

Convergence and pitfalls of density functional perturbation theory phonons calculations from a high-throughput perspective

Guido Petretto*, Xavier Gonze, Geoffroy Hautier, Gian-Marco Rignanese*

Institute of Condensed Matter and Nanoscience (IMCN), Université catholique de Louvain, B-1348 Louvain-la-neuve, Belgium

ARTICLE INFO

Article history:

Received 26 September 2017

Received in revised form 16 December 2017

Accepted 16 December 2017

Keywords:

Phonons

Density functional perturbation theory

High-throughput

ABSTRACT

The diffusion of large databases collecting different kind of material properties from high-throughput density functional theory calculations has opened new paths in the study of materials science thanks to data mining and machine learning techniques. Phonon calculations have already been employed successfully to predict materials properties and interpret experimental data, e.g. phase stability, ferroelectricity and Raman spectra, so their availability for a large set of materials will further increase the analytical and predictive power at hand. Moving to a larger scale with density functional perturbation calculations, however, requires the presence of a robust framework to handle this challenging task. In light of this, we automatized the phonon calculation and applied the result to the analysis of the convergence trends for several materials. This allowed to identify and tackle some common problems emerging in this kind of simulations and to lay out the basis to obtain reliable phonon band structures from high-throughput calculations, as well as optimizing the approach to standard phonon simulations.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

In the latest years, the development of efficient density functional theory (DFT) software packages and the constant increase of the computational capabilities of modern supercomputers have opened new possibilities in the field of material science, including the discovery of new materials based on the high-throughput screening of large number of compounds [1]. Following this trend, several open databases have been created containing a number of properties obtained from DFT calculations [2–6] thanks to *ad hoc* frameworks developed to help the automation of the whole process [7–11]. High-throughput techniques have not only allowed to search for materials candidates for a specific application but also to use data mining and machine learning techniques to understand trends and limitations in materials science [12–17].

Building on the experience gained so far and on the tools made available, it is now possible to address the calculation of more and more involved quantities for a vast number of materials, with efforts that are ongoing to analyze properties obtained with calculations beyond the ground state [18–23]. Vibrational properties in solids are essential to many applications as they govern several materials properties (e.g. vibrational entropy [24,25], thermal con-

ductivity [26–28], ferroelectric and ferroelastic transitions [29,30]). Phonon computations can be performed very efficiently using the Density functional perturbation theory (DFPT) [31] providing a computational access to these important properties.

Before moving to large scale calculations it is mandatory to implement a robust procedure to handle the whole process and provide a reliable set of input parameters that suits most of the possible cases that will be considered. In this regard, it has been shown how the validation of the results is an element of great relevance when approaching the high-throughput regime [20,32].

In this paper we present the outcome of a study concerning the convergence rate of phonon related quantities with respect to the sampling of the Brillouin zone alongside a validation of the results generated within an automated process. In addition, since performing a considerable amount of simulations provides a stress test for the DFT and DFPT code, we have been able to identify some subtleties and pitfalls, that tend to show up under certain conditions and that may go unnoticed in a standard approach. Given the substantial difference in terms of requirements between metals and semiconductors and the difficulty of expressing a general recipe that would allow to capture peculiar phenomena as the Kohn anomalies, we have limited our analysis to the case of semiconducting materials.

In the following, after the definition of the theoretical framework (Section 2), we will first discuss the evolution of the phonon frequencies with respect to the density of k -points and q -points in

* Corresponding authors.

E-mail addresses: guido.petretto@uclouvain.be (G. Petretto), gian-marco.rignanese@uclouvain.be (G.-M. Rignanese).

Sections 3 and 4, respectively, highlighting the problems related with the sampling. In the second part we proceed to demonstrate the reliability of our results through a validation process, involving also a comparison with experimental data (Section 5).

2. Formalism and methodologies

In this study we considered a set of 48 semiconducting materials. These were chosen to cover a variety of elements of the periodic table and all the crystal systems, with different system sizes and band gaps. Detailed information about the properties of each material can be found in the [Supplemental Material \(SM\)](#).

