The Infrared spectrum of very large (periodic) systems: global versus fragment strategies-the case of three defects in diamond

Fabien Pascale, Simone Salustro, Anna Maria Ferrari, Michel Rérat, Philippe D'arco, Roberto Dovesi

Noname manuscript No.
(will be inserted by the editor)

The Infrared spectrum of very large (periodic) systems. Global vs. fragment strategies. The case of three defects in diamond.

Fabien Pascale · Simone Salustro · Anna Maria Ferrari · Michel Rérat · Philippe D'Arco · Roberto Dovesi

the date of receipt and acceptance should be inserted later

E-mail: fabien.pascale@univ-lorraine.fr September 19, 2018

Abstract The calculation of the full vibrational spectrum (Infrared or Raman) of very large systems (say larger than one thousand atoms) is not only very expensive, but also of relatively low interest, as in many (most of the) cases only a subset of modes, well separated from the large, diffuse bands resulting from the superposition of thousands of peaks, is used for the spectroscopic characterization of the specific system under study. Here a *fragment* strategy, consisting in computing and diagonalizing a reduced (in size) Hessian matrix centered around the zone of interest, is illustrated, and its accuracy and efficiency documented, by comparison with the *full Hessian diagonalization* (FHD) scheme. Three test cases are considered, showing different vibrational features. They are defects in diamond: the VN_3H defect (V stands for the vacancy), where the interesting point is the characterization of the bending and stretching modes of H, well separated from the large band resulting from the perturbation of the diamond manifold; the VH_4 defect (four H atoms in the vacancy, with vibrational modes related to H appearing both at high and low wavenumbers); the I_{2N} interstitial defect, with modes in which the N atoms are involved, appearing at wavenumbers not far from the manifold of the perfect diamond modes. So the three cases, apparently similar, explore three different situations of interest for the *fragment* strategy: i) localized modes very well separated from the large diamond continuous band (VN_3H); ii) modes at upper border of the large diamond continuous band (I_{2N}); a case in which the modes of interest appear both as separated from, and merged with, the large continuous band (VH_4).
It turns out that in all cases relatively small fragments, containing from 2 to 40 atoms, permit to reproduce with high accuracy (the difference with respect to the *FHD* being always smaller than 5 cm^{-1} for the wavenumbers, and a few percent for the IR intensity) the spectral feature(s) of interest, at a computational cost that is only a small fraction of the one required by the FHD.

Fabien Pascale (E-mail: fabien.pascale@univ-lorraine.fr)
Laboratoire de Physique et Chimie Théoriques, CNRS, UMR 7019. Vandœuvre-lès-Nancy, 54506 France

Simone Salustro · Anna Maria Ferrari · Roberto Dovesi
Dipartimento di Chimica and NIS (Nanostructured Interfaces and Surfaces) Centre, Università di Torino, via Giuria 5, I-10125 Torino, Italy

Michel Rérat
CNRS / Université de Pau et des Pays de l'Adour, IPREM, UMR5254, 64000 Pau, France

Philippe D'Arco
Institut des Sciences de la Terre de Paris (UMR 7193 UPMC-CNRS), UPMC, Sorbonne Universités, Paris (France)

1 Introduction

The frontier separating the domains of systems that can be treated at the quantum mechanical level, or with semi-empirical methods or classical force fields or parametrized molecular dynamics, is rapidly moving, thanks to the rapid progress of hardware and (the less rapid) progress of software. Quantum mechanical codes (see for example Refs. [1, 2, 3, 4]), based mostly on the various flavours of DFT, are becoming more general and more efficient in general, although large differences exist among them with respect to many features (one example is the degree of parallelism of these codes). The problem of the feasibility of calculations for very large systems and of the related numerical accuracy, discussed here, has been a constant concern of János Ángyán, to which this Special Issue is dedicated. János devoted to this subject general and important publications [5]. A second area of common interest with János is related to the performance of hybrid functionals, that have been constantly used by the

present authors for many years, also when they were not very popular in the solid state community. János published many illuminating papers [6, 7] devoted to this subject.

In the investigation of crystalline solids, in spite of the fact that a large fraction of the literature is still focused on relatively simple and low-cost properties, like the band gap and the density of states, for which a single SCF calculation is required, the interest and the computational effort is expected to move towards more expensive (and more interesting, in the authors' opinion) observables, such as the physical tensors (elastic, dielectric, piezoelectric, photoelastic, hyperpolarizability) and the vibrational properties (wavenumbers, Infrared and Raman intensities). Some of the present authors have shown that it is possible [8], at relatively low computational cost, to obtain the full set of vibrational wavenumbers of quite large unit cells (say 1000-3000 atoms), and with a relatively small additional cost, to build the full IR spectrum of these systems. The Raman spectrum is, at this stage, more expensive, but can be obtained for unit cells of up a few hundredth atoms [8]. The rapid increase of the power of supercomputers, the parallel structure of the CRYSTAL code [9] to which we are referring to, the availability of a Multi-Task option, are features that will permit in short to target the vibrational spectrum of systems in the range of 10.000 atoms, if a supercomputer is available (see for example the European projects in this domain at `http://www.prace-ri.eu/`).

