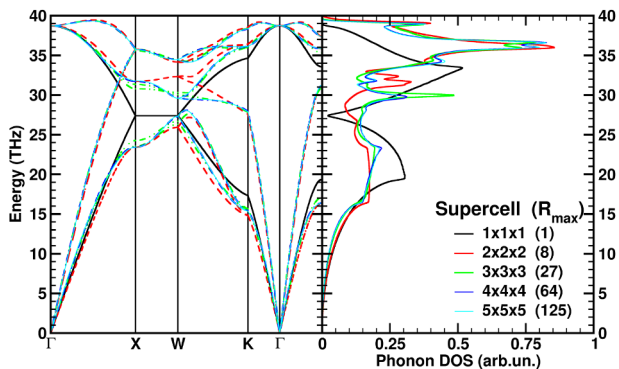

Fig. 2. The vibrational band structure and resulting vibrational density of states (DOS) as obtained with Eq. (5) for pristine bulk diamond. Different color curves are used to show the convergence of the vibrational band structure and DOS with regard to the supercell size. The supercells are constructed starting from the primitive 2-atom unit cell. The supercells contain 2 (1×1), 16 (2×2), 54 (3×3), 128 (4×4) and, 250 (5×5) atoms.

theoretically an RMSD of zero. A defect atom, on the other hand, has a positive non-zero RMSD. However, as this is an unbound function of which the value strongly depends on the shape of the spectra $\widetilde{\omega}$ and $\widetilde{\omega}_k$, the information gained is too limited for our purpose.

Alternately, the overlap of the (normalized) phonon spectrum obtained for the host system and atom k of the defect system presents a bound function with an upper value of 1 (or 100%):

$$\chi_k = \left(\int_\nu \min\left(\widetilde{\omega}(\nu),\ \widetilde{\omega}_k(\nu) \right) d\nu \right) \times 100\%. \tag{10}$$

In this case, a *host atom* shows 100% overlap while a *defect atom* shows a lower value [49]. A value of 0% could in theory be obtained for a defect atom which gives rise only to vibrational contributions outside the host spectrum. The substitutional Eu atom in diamond, which we will discuss later, approaches this theoretical limit with $\chi_{Eu} = 10\%$.

We noted earlier that the vibrational spectrum and DOS for systems with a small unit cell (such as prospective host systems) may require rather large supercells to present a converged picture (*cf.*, Fig. 2). Defects, in contrast, are modelled using large supercells to approximate the experimentally relevant "low" defect concentrations. As a result, longer ranged vibrational interactions are by default incorporated for such systems, leading to a more converged phonon spectrum than is the case for a small host system unit cell (*cf.*, black curve in Fig. 2). It is therefore essential to obtain a sufficiently converged host *reference* spectrum, $\widetilde{\omega}(\nu)$, (*cf.*, the $5 \times 5 \times 5$ spectrum in Fig. 2)) to avoid artificial overlap mismatch when calculating χ_k.

Fortunately, this mismatch can be quantified by calculating the convergence of χ_k of the host spectrum itself. In the case of diamond, this is shown in Fig. 3. E.g., the χ_k of a 2-atom diamond system gives a mismatch of about 20%. As such, a "host atom" in a defective system, when using the 2-atom reference data, would have a χ_k of about 80%, instead of the theoretical maximum of 100%. This is seen in Fig. 4, showing $\chi_{C,host}$ of the host C atoms in the defect systems to be in the range of 79–84% for the $1 \times 1 \times 1$ reference spectrum. As the computational cost—even at the quantum-mechanical level—of obtaining well converged reference host spectra is not extreme when taking advantage of (translational) symmetry, we consider this not to be a limitation in the practical application of the presented method. We therefore assume in the following that the reference spectra are sufficiently converged.

For the "host atoms" in defect systems, long ranged interactions impact their expected χ_k as well. As can be seen in Fig. 4, this leads to a leveling of the host atom χ_k as function of the reference spectra used.