All the simulations were performed with the ABINIT software package [33–35], which relies on the DFPT formalism to carry on calculations related to phononic and electric field perturbations [31,36,37]. The Perdew-Burke-Ernzerhof (PBE) [38] was used as exchange–correlation (xc) functional. While this standard approximation has been demonstrated to underperform for phonon frequencies compared to other exchange–correlation functionals, it still produces results in reasonable agreement with experimental data [39] and we assume that our conclusions will still hold for other local and semilocal xc functional approximations (e.g., LDA and PBEsol).

Optimized norm-conserving pseudopotentials (ONCV) [40] were used for all the elements treating semi-core states as valence electrons for transition metals (as available from the pseudopotentials table Pseudo-dojo version 0.2 [41]). The cutoff was chosen independently for each material according to the values suggested in the Pseudo-dojo (see SM). These pseudopotentials and the cutoff values have been carefully tested with respect to all electron codes [42] and with respect to the fulfillment of the acoustic sum rule (ASR). Checks on the ASR and the charge neutrality [37], that are sensitive to the cutoff, were also performed on each material and the convergence with respect to the cutoff was further verified for problematic cases.

The splitting between longitudinal and transverse optical mode (LO-TO) was taken into account through the calculation of the non-analytic term containing the Born effective charges Z^* (BECs) [37]. Phonon frequencies at generic q -points were obtained through Fourier interpolation and, in that case, the ASR and the charge neutrality at the Γ point were enforced explicitly.

For all the materials, the unit cells were standardized according to Ref. [43] and each was relaxed until all the forces on the atoms were below 10^{-6} Ha/Bohr and the stresses were below 10^{-4} Ha/Bohr³.

All the calculations were carried out using Fireworks as a workflow manager [7] integrated with the different libraries: pymatgen [8], abipy [44] and abiflows [45].

The Monkhorst-Pack grids used to sample the Brillouin zone were distinguished considering the number of the points per reciprocal atom for both the electron wavefunctions (referred to as k -points) and for the phonon wavefunctions (referred to as q -points), respectively. These quantities, the number of points in the full Brillouin zone times the number of atoms, are labelled $kpra$ and $qpra$, respectively.

For the statistical analysis, we identified, for each material, a reference grid (leading to converged phonon frequencies, see below). For each other sampling density the errors were defined as the absolute value of the difference between the phonon frequencies ω calculated with that grid and those obtained with the reference grid. In this context, when analyzing the k -point convergence we considered the frequencies at the high symmetry points of the Brillouin zone [43], while when varying the q -points grid the convergence was checked for the interpolated values on very dense regular grids of q -points (100,000 $qpra$). In the latter case, q -points belonging to the dense grids explicitly obtained from the DFPT

calculation were excluded from the statistical analysis, as their error would be zero by definition.

For each specific grid we considered mainly the mean absolute error (MAE) and the mean absolute relative error (MARE) over all the frequencies. The choice of focusing on averaged errors stems from the fact that many quantities derived from phonon calculations that can be compared to experimental results are obtained integrating over the whole Brillouin zone (e.g. DOS and thermodynamic properties [46]). In addition, in order to get a better insight on the presence of possible problematic regions, the maximum absolute error (MxAE) and the maximum absolute relative error (MxAE) were also considered.

The reference configuration was selected independently for each material and for the k and q grids, by considering progressively increasing densities in the reciprocal space, until a convergence was reached based on the MAE and MARE.

3. Electron wavevector grid convergence

Before any statistical analysis, we would like to highlight the importance of the choice of the k -point grid itself. For a subset of the materials considered (Si, BP and ZnO) the k -point dependence of the phonon frequencies was investigated for a list of Γ centered grids, thus satisfying the same symmetries of the crystal, and for the same grids with a shift that breaks the symmetry. The values of the MxAE with respect to a reference grid of approximately 5500 $kpra$ are reported in Fig. 1, showing that the symmetric set of grids can reduce the error by up to two orders of magnitude compared to the non-symmetric one. This demonstrates the importance of choosing a k -point grid that respects the symmetries of the system to improve the rate of convergence. In the following, the k -point grids were thus chosen with appropriate shifts satisfying this criterion.