However, a question might be raised about the need of a complete determination of the vibrational spectrum of systems of such dimensions, in particular for comparing IR and Raman spectra. Ten thousand atoms generate three times more vibrational modes, that would generate a continuous band in many part of the explored wavenumber range. Such a continuum band is essentially useless from the point of view of the characterization of the system (obviously this is not the case for the thermodynamic properties).

This approach is based on the hypothesis that all the modes of interest (usually separated in energy from the large band(s) characterizing the huge system), are local in character, so that the vibrational eigenvectors involve only atoms within a given radius from *the center* of the zone of interest (in the present examples: the point defect region).
The problem of the *partition* of the very expensive calculation of frequencies and IR intensities of very large molecular systems (and polymers, in some cases) has been tackled in recent years in a series of publications [10, 11, 12], that are however formulated along lines that differ from the present ones. One additional major difference is that the present scheme is applicable to three dimensional periodic defective solids.
The *fragment* strategy should not be confused with the cluster approach, in which from the beginning a subset of atoms is extracted from the whole (infinite) system. In this latter case many problems (border termination, saturation, loss of the infinite nature of the system) make the model more delicate and doubtful (see for example Ref.[13]).
The accuracy and efficiency of the *fragment* scheme is discussed by comparison with the *full Hessian diagonalization (FHD)* scheme with reference to three cases. They all refer to defects in diamond, for which one of the crucial variables is the dilution of the defect in the infinite solid. The defects are investigated with supercells (a periodic replica of a large unit cell containing the defect at the center); the larger is the supercell, the smaller is the concentration of the defect and the lateral interaction among defects. The first defect is VN_3H [14] (V stands for the vacancy, surrounded by three N and one C atoms; the latter is saturated with a H atom); here the interest is in the characterization of the bending and stretching modes of the hydrogen atom, that are well separated from the large band resulting from the perturbation of the diamond manifold; the second is the I_{2N} interstitial defect [15], with the modes generated by the two nitrogen atoms appearing at wavenumbers not far from the manifold of perfect diamond; the third is VH_4 [16] (four H atoms in the vacancy), with vibrational modes related to H appearing both at high and low wavenumbers.
All the discussed cases show that relatively small fragments, containing from say 2 (in the limiting case of the CH stretching) to 40 atoms, permit to reproduce with high accuracy (the difference with respect to the FHD being always smaller than 5 cm^{-1}) the spectral feature(s) of interest, at a computational cost that is only a small fraction of the one required by FHD .
The structure of the paper is the following: in section 2 the adopted model and the computational parameters are defined. In section 3 the IR vibrational spectrum (wavenumbers and intensities) obtained with fragments of increasing size are compared with the ones obtained from the diagonalization of the complete Hessian of the system (*FHD*). In section 4 the cost of the various steps of the calculation for the two strategies are compared. Finally, in section 5, some conclusions are drawn.

2 Computational Methods

Calculations have been performed by using the B3LYP global hybrid functional [17, 18] as implemented in the CRYSTAL program [1].

An *all electron* basis set of Gaussian-type functions has been adopted (Pople's 6-21G [19]) for carbon and nitrogen atomic species; the exponent of the most diffuse *sp* shell is 0.23 (C) and 0.30 (N) $Bohr^{-2}$. The (6)-31G basis set [20] has been used for hydrogen. In order to simulate the presence of a vacancy in the defective structure, a complete removal of the atom of interest (nucleus, electrons, basis set) has been performed.

The Coulomb and exchange infinite lattice series are controlled by five parameters, T_i, which have been set to 8