For the point-defects modeled with a 64 atom conventional supercell (which is relatively small for a defect cell), the χ_k of the host atoms
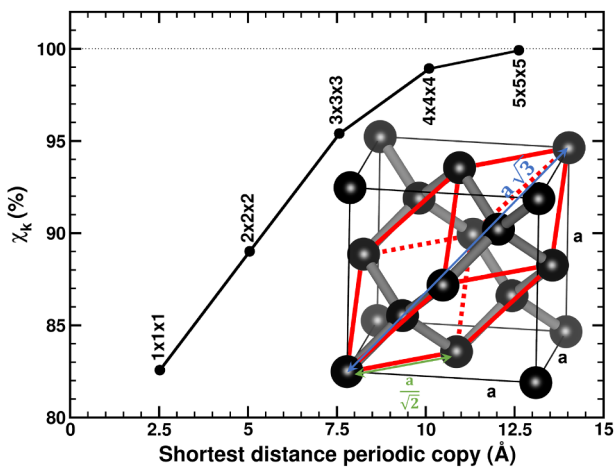

Fig. 4. Convergence of χ_C of C host atoms in the different point-defect systems as function of supercell used in creating the reference host spectrum. The overlap is calculated as the average of the overlap χ_C for all atoms –5.0 Å removed from the center of the point-defect in the 64 atom conventional supercells. The empty red squares show the values obtained for the C_i point-defect in a 216 atom conventional supercell, for comparison.

ranges between 89 and 92%, which is consistent with the convergence of the host atom overlap in the primitive $2 \times 2 \times 2$ cell (*cf.*, Fig. 3). For larger defect cells, the overlap of the host atoms increases, as can be seen for the example of the C_i point-defect (*cf.*, Fig. 4). Placing the point-defect in a conventional $3 \times 3 \times 3$ supercell, χ_k of the host atoms increases to about 95%. This, in turn, is consistent with the convergence of the reference host atom in the primitive $3 \times 3 \times 3$ cell (*cf.*, Fig. 3). More interestingly, this increase is not due to a gradual increase in χ_k for atoms ever farther away from the point-defect, but rather due to a general upward shift of the overlap of the *non-defect* atoms, as can be seen in Fig. 5.

In contrast, χ_k remains the same for the defect atoms, indicating χ_k of defect atoms to rapidly converge with regard to system size. As such, χ_k is shown to be a useful measure to effectively determine the atoms belonging to the defect, and even to which degree. This allows for the

Fig. 3. Convergence of the diamond vibrational spectrum as function of the supercell size, using χ_k as quality measure. The vibrational DOS of the $5 \times 5 \times 5$ supercell is used as reference.
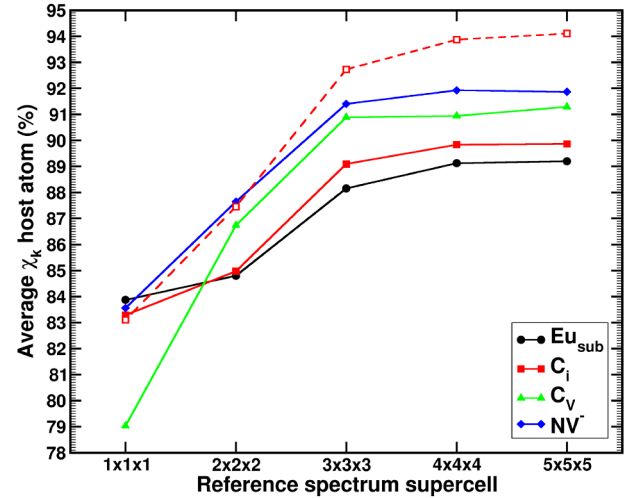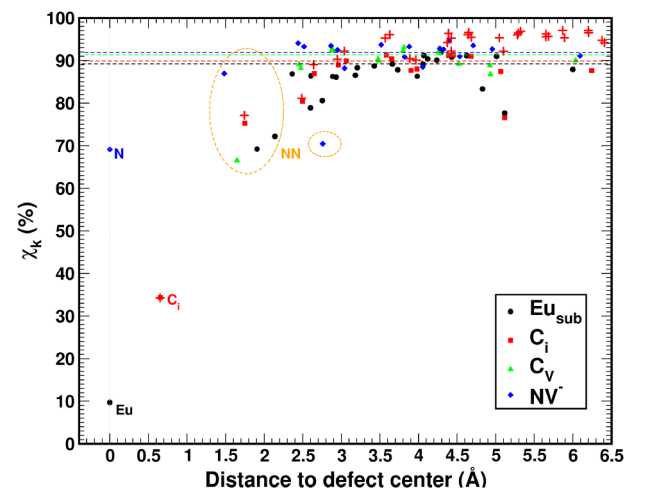