The relative and absolute errors for the k -point grids were calculated as described in Section 2 for each different grid available. Considering a threshold ε for the maximum error tolerated and given a specific value \overline{kpra} of the k -points per reciprocal atom, we analyzed the fraction F of materials for which the results are converged within the limits ε and \overline{kpra} . More precisely, we define F as:

$$F(\overline{kpra}, \varepsilon) = \frac{\sum_{i=1}^{48} [|\text{Err}_i(\kappa_i)| < \varepsilon]}{48}, \quad (1)$$

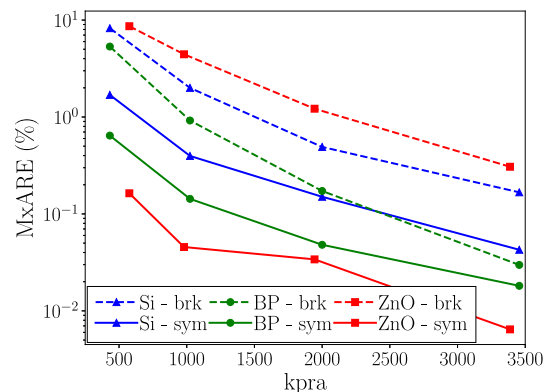


Fig. 1. Evolution of the maximum relative error for different k -point samplings for Si (blue triangle), BP (green circle) and ZnO (red square). The reference grids have approximately 5500 $kpra$. Solid lines represent the error for a Γ centered grid, while dashed lines are for shifted grids, preserving and breaking the symmetries (sym and brk) of the crystals, respectively. The lines are a guide to the eye. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where i labels the 48 materials, κ_i is the largest grid available lower than \overline{kpra}

$$\kappa_i = \max_{kpra_i < \overline{kpra}} kpra_i, \quad (2)$$

$\text{Err}_i(\kappa_i)$ represents either the MAE, MARE, MxAE or MxARE, while the Iverson bracket $[P]$ for a proposition P is 1 if P is true and 0 otherwise.

The values of F as a function of $kpra$ and ε for the different types of errors are reported in Fig. 2 and can be used to identify an appropriate value for the k -point density. The mean errors, both relative and absolute, tend to converge rather fast and for grid denser than 1000 $kpra$ all the materials can be considered converged with a MAE below 1 cm^{-1} and a MARE of 1.5%. However, this represents the largest mean error and, at this grid density 90% of the materials are actually converged at a much lower error threshold (0.3 cm^{-1} and 0.15%).

It should be noted that F is not strictly a monotonic function of $kpra$ for a fixed value of ε , reflecting the oscillations in the error that show up in a few cases. These oscillations are eventually smoothed out for denser grids and, in the cases where this phenomenon is more relevant (AgCl and ScF_3), their origin can be traced back to a large error on a particular mode for a particular q -point, suggesting that the impact on the overall band structure is quite limited.

This can be seen also in the case of the maximum errors, where these large oscillations affects just a couple of cases, while almost all the materials show rather small values of MxRE and MxARE at 1000 $kpra$.

These results can be compared to the rate of convergence of the total energy per atom E_{at} . In Fig. 3 we report the values of F for the absolute error (AE), showing how, with a $kpra$ below 800, 95% of the materials are converged within a threshold of 5 meV/atom. This demonstrates that, in general, a denser sampling of the Brillouin zone is required to obtain accurate perturbations, compared to the total energy.