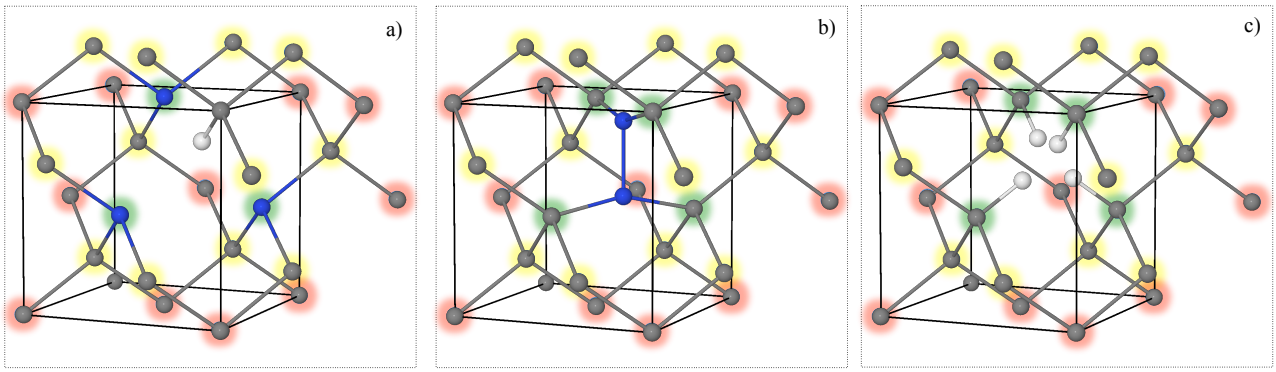

Fig. 1 Conventional cell of diamond where the defective site is shown. Nitrogen is blue, hydrogen is white. Local fragments for VN_3H (a), I_{2N} (b) and VH_4 (c) systems are indicated. The small fragment (SF) involves atoms not highlighted; the medium fragment (MF) involves also atoms highlighted in green; the large fragment (LF) includes also atoms highlighted in yellow. The number of atoms included are 2, 5 and 17 in (a), 2, 6 and 18 in (b) and 4, 8 and 20 in (c). The big fragment (BF), not shown in the figure, includes one further set of neighbors, for a total number of 41 for VN_3H, 42 for I_{2N} and 44 for VH_4.

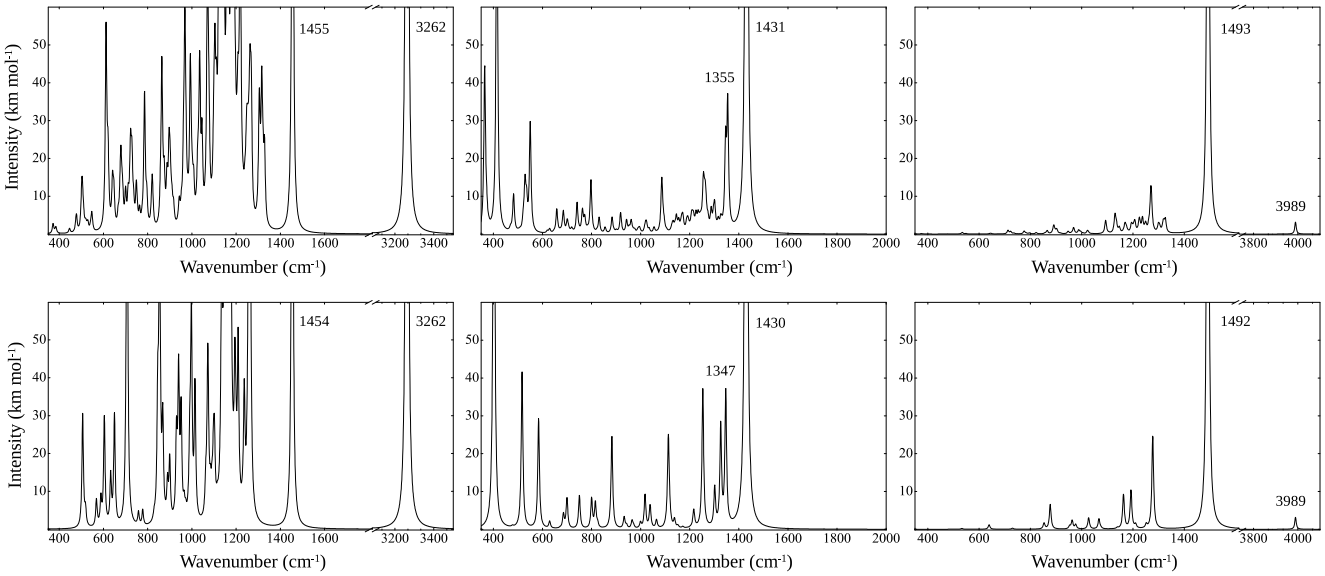

Fig. 2 B3LYP simulated IR spectra of the VN_3H (left), I_{2N} (center) and VH_4 (right) defects obtained with the FHD (top) and the *fragment* (bottom) strategies. The big fragment, BF, has been used. Values refer to the S_{216} supercell.

(T_1-T_4) and 16 (T_5). The convergence threshold on energy for the self-consistent-field (SCF) procedure has been set to 10^{-8} hartree for structural optimizations, and to 10^{-10} hartree for vibration frequency calculations.