Fig. 5. The overlap χ_k for each atom k of the diamond point-defect systems modeled using a small $2 \times 2 \times 2$ conventional cell (64 atoms). The atoms forming the defect centers are indicated, as well as their nearest neighbours (NN). The horizontal dashed lines show the average χ_k value found for atoms in a range of 3–5 Å from the defect center for the $2 \times 2 \times 2$ defect cells. The red crosses show the result for a C_i point-defect using a $3 \times 3 \times 3$ conventional cell (216 atoms), for comparison.

efficient calculation of the fragment spectra of a defect in a larger supercell system. Furthermore, the χ_k values also allow for the systematic improvement of such fragment spectra (*cf.* below and Fig. 7) without the need for *ad hoc* selection [35]. Instead of defining a purely spatial threshold function, the threshold for inclusion can now be directly related to the atom's contribution to the defect spectrum. Alternately, it is also possible to directly construct a defect spectrum from the atom projected spectra, using χ_k for selection purposes. The resulting spectrum will contain all (defect) features missing in the host spectrum. More importantly, this defect spectrum also contains contributions due to the interaction of the defect with the host lattice. These are incorporated through the (defect-) atom projection of (delocalized) host system modes. Furthermore, the partitioning of the system into a host and defect fraction could be used as a platform to calculate derived thermodynamic contributions due to the defect (which goes beyond the scope of the current work). Also, due to the small size of the fragment to consider, one could more easily move beyond the standard harmonic approximation [17,50,51].

4. Diamond based defects

To evaluate our method, four different point-defects in diamond are considered.

- The substitutional Eu dopant in diamond (Eu_{sub}): This heavy lanthanide dopant gives rise to low lying atomic phonon bands with a clearly distinguishable peak in the phonon spectrum [15].
- The $\langle 001 \rangle$ split interstitial (C_i): This intrinsic point-defect places two C atoms at a single site. It provides a local breaking of the symmetry with limited change in the chemical environment. As a result, two very distinct optical phonon peaks are created well above the bulk spectrum.
- The neutral C vacancy (C_V): This intrinsic point-defect is obtained by removing a single C atom, and as a result, it resembles pristine diamond most closely [38].
- The negatively charged nitrogen-vacancy centre (NV^-): One of the most discussed and studied point-defects in diamond. This defect presents a combination of a substitutional dopant and a carbon vacancy. Due to a mass comparable to that of C, the N atom gives rise to a spectrum comparable to that of C itself, making it challenging to extract while being of great interest for applications.

Our choice to consider point-defects in this section, is rooted in their intrinsic simplicity for practical purposes. *E.g.*, it straightforward to define a distance measure with regard to a point-defect. However, we expect our method to be equally well suited to deal with more complex defects such as defect-clusters. One could even image its use for extended defects such as defect lines, or interfaces such as grain boundaries or surfaces. Of course in the later cases, one will require larger cells to deal with the larger sizes of the defects. The application of this method to surfaces poses some challenges as well as promising capabilities. One of the challenges lies in the comparison of a 2D surface Brillouin zone to that of a 3D bulk Brillouin zone. Notable exceptions being the true 2D materials such as graphene and 2D-dichalcogenides [52–55]. A promising capability would be the fingerprinting surface defects. But this method could also be used to fingerprint surface reconstructions such as dimer-row reconstructions or the formation of nanowire arrays, providing a new tool for corroborative modeling of experimental systems [56,57].