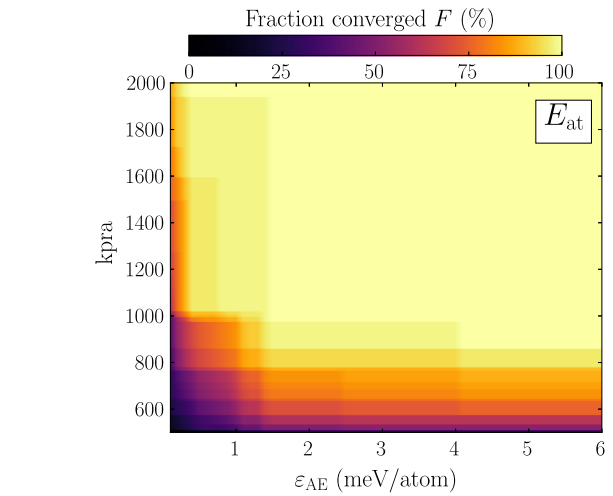


Fig. 3. Fraction of materials converged F (see Eq. (1)) with respect to the energy per atom E_{at} as a function of the k -point density $kpra$ and maximum error threshold allowed for the absolute error ε_{AE} .

lounin zone is required to obtain accurate perturbations, compared to the total energy.

In the case of polar materials, the Γ point deserves further discussion. The error on the modes that have an LO-TO splitting is systematically larger compared to the error obtained neglecting the nonanalytical contribution. This is due to a slower convergence of the derivative with respect to the electric field. This should be usually taken into account when studying the phonon frequencies just at Γ , but it should also be mentioned that the maximum errors shown in Fig. 2 are coming from modes at q -points away from the origin for most of the materials. This means that when considering the whole Brillouin zone a denser k -point grid is not needed for the electric field derivatives.

As a further check, we explored the convergence of the BECs and dielectric constant ϵ for all the polar materials considered here. The fraction F for the MARE and MxARE related to these two quantities are shown in Figs. 4 and 5, respectively. In general, the rate of convergence turned out to be slightly slower compared to the phonon frequencies shown in Fig. 2, but at 1000 $kpra$ for these quantities roughly 85% of the materials are converged with a tolerance of 5% on the MxARE. This k -point density is then suitable in general also for quantities purely related to the electric perturbations.

As a final point, we confirm that k -points have little or no influence on the breaking of the ASR, when it is not imposed explicitly.

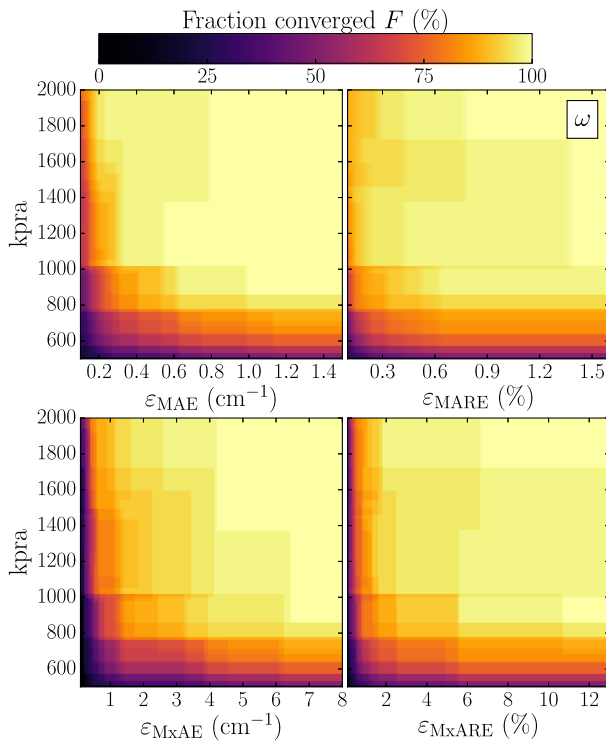


Fig. 2. Fraction of materials converged F (see Eq. (1)) with respect to the phonon frequencies ω as a function of the k -point density $kpra$ and maximum error threshold allowed ε for the different kind of errors MAE, MARE, MxAE, MxARE.