The DFT exchange-correlation contribution and its gradient are evaluated by numerical integration over the unit cell volume. The generation of the integration grid points in CRYSTAL is based on an atomic partition method, originally developed by Becke [21] for molecules and furtherly extended to periodic systems. Within this scheme the unit cell is partitioned into atomic volumes centered on the nuclei, where each point is associated to a weight. Radial and angular points for the integration grid are generated through Gauss-Legendre radial quadrature and Lebedev two-dimensional angular point distributions. The choice of a suitable grid is crucial both for numerical accuracy and need of computational resources. In this study the default [22] pruned grid with 75 radial and 974 angular points has been used, whose accuracy can be measured by comparing the integrated charge density of N_i= 1294.030 for the VH_4 supercell containing 216 atoms, with the total number of 1294 electrons in the unit cell. As anticipated before, a periodic supercell approach is used in order to simulate different defect concentrations. In this work two cubic supercells have been considered, containing respectively 64 (S_{64}) and 216 (S_{216}) atoms. A Γ-centered Pack-Monkhorst grid [23] for sampling the reciprocal space

has been used, consisting of 4×4×4=64 (S_{64}) and 2×2×2=8 (S_{216}) **k**-points in the First Brillouin Zone.

2.1 Harmonic frequencies and the IR spectra

Frequencies at the Γ point are obtained within the harmonic approximation by diagonalising the mass-weighted Hessian matrix, W, whose elements are defined as

$$W^{\Gamma}_{\alpha i,\beta j} = \frac{H^{\mathbf{0}}_{\alpha i,\beta j}}{\sqrt{M_\alpha M_\beta}} \quad \text{with} \quad H^{\mathbf{0}}_{\alpha i,\beta j} = \left(\frac{\partial^2 E}{\partial u^{\mathbf{0}}_{\alpha i} \partial u^{\mathbf{0}}_{\beta j}} \right) , \quad (1)$$

where M_α and M_β are the masses of atoms associated with the i and j atomic coordinates.

Energy first derivatives with respect to the atomic positions, $v_{\alpha,j} = \partial E / \partial u_{\alpha,j}$, are calculated analytically for all the $u_{\alpha,j}$ coordinates (E is the total energy, $u_{\alpha,j}$ is the displacement coordinate with respect to the equilibrium, α labels the atoms), whereas second derivatives at $\mathbf{u} = \mathbf{0}$ are calculated numerically using a single displacement along each coordinate (the central point and one point on the right side of the parabola):

$$\left[\frac{\partial v_{\alpha j}}{\partial u_{\beta i}} \right] \approx \frac{v_{\alpha j}(0,\dots,u_{\beta i},\dots) - v_{\alpha j}(0,\dots,0,\dots)}{u_{\beta i}} \quad (2)$$

Previous calculations [24] have shown that in bulk systems the influence of both u and N is very small (less than 1 cm^{-1}) when H atoms are not present. For the C-H, N-H and O-H stretching modes anharmonicity is very large [25]. In the present discussion, however, as we are interested in the *wavenumber difference* between the FHD and *fragment* approaches, the amount of anharmonicity is irrelevant, as frequencies are evaluated in exactly the same way for the two schemes.

In order to limit the frequency calculations only to a moiety of the system, the `FRAGMENT` option, available in the CRYSTAL code, has been used. It permits to select the atoms to be considered in the construction of the *reduced* Hessian in equations 1 and 2. With reference to the diamond unit cell reported in the first panel of Figure 1, in this work four different fragments of increasing size have been considered. The smallest fragment (SF) (see Figure 1) includes only the atoms involved in the defect, not highlighted in the figure. The medium (MF), large (LF) and big (BF) fragments (the latter is not reported in the figure) , include the first (4 atoms), second (12) and third (24) shells of neighbors of the defect.

2.2 The Berry phase scheme for the IR intensity

IR intensities have been evaluated by using a computational scheme [26] based on the Berry Phase [27]. The derivatives of the dipole moment with respect to the cartesian coordinates of the unit cell atoms are evaluated numerically, through a scheme similar the the one used for the numerical derivative of the gradient for obtaining the Hessian matrix (see equation 2). It is on the basis of this similarity that it is possible to adopt a *fragment* strategy for the evaluation of the IR intensity. The crucial step, that consists in evaluating the Berry phase difference between the reference geometry and the one in which one atom has been displaced along x, y or z, can be limited to the subset of atoms belonging to the *fragment*.
As regards Raman intensities, only an analytical scheme is available at present in the CRYSTAL code, that does not fit the *atoms of the fragment only* logic described above. This means that the costs of the Raman intensity for the complete unit cell and for the *fragment* are the same, with no advantages then for the latter.
Schemes implying the numerical derivative of the polarizability (evaluated through the CPHF algorithm) with respect to the cartesian coordinates of the atoms (as in equations 2) are possible, that would permit to extend the *fragment* strategy also to Raman intensities. As they are not yet implemented in the code, in the following the *fragment* strategy will be documented for frequencies and IR intensities only.