These examples, although very interesting in their own right, go beyond the scope of the current work.

4.1. Defect phonon spectra

The overlap χ_k is calculated for each atom in the defect systems. In Fig. 5, χ_k is shown as function of the distance of atom k to the center of
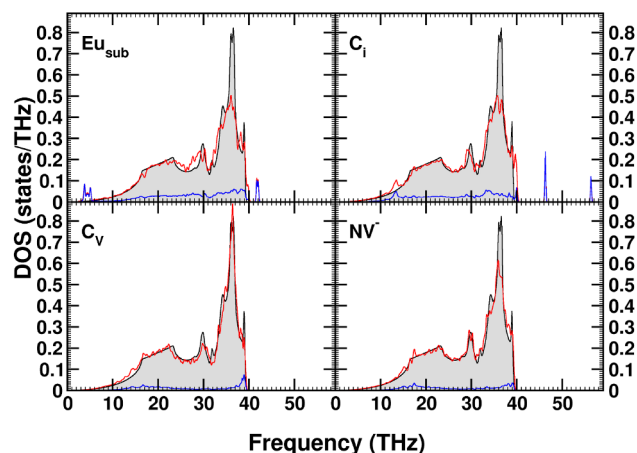

Fig. 6. The phonon spectra of the four diamond point-defects (red curves) and the partial phonon spectrum due to the point-defect (blue curve) with a threshold of $\chi_k < 85\%$. The bulk diamond spectrum (black curve) is given as reference.
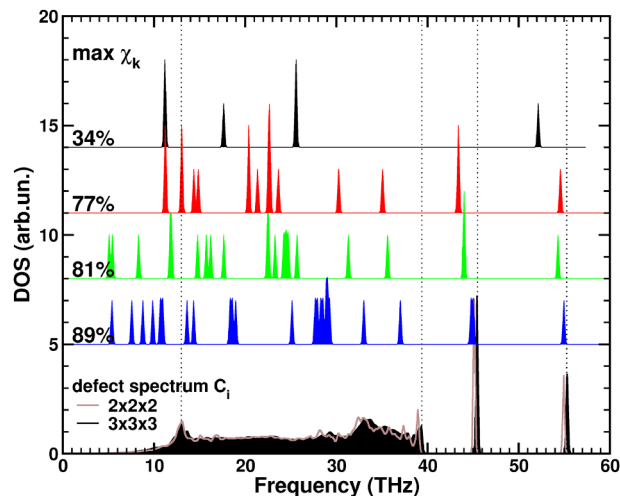

Fig. 7. The phonon spectrum of the C_i point-defect as obtained using a 65 ($2 \times 2 \times 2$) and a 217 ($3 \times 3 \times 3$) atom supercell, with $\chi_k < 85\%$. Fragment spectra obtained using only atoms with $\chi_k < 34, 77, 81,$ and 89% in the $3 \times 3 \times 3$ system, are shown in comparison. Dotted lines indicate the position of specific defect spectrum features.

Table 1
CPU time (days) required to generate the Hessian matrix of the C_i point-defect in diamond, using first-principles quantum-mechanical calculations.

System	CPU time (days)
$3 \times 3 \times 3$ full spectrum	3322
$2 \times 2 \times 2$ full spectrum	137
Fragment 34%	32
Fragment 77%	83
Fragment 81%	138
Fragment 89%	251

the point-defect. The atoms at the center of the point-defect are indicated, as well as the nearest neighbour (NN) atoms. In Fig. 6, the phonon spectrum of the different point-defects is presented in comparison to the phonon spectrum of pristine diamond. Both the Eu_{sub} and the C_i point-defect give rise to clearly distinguishable phonon peaks, which show little to no overlap with the host phonon spectrum. This results in very low χ_k values for Eu and the interstitial C atoms, as is