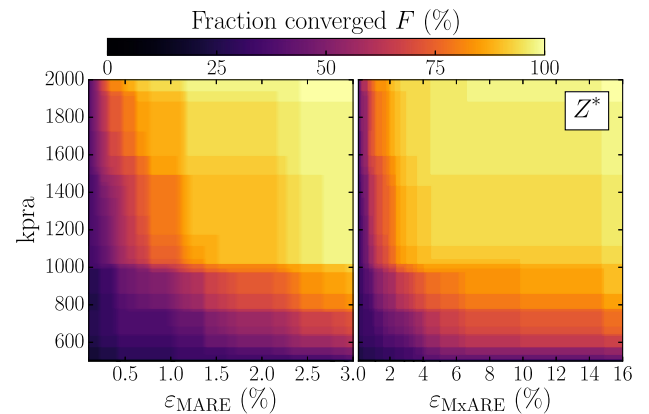


Fig. 4. Fraction of materials converged F (see Eq. (1)) with respect to the BECs Z^* as a function of the k -point density $kpra$ and maximum error threshold allowed ε for the different kind of relative errors MARE and MxARE.

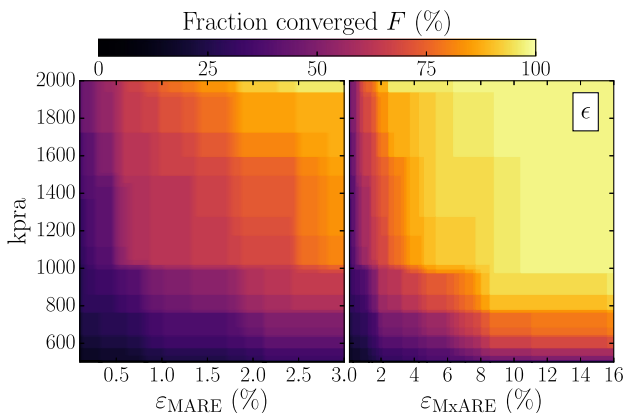


Fig. 5. Fraction of materials converged F (see Eq. (1)) with respect to the dielectric tensor ϵ as a function of the k -point density $kpra$ and maximum error threshold allowed ϵ for the different kind of relative errors MARE and MxARE.

4. Phonon wavevector grid convergence

Given the conclusions from Section 3, we chose k -point grids with densities larger than 1000 $kpra$ for all the materials and the phonon frequencies were calculated using different sets of phonon wavevector (q -points) samplings. The aim was to identify a value of $qpra$ that would provide converged results for the phonon frequencies obtained through Fourier interpolation over the whole Brillouin zone.

The values of the fraction of converged materials F , as defined in Eq. (1), are shown in Fig. 6 for the MAE, MARE, MxAE and MxARE as a function of $qpra$ and ϵ . Comparing these results with the case of the k -point sampling, it can be seen immediately that the rate of convergence with respect to the q -point density is slower and in particular that there is a strikingly different behavior of the maximum errors.