3 Results

Let us summarize the situation of the IR and Raman spectra for the various cases here considered. The starting point is the perfect IR diamond spectrum, that is completely flat in the full wavenumber range, for symmetry reasons. The corresponding Raman spectrum shows a single peak at 1332 cm^{-1} (experimental value; the calculated peak with the present basis set and functional is at 1317 cm^{-1}).
When the vacancy is inserted (VN_3H and VH_4) and the symmetry is reduced (VN_3H and I_{2N}), many *pure diamond* IR peaks, that are forbidden in the perfect system, become visible, and form a large band spanning from about 400 to 1330 cm^{-1}. The defect modes, that appear above this upper limit as isolated peaks, are reported in Table 1. These modes are easily identified as they appear *above* 1330 cm^{-1}. This attribution is confirmed by looking at the graphical animation of the modes, shown at `http://www.crystal.unito.it`. This latter tool permits also to identify some of the modes that appear at the lower extreme of the pure diamond band; these modes (at 414 cm^{-1} for I_{2N} and at 342 cm^{-1} for VH_4 for the S_{216} supercell), that have certainly a much larger collective nature, are also reported in the table, in order to include in the analysis also cases that are difficult or *impossible* to be tackled with the *fragment* strategy (a fragment obviously cannot describe a fully delocalized mode). Table 1 reports then 3 modes for VN_3H (the 3 degrees of freedom of the H atom), 6 modes for I_{2N} (the six modes of the two N atoms) and the 12 modes of the 4 H atoms.

Table 1 B3LYP wavenumbers (in cm^{-1}) and IR intensities (in km/mol) as obtained for small (SF), medium (MF), large (LF) and big (BF) fragments extracted from the S_{64} and S_{216} supercells. For comparison, frequencies obtained through the full Hessian diagonalization (*FHD*) are reported.

System	Supercell	IrRep	SF		MF		LF		BF		FHD	
			ν	*I*	*ν*	*I*	*ν*	*I*	*ν*	*I*	*ν*	*I*
VN_3H	64	E	1415	*249*	1416	*251*	1452	*213*	1459	*202*	1460	*206*
		A_1	3250	*612*	3250	*619*	3250	*619*	3250	*619*	3250	*619*
	216	E	1411	*240*	1411	*243*	1447	*206*	1454	*194*	1455	*194*
		A_1	3262	*601*	3262	*608*	3262	*608*	3262	*608*	3262	*608*
I_{2N}	64	E	509	*298*	467	*236*	422	*195*	394	*141*	396	*173*
		B_2	906	*34*	1301	*53*	1330	*46*	1339	*40*	1349	*34*
		E	1139	*173*	1415	*363*	1416	*355*	1420	*352*	1422	*347*
		A_1	1751	*0*	1757	*0*	1761	*0*	1761	*0*	1761	*0*
	216	E	513	*295*	471	*235*	426	*194*	402	*151*	414	*104*
		B_2	909	*34*	1306	*50*	1337	*43*	1347	*36*	1355	*33*
		E	1146	*179*	1425	*371*	1425	*362*	1430	*358*	1431	*352*
		A_1	1763	*0*	1768	*0*	1773	*0*	1773	*0*	1773	*0*
VH_4	64	T_1	861	*0*	750	*0*	494	*0*	412	*0*	352	*0*
		E	1320	*0*	1331	*0*	1387	*0*	1398	*0*	1400	*0*
		T_2	1382	*233*	1423	*285*	1478	*244*	1486	*233*	1488	*231*
		T_2	3853	*14*	3974	*4*	3974	*4*	3974	*4*	3974	*4*
		A_1	4199	*0*	4276	*0*	4276	*0*	4276	*0*	4276	*0*
	216	T_1	853	*0*	749	*0*	496	*0*	422	*0*	342	*0*
		E	1322	*0*	1333	*0*	1389	*0*	1399	*0*	1402	*0*
		T_2	1387	*240*	1428	*292*	1484	*252*	1492	*239*	1493	*237*
		T_2	3868	*12*	3989	*3*	3989	*3*	3989	*3*	3989	*3*
		A_1	4221	*0*	4298	*0*	4298	*0*	4298	*0*	4298	*0*

Let us consider first the VN_3H case. The Table 1 shows that the stretching (A_1) and bending (E) modes of H blue shift by 12 and red shift by 5 cm^{-1}, respectively, when the dilution of the defect increases (supercells S_{64} and S_{216}, respectively). So probably only at S_{512} or S_{1000} the interaction of modes belonging to different cells falls below 1-2 cm^{-1} . Working now at fixed defect concentration, the Table 1 shows that the *fragment* wavenumbers coincide with the FHD ones for the stretching already for the SF case. The bending mode is less localized, and requires a larger *fragment*: the difference with respect to FHD, that is as large as 45 cm^{-1} at the SF level, reduces to 8 and 1 cm^{-1} at the LF (17 atoms) and BF (41 atoms) level. The trend is the same for S_{64} and S_{216}.