seen in Fig. 5. The shells of NN and next-NN atoms already show rather large χ_k values of 70–80%, indicating that their atom-projected spectra are harder to distinguish from that of a host atom, but still clearly different. In contrast, the C_V point-defect shows a phonon spectrum quite similar to that of the host system, making it hard to indicate the differences and their sources. However, looking at Fig. 5, four atoms stand out clearly with a $\chi_k \approx 70\%$: the four atoms surrounding the vacancy. Next-NN C atoms present a converged host character, showing this point-defect to be strongly localized on the vacancy and its four surrounding atoms. Considering the projected phonon DOS associated with these atoms (blue curve in the bottom left panel of Fig. 6), it becomes clear that the defect spectrum consists of a peak at the high end of the spectrum, and two peaks around 14.5 and 17 THz. Turning our attention to the NV^- point-defect, we notice in Fig. 5 that for the N atom, as well as for the three C atoms surrounding the vacancy, $\chi_k \approx 70\%$, similar as for the C_V point-defect. The NN C atoms surrounding the N atom, on the other hand, present a high χ_k associated with host atoms. The defect spectrum is also quite similar to that of the C_V point-defect, with additional peaks in the range of 14–17 THz. However, in contrast to the C_V point-defect, the peak at the high end of the spectrum is much less pronounced.

4.2. Comparison to the fragment spectrum: the $\langle 001 \rangle$ split interstitial

In Fig. 7 the defect spectrum of the C_i point-defect is shown, obtained in both a *smaller* $2 \times 2 \times 2$ (brown curve) and *larger* $3 \times 3 \times 3$ (black) diamond supercell. This shows that the $2 \times 2 \times 2$ supercell is sufficient to construct a well-converged defect spectrum. The two optical peaks at about 45 THz and 55 THz are found to be within 0.5 THz of the results obtained with the $3 \times 3 \times 3$ supercell, while the feature at 13 THz shows no visible deviation. Furthermore, the broad band, due to defect-host system interactions, is well converged. It is important to note that the computational cost for generating the first-principles Hessian matrix within a periodic plane waves approach (shown in Table 1), for the $2 \times 2 \times 2$ supercell is 24× lower than for the larger supercell, making this a cost-efficient approach.

The defect spectrum is also compared to different fragment spectra obtained using the $3 \times 3 \times 3$ supercell. The atoms belonging to the fragment are determined using their χ_k value: $\chi_k \leqslant 34\%$ (2 atoms), 77% (6 atoms), 81% (10 atoms), and 89% (18 atoms). The resulting defect spectra obtained using the fragment approach are shown in Fig. 7. All fragments (except the smallest 2-atom fragment) give rise to the two optical modes, and it is only the largest fragment which positions them with an accuracy comparable to the $2 \times 2 \times 2$ defect spectrum (at almost twice the computational cost). More interestingly, the feature at 13 THz is not retrieved in the fragment spectra, neither is the broad interaction band.

5. Conclusions

In this work, a method is presented for determining the phonon-spectrum of a *defect* using relatively small periodic first-principles calculations. Our method provides a quantitative measure for assigning atoms to a defect. This allows it to be used in tandem with a fragment approach to efficiently obtain incrementally more accurate fragments in much larger supercells. Alternately, combining the atom projected vibrational spectra of the *defect* atoms gives rise to a quickly converging defect spectrum which combines the defect specific features of the spectrum with the contributions due to defect-host interactions. The resulting partitioning of the system spectrum into a host and defect component opens up the possibility for similar partitioning of properties derived from the phonon spectrum, which is the subject of ongoing research. The presented methods have been implemented in the HIVE package [58].