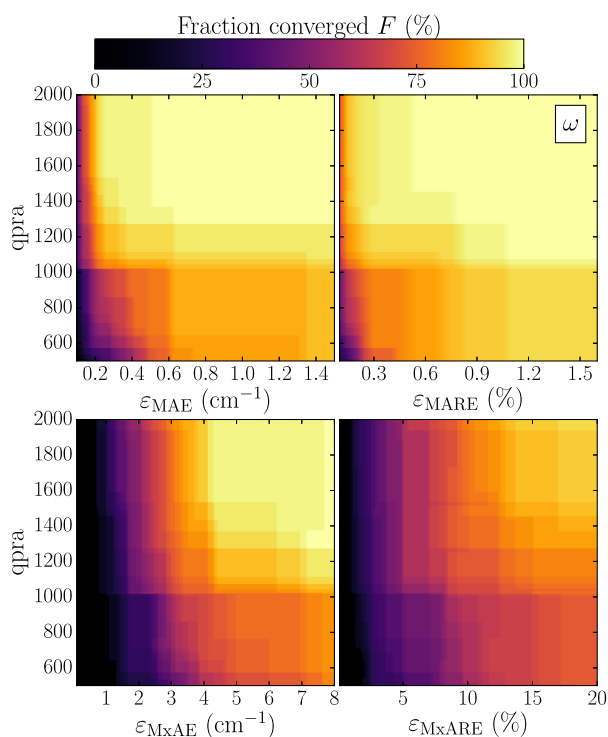


Fig. 6. Fraction of materials converged F (see Eq. (1)) with respect to the phonon frequencies ω as a function of the q -point density $qpra$ and maximum error threshold allowed ϵ for the different kind of errors MAE, MARE, MxAE, MxARE.

The presence of large maximum errors can be partially understood on the basis of two considerations: first, the fact that values are obtained through an interpolation could introduce errors related to the interpolation scheme. Second, evaluating the errors on a very large set of points increases the probability of coming across a small region of the Brillouin zone that is poorly described, possibly leading to a large maximum error. Under these assumptions it is reasonable to expect that the average error will not be increased substantially.

However, while these factors definitely play a role, analyzing each material individually, we also identified a series of problems giving a strong contribution to the maximum error.

One of these problems, present in different materials, shows up when the q -point grid becomes denser than the k -point grid. The frequencies tend to deviate from the linear trend that is expected for the acoustic modes close to Γ . The problem becomes more and more evident as the q grid gets denser. Due to the interpolation, these deviations tend to propagate also further away from the close proximity of Γ (see Fig. S1 of the SM for an example). This can lead to a large maximum error, as we have defined it, due to an incorrect value of the reference frequencies. In particular, since this involves mainly small frequencies for few modes and few q -points, the MxARE tends to be very large, while the average errors still keeps improving with increasing $qpra$. Despite this, we explicitly verified in the cases of K_2O , FeS_2 , RbI and $RbYO_2$, that changing the k -point grid to match exactly the values of the q -point grid eliminates completely these deviations. These calculations show that the frequencies (even close to Γ) were already well converged even with a lower $qpra$ (see Fig. S2 of the SM).

Although we were not able to identify the origin of the problem univocally there are various possible sources of error for $\mathbf{q} \rightarrow 0$. First the wavelength of the phonon perturbations is longer than the electronic one. There is also the need for additional calculations to handle incommensurate grids. Moreover, we have observed that adding an appropriate cutoff in real space on the IFCs corrects the behavior close to Γ (see Fig. S3 of the SM), suggesting that some numerical inaccuracies at long wavelength should be involved.

We point out that the value of such a cutoff on the IFCs should be tuned to get satisfactory results of the phonon frequencies. So, while this may be a viable option to improve the results in standard phonon calculations, its use in an high-throughput process is limited by the need of an automated determination of a suitable value of the cutoff, that may be not trivial.

A second problem appears in $AgCl$. On top of some inaccuracies as those discussed above, this material has its relaxed PBE lattice parameters close to a phase transition [47]. As a consequence some acoustic modes are softened along the Γ -K direction and at L, with the onset of small dips in the phonon band structure (see Fig. S4 of the SM). These irregularities may be not properly described with a Fourier interpolation on a regular grid of q -points, so large MxRE and MxARE arise in those zones due to the different interpolations obtained with different q -point grids.

Aside from these problems, we also highlight the fact that deviations from the linear trend for acoustic modes at Γ could signal a lack of convergence of the energy cutoff or the k -point sampling, and that can be used as an indicator to identify unconverged calculations during an high-throughput screening.

Based on these considerations about the maximum errors and on the results of Fig. 6, we conclude that nonetheless a larger $qpra$ is needed in general with respect to the $kpra$, with an optimal value set at around 1500 points per reciprocal atom. With this value all the materials could be considered converged within a MAE of 0.5 cm^{-1} and a MARE below 0.6%. Such a $qpra$ value also leads to satisfactory maximum errors, with the exception of the cases discussed above.