The pattern is similar for the three modes of I_{2N} at 1773, 1431 and 1355 cm^{-1} (S_{216} supercell; also in this case the trend is the same for S_{64}); these modes are well localized and above the diamond large band. As regards the E mode at 414 cm^{-1} (S_{216}), the convergence of the various *fragments* is slower but satisfactory (LF and BF are at 426 and 402 cm^{-1}, in both case within 12 cm^{-1} of the FHD value)

The VH_4 case behaves very similarly to the other two cases. The four H stretching are very localized, and the MF results (8 atoms, the 4 H and the 4 C to which they are connected) already coincide with the FHD data. For the five bending modes above 1300 cm^{-1} (E and T_2 symmetry) it is necessary to reach the BF for reducing the difference to 3 cm^{-1}. The mode at very low wavenumber is much more delocalized; it decreases from 853 to 749 to 496 and finally to 422 in going from SF to BF. This value remains however about 80 cm^{-1} larger than the FHD result for S_{216}.

It is instructive to consider supercells larger than S_{216} for analyzing the dependence of the defect modes on the defect-defect interaction. Data for S_{512} and S_{1000} have been collected (see reference [16], and unpublished results), The A_1 stretching *breathing* in phase mode of the four H atoms introduces a relatively large perturbation, so that some further shift is observed (4298, 4306 and 4309 for S_{216}, S_{512} and S_{1000}). For the other high wavenumber modes, the difference between S_{216}, S_{500} and S_{1000} is 7 and 2 (T_2 stretching), 2 and 1 (T_2 bending) and 1 and 0 cm^{-1} (E bending). The lowest frequency of the full set of modes of the supercell, on the contrary, changes from 342 (S_{216}) to 289 (S_{512}) to 236 (S_{1000}) cm^{-1}, confirming its collective, very diffuse nature. If it continues to change also for concentrations described by supercells as large as S_{512} and S_{1000}, it cannot obviously be described accurately with any *fragment* of one of these cells.

Let us consider now the IR intensities. As they depend on the eigenvectors, whereas the wavenumbers are the eigenvalues,

one would expect that the effect of the perturbation (*i.e.* the propagation from the center of the defect) is larger for the former than for the latter. Actually, the FHD intensities turn out to be very well reproduced from the *fragments* in most of the cases.
In the VN_3H case, the difference between the BF and FHD intensities is 4/206 for the E mode and 0/619 km/mol for the A_1 mode, that is smaller than 2 % (S_{64} supercell; it goes to zero for both the A_1 and E modes in S_{216}).
For I_{2N} and the S_{64} supercell, the most intense peak at 1422 cm^{-1} shows a BF intensity of 352 km/mol, to be compared to 347 from FHD (1% error). The difference is larger for much less intense peaks (6 km/mol for the mode at 1349 cm^{-1}, 40 vs 34 km/mol). For the low frequency and very diffuse mode at 396 cm^{-1}, as expected, the difference is much larger (141 vs 173 km/mol, 18 % difference).
The VH_4 behavior is similar to the one of the other two defects: the intensity of the only IR active mode, at 1488 cm^{-1} (S_{64}), is 233 for BF and 231 km/mol for FHD , the difference being as small as 1%.
The above comments remain essentially unaltered when comparing the various *fragments* with the FHD obtained with S_{216}. It should be noticed that also with the very cheap MF the agreement with FHD is very satisfactory.
Figure 2 compares the FHD (top) and *fragment* (bottom, BF) spectra obtained with the largest supercell S_{216} for the three defects. The quantitative agreement, evident from Table 1 for the high wavenumbers, is here visually evident. Also the overall structure of the spectrum is well reproduced.