CRediT authorship contribution statement

Danny E.P. Vanpoucke: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Bonnot, Surf. Coat. Technol. 45 (1991) 343.
[2] C.R. Jacob, M. Reiher, J. Chem. Phys. 130 (2009) 084106 .
[3] Z.V. Zivcova, O. Frank, S. Drijkoningen, K. Haenen, V. Mortet, L. Kavan, RSC Adv. 6 (2016) 51387.
[4] D. Law, D. Zhou, Y. Qiu, Y. Chen, G.G. Zhang, L. Yu, R.V. Mantri (Eds.), Developing Solid Oral Dosage Forms, second ed., Academic Press, Boston, 2017, pp. 59–84 ISBN 978-0-12-802447-8.
[5] M. Corva, A. Ferrari, M. Rinaldi, Z. Feng, M. Roiaz, C. Rameshan, G. Rupprechter, R. Costantini, M. Dell'Angela, G. Pastore, et al., Nat. Commun. 9 (2018) 4703.
[6] D.G. Baranov, M. Wersäll, J. Cuadra, T.J. Antosiewicz, T. Shegai, ACS Photonics 5 (2018) 24.
[7] B.M. Weckhuysen, Z. Ozturk, R. Brand, J. Boereboom, F. Meirer, Chem.: Eur. J. 25 (2019) 8070.
[8] Y. Ozaki, Bull. Chem. Soc. Jpn 92 (2019) 629.
[9] G. Roma, Model. Simul. Mater. Sci. Eng. 27 (2019).
[10] W. Windl, P. Pavone, K. Karch, O. Schütt, D. Strauch, P. Giannozzi, S. Baroni, Phys. Rev. B 48 (1993) 3164.
[11] M. Lazzeri, S. de Gironcoli, Surf. Sci. 402–404 (1998) 715.
[12] G. Roma, Physica Status Solidi (a) 213 (2016) 2995.
[13] C. Crépisson, M. Blanchard, M. Lazzeri, E. Balan, C. Sanloup, Geochim. Cosmochim. Acta 222 (2018) 146.
[14] S. Salustro, A.M. Ferrari, F.S. Gentile, J.K. Desmarais, R. Rérat, R. Dovesi, J. Phys. Chem. A 122 (2018) 594.
[15] D.E.P. Vanpoucke, S.S. Nicley, J. Raymakers, W. Maes, K. Haenen, Diam. Relat. Mater. 94 (2019) 233.
[16] G. Petretto, X. Gonze, G. Hautier, G.-M. Rignanese, Comput. Mater. Sci. 144 (2018) 331.
[17] P.T. Panek, C.R. Jacob, J. Phys. Chem. Lett. 7 (2016) 3084.
[18] T.Q. Teodoro, M.A.J. Koenis, S.E. Galembeck, V.P. Nicu, W.J. Buma, L. Visscher, J. Phys. Chem. Lett. 9 (2018) 6878.
[19] K. Kunc, R.M. Martin, Phys. Rev. Lett. 48 (1982) 406.
[20] M. Lazzeri, P. Thibaudeau, Phys. Rev. B 74 (2006) 140301 .
[21] A. Togo, L. Chaput, I. Tanaka, Phys. Rev. B 91 (2015) 094306 .
[22] E. Londero, G. Thiering, L. Razinkovas, A. Gali, A. Alkauskas, Phys. Rev. B 98 (2018).
[23] H. Lachenani, A. Larabi, N. Gabouze, Silicon 11 (2019).
[24] S. Ahmad, I. Georgieva, M. Hanif, M. Monim-ul Mehboob, S. Munir, A. Sohail, A.A. Isab, J. Mol. Mod. 25 (2019).
[25] E.T. Sato, H. Martinho, Biomed, Opt. Express 9 (2018) 1728.
[26] C. Herrmann, J. Neugebauer, M. Reiher, New J. Chem. 31 (2007) 818.
[27] P. Bouř, T.A. Keiderling, J. Chem. Phys. 117 (2002) 4126.
[28] J.D. Head, Int. J. Quantum Chem. 65 (1997) 827.
[29] A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, M. Waroquier, J. Chem. Phys. 126 (2007) 224102 .
[30] R. Terrett, R. Stranger, T. Frankcombe, R.J. Pace, Phys. Chem. Chem. Phys. 19 (2017) 6654.
[31] P. Bouř, J. Sopková, L. Bednárová, P. Maloň, T.A. Keiderling, J. Comp. Chem. 18 (1997) 646.
[32] S. Yamamoto, X. Li, K. Ruud, P. Bouř, J. Chem. Theory Comput. 8 (2012) 977.
[33] S. Yamamoto, M. Miyada, H. Sato, H. Hoshina, Y. Ozaki, J. Phys. Chem. B 121 (2017) 1128.
[34] M.W.D. Hanson-Heine, F.S. Husseini, J.D. Hirst, N.A. Besley, J. Chem. Theory Comput. 12 (2016) 1905.
[35] F. Pascale, S. Salustro, A.M. Ferrari, M. Rerat, P. D'Arco, R. Dovesi, Theor. Chem. Acc. 137 (2018).
[36] G. Kresse, D. Joubert, Phys. Rev. B 59 (1999) 1758.
[37] J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77 (1996) 3865.