4 The cost of the calculation

We can now consider the cost of the calculation in the *Fragment* and FHD cases. The cost can be split in three parts:

- The Optimization step. Here the cost is the same for the two cases, as they differ only at the level of the construction of the Hessian matrix. The cost of this step is usually a small fraction of the overall cost, in particular when the system has some symmetry, that is fully exploited in CRYSTAL.
- The construction of the Hessian in the FHD strategy implies to perform, in the most general case, 3·N·(SCF+G) calculations (G stands for gradient), where N is the number of atoms. If the system has some point symmetry, this number can drastically reduce: for example in the case of the three diamond defects and of the S_{216} supercell, this number reduces from about 3·216 = 648 to 125 (VN_3H), 102 (I_{2N}) and 41 (VH_4), thanks to the symmetry, that is C_{3v}, D_{2d} and T_d for the 3 cases.
  However this reduced number of SCF+G calculations is performed with essentially no symmetry, as the displacement of one atom reduces in most of the cases the symmetry to P1.
  So this turns out to be by far the expensive step, due to this loss of symmetry (we recall that symmetry is fully exploited in the CRYSTAL code, and that the cost of the SCF+G calculation is inversely proportional to the number of point symmetry operators).
  When the *fragment* option is used, this step of the calculation costs 3·M·(SCF+G), where M is the number of atoms in the fragment. As an example, for the 4 different fragments (SF, MF, LF and BF) of the I_{2N} defect, and that contain 2, 6, 18 and 42 atoms respectively, the number of SCF+G calculations (6, 18, 54 and 126 without symmetry exploitation) reduces to 2, 4, 9 and 19. If however the reference cell would be S_{512} or S_{1000}, say, the number of SCF+G calculations with *fragment* would remain the same (obviously the cost of each SCF+G is now larger, being the cell larger), whereas the number of SCF+G of FHD would increase by a factor 2 and 4. Table 1 shows that quite small *fragments* are sufficient in most of the cases, indicating then that the saving factor obtained by using the *fragment* option can be as large as two orders of magnitude or more.
- The calculation of the IR intensities through the Berry phase scheme is very fast. Let us consider for example the VH_4 defect, S_{64} (then 67 atoms in the cell), for which a total of 16 SCF+G displacements must be performed. At each one of these steps, the Berry phase must be evaluated. When running on 32 cores, the cost of the SCF+G is about 700 seconds, and the cost of the Berry phase calculation is just 7 seconds (1% of the total). For the S_{216} case (then 219 atoms, about 3.4 times larger than S_{64} in number of atoms), one SCF+G calculation for a displacement costs 2500 seconds, and the Berry phase just 16 seconds, less than 1%. The real bottle-neck becomes the memory occupation for systems with more than, say, 500 atoms, and this would require restructuring the algorithm, that will be the next implementation.

5 Conclusions

In the characterization of large unit cell systems, containing hundreds or thousands of atoms, through the IR spectra, the construction of the full Hessian matrix is not only extremely expensive, but also not very useful, as the fingerprints of the system, if any, are related to specific functional groups. It has been shown here that a *fragment* strategy, consisting in computing a Hessian matrix involving only a subset of atoms *around* the functional group of interest, permits to limit the cost of the calculation to an amount that is independent from the size of the real system, at variance with respect to the full Hessian diagonalization FHD scheme that grows rapidly with the system size. The three examples considered here document that also relatively small subsets of atoms (say less

than 40) reproduce the wavenumbers with a negligible error when compared with the FHD scheme. Also the IR intensities are well reproduced, with differences in percentage never larger than 10% for the high frequency modes.
The *fragment* strategy seems then quite effective, at least for the cases here discussed (defects in semiconductors). Its applicability and effectiveness to a larger family of systems will be investigated in the near future.

ACKNOWLEDGEMENTS

FP: this work was granted access to the HPC resources of [CINES/IDRIS/TGCC] under the allocation 2018- [A0040810471] made by GENCI. High Performance Computing resources were partially provided by the EXPLOR centre hosted by the University de Lorraine.

References

1. R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman. Quantum-Mechanical Condensed Matter Simulations with CRYSTAL. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8:e1360, 2018. URL `https://doi.org/10.1002/wcms.1360`.
2. G. Kresse and J. Furthmüller. Efficient Iterative Schemes for *Ab Initio* Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B*, 54:11169–11186, 1996.
3. Jürg Hutter, Marcella Iannuzzi, Florian Schiffmann, and Joost VandeVondele. CP2K: Atomistic Simulations of Condensed Matter Systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(1):15–25, 2014.
4. P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, et al. QUANTUM ESPRESSO: a Modular and Open-Source Software Project for Quantum Simulations of Materials. *J. Phys. Condens. Matter*, 21(39):395502, 2009.
5. N. Ferre and J.G. Ángyán. Approximate Electrostatic Interaction Operator for QM/MM calculations. *Chem. Phys. Lett.*, 356:33–339, 2002.
6. J. Paier, M. Marsman, K. Hummer, G. Kresse, I.C. Gerber, and J.G. Ángyán. Screened Hybrid Density Fuctionals Applied to Solids. *J. Chem. Phys.*, 124:154709, 2006.
7. G.I. Csonka, J.P. Perdew, A. Ruzsinsly, P.H.T. Philipsen, H.T. Pier, S. Lebègue, J. Paier, O.A. Vidrov, and J.G. Ángyán. Assessing the Performance of Recent Density Functional for Bulk Solids. *Phys. Rev. B*, 79:155107, 2009.
8. S. Salustro, A. M. Ferrari, F. S. Gentile, J. K. Desmarais, M. Rérat, and R. Dovesi. Characterization of the B-center Defect in Diamond Through the Vibrational Spectrum. A Quantum Mechanical Approach. *J. Phys. Chem. A*, 2017. doi: 10.1021/acs.jpca.7b11551.
9. A. Erba, J. Baima, I. Bush, R. Orlando, and R. Dovesi. Large-Scale Condensed Matter DFT Simulations: Performance and Capabilities of the CRYSTAL Code. *J. Chem. Theor. Comput.*, 13(10):5019–5027, 2017.
10. C. R. Jacob and M. Reiher. Localizing Normal Modes in Large Molecules. *J. Chem. Phys.*, 130:084106, 2009.
11. P. Bouř and T. A. Keiderling. Partial Optimization of Molecular Geometry in Normal Coordinates and Use as a Tool for Simulation of Vibrational Spectra. *J. Chem. Phys.*, 117:4126, 2002.
12. S. Yamamoto and P. Bouř. *Frontiers of Quantum Chemistry: Calculation of Vibrational Spectra of Large Molecules from Their Fragments*. . Springer Nature Singapore, 2018. URL `https://doi.org/10.1007/978-981-10-5651-2`.
13. S. Salustro, A. M. Ferrari, R. Orlando, and R. Dovesi. Comparison Between Cluster and Supercell Approaches: the case of Defects in Diamond. *Theor. Chem. Acc.*, 4(136):1–13, 2017.
14. F. Gentile, S. Salustro, M. Causá, A. Erba, P. Carbonnière, and R. Dovesi. The VN_3H Defect in Diamond. A Quantum Mechanical Investigation of the Structural, Electronic and Vibrational Properties. *Phys. Chem. Chem. Phys.*, 1(4):1–2, 2017.
15. S. Salustro, F. Pascale, W. Mackrodt, C. Ravoux, A. Erba, and R. Dovesi. Interstitial Nitrogen Atoms in Diamond; A Quantum Mechanical Investigation of its Electronic and Vibrational Properties. *Phys. Chem. Chem. Phys.*, xx:yy, 2018.
16. F. S. Gentile, S. Salustro, J.K. Desmarais, A.M. Ferrari, P. D'Arco, and R. Dovesi. Vibrational Spectroscopy of Hydrogens in Diamond: A Quantum Mechanical Treatment. *PCCP*, 20:11930–11940, 2018. doi: 10.1039/C8CP00596F.
17. A. D. Becke. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.*, 98(7):5648–5652, 1993.
18. C. Lee, W. Yang, and R. Parr. Development of the Colle-Salvetti Correlation-Energy Formula Into a Functional of the Electron Density. *Phys. Rev. B*, 37(2):785–789, 1988.
19. J. S. Binkley, J. A. Pople, and W. J. Hehre. Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements. *J. Am. Chem. Soc.*, 102(3):939–947, 1980.
20. W. J. Hehre, R. Ditchfield, and J. A. Pople. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.*, 56(5):2257–2261, 1972.
21. A. D. Becke. A Multicenter Numerical Integration Scheme for Polyatomic Molecules. *J. Chem. Phys.*, 88(4):2547–2553, 1988.
22. R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, Ph. D'Arco, and M. Llunell. *CRYSTAL 2014 User's Manual*. . University of Torino, Torino, 2013.
23. H. J. Monkhorst and J. D. Pack. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B*, 13(12):5188, 1976.
24. F. Pascale, C. M. Zicovich-Wilson, F. Lòpez Gejo, B. Civalleri, R. Orlando, and R. Dovesi. The Calculation of the Vibrational Frequencies of the Crystalline Compounds and its Implementation in the CRYSTAL Code. *J. Comput. Chem.*, 25(6):888–897, 2004.
25. P. Ugliengo, F.Pascale, M. Mérawa, P. Labéguerie, S. Tosoni, and R. Dovesi. Infrared Spectra of Hydrogen-Bonded Ionic Crystals: *Ab Initio* Study of $Mg(OH)_2$ and β-$Be(OH)_2$. *J. Phys. Chem. B*, 108(36):13632–13637, 2004.
26. S. Dall'Olio, R. Dovesi, and R. Resta. Spontaneous Polarization as a Berry Phase of the Hartree-Fock Wavefunction: the Case of KNbO3. *Phys. Rev. B*, 56:10105–10114, 1997.
27. R.D. King-Smith and D. Vanderbilt. First-Principles Investigation of Ferroelectricity in Perovskite Compounds. *Phys. Rev. B*, 49: 5828, 1994.