[38] D.E.P. Vanpoucke, K. Haenen, Diam. Relat. Mater. 79 (2017) 60.

[39] M. Born, K. Huang, Dynamical Theory of Crystal Lattices, International series of monographs on physics, Clarendon Press, 1988.

[40] J. Stouten, D.E.P. Vanpoucke, G. Van Assche, K.V. Bernaerts, Macromolecules 53 (2020) 1388–1404, https://doi.org/10.1021/acs.macromol.9b02659.

[41] G.J. Ackland, M.C. Warren, S.J. Clark, J. Phys.: Condens. Matter 9 (1997) 7861.

[42] Although charged defects require very large cells (> 20 Å) to obtain converged results with regard to defect energies, the qualitative picture presented by the overlap χ_a is obtained already for rather small cells, as shown by the example of the negatively charged NV center.

[43] D.E.P. Vanpoucke, K. Lejaeghere, V.V. Speybroeck, M. Waroquier, A. Ghysels, J. Phys. Chem. C 119 (2015) 23752.

[44] In this specific case we are referring to a supercell of the defect system. I.e. a cell containing multiple periodic copies of the defect. In such a system, we can use Eq. (4) directly on the dynamical matrix as a whole. In contrast, as it is a supercell of the defect system, the dynamical matrix of the whole supercell can be split into smaller matrices related to constituent unitcells. These will then serve as the matrices summed over in Eq. (5).

[45] X. Qian, R. Yang, Phys. Rev. B 98 (2018) 224108 .

[46] In numerical packages, the obtained eigenvectors, from for example a singular value decomposition, will be normalized by default.

[47] I. Rosenblum, J. Adler, S. Brandon, Comput. Mater. Sci. 12 (1998) 9.

[48] S.P. Adiga, V.P. Adiga, R.W. Carpick, D.W. Brenner, J. Phys. Chem. C 115 (2011) 21691.

[49] As defects generally slighly modify the lattice parameters in DFT calculations, they also lead to a very slight shift of the spectral lines. To deal with this miss-alignment we first scale the host-or defect-system (full system) spectrum to maximize their overlap. This fitted scaling factor is then applied on the atom-projected spectra before the calculation of the overlap χ_k.

[50] G.M. Piccini, J. Sauer, J. Chem. Theory Comput. 10 (2014) 2479.

[51] K.D. Wispelaere, L. Vanduyfhuys, V.V. Speybroeck, C.R.A. Catlow, V.V. Speybroeck, R.A. van Santen (Eds.), Modelling and Simulation in the Science of Micro- and Meso-Porous Materials, Elsevier, 2018, pp. 189–228.

[52] S. Sakong, P. Kratzer, J. Chem. Phys. 133 (2010) 054505 .

[53] J. Pesic, V. Damljanovic, R. Gajic, K. Hingerl, M. Belic, EPL 112 (2015) 67006.

[54] J. Ibanez, T. Wozniak, F. Dybala, R. Oliva, S. Hernandez, R. Kudrawiec, Sci. Rep. 8 (2018) 12757.

[55] P. Serafini, A. Milani, M. Tommasini, C. Castiglioni, C.S. Casari, Phys. Rev. Mater. 4 (2020) 014001 .

[56] D.E.P. Vanpoucke, G. Brocks, Phys. Rev. B 81 (2010) 035333 .

[57] D.E.P. Vanpoucke, J. Phys.: Cond. Matter 26 (2014) 133001 .

[58] D.E.P. Vanpoucke, HIVE v4.x, 2020.https://github.com/DannyVanpoucke/HIVE4-tools.