Due to the problems highlighted above, it would also be convenient to raise the value of k_{prra} as well, in order to always work with q -point and k -point grids of the same size.

The conclusion of the present study also applies to the computation of phonon frequencies using the supercell method [48,49], Fourier extrapolated using interatomic force constants. Indeed, the present DFPT approach and the supercell method do not differ at the level of their interpolation scheme. This means that the above-mentioned MAE and MARE targets or maximum errors need supercells of 1500 atoms to be attained. The current practice with supercell calculations, however, usually rely on supercells with less than 200 atoms. Hence, the accuracy of such studies is quite limited. This illustrates the power of the DFPT approach, that uses only the primitive cell, even for incommensurate q -point wavevectors.

5. Validation

To validate our set of results and highlight the effects of the problems discussed in the previous sections further checks have been performed. One aimed at the analysis of the pathological region in the proximity of Γ and the other aimed at testing the accuracy in comparison with experimental data.

5.1. Sound velocity

The first test is based on the calculation of the sound velocity of the materials, as obtained from the slope of the acoustic mode close to the Γ point. Since not many experimental data are available, here we focus on establishing the precision of our results, rather than the accuracy. We then calculated the same sound velocities from the DFPT calculations of the elastic constants and the help of the Christoffel tensor. These quantities should match for converged calculations and allow to verify the trends close to the origin of the Brillouin zone. In fact, apart from the problems mentioned in Section 4, the linear behavior of the acoustic modes can eventually break down for small q , altering the value of the slope extracted from the interpolation.

More in detail, elastic constants were calculated with the same approximations used to run the phonon simulations in Section 4, while to extract the sound velocity from phonon frequencies q -point grids of approximately 1500 q pr were used. We compared the values of the speed of sound for the three acoustic modes averaged over the three cartesian directions. The distribution of the relative difference between the two different types of calculations is shown in Fig. 7.

The agreement is generally good, with more than 80% of the evaluated sound velocities having a difference lower than 5%

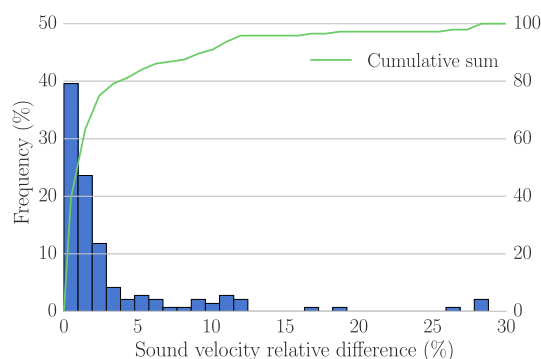


Fig. 7. Distribution of the relative difference between the sound velocities obtained from the slope of the acoustic modes and from the elastic tensor. Each values considered is for one of the three acoustic modes, averaged over the three cartesian direction.

between the two methods and more than 90% below 10% difference.

The outliers with large differences can be easily understood. The first is ZnSe, for which the cutoff energy is not large enough for the calculation of the elastic tensor, while phonon bands are converged. As the cutoff is increased the agreement is in line with other materials. The second is AgCl, that was discussed in the previous section and that shows problems close to Γ coming from the q -point sampling and from the presence of soft modes in the Brillouin zone. This is likely to lead to inaccuracies in the interpolated values of the sound velocity.

The other smaller differences can be rationalized in terms of the convergence with respect to the k -points, q -points and energy cutoff as well as the small numerical inaccuracies in the interpolation close to Γ . We can thus conclude that the precision achieved with the suggested densities is satisfactory even for the description of the central region of the Brillouin zone, in particular having high-throughput calculations in mind.

5.2. Thermodynamic properties

The second part of the validation concerns the precision achieved in comparison with experimental data. In particular we used as a target the values of the entropy S at room temperature (i.e. around 300 K, matching the experimental data available). This can be obtained from the normalized phonon density of states $g(\omega)$ [46]:

$$S(T) = 3nNk_B \int_0^{\omega_L} \left(\frac{\hbar\omega}{2k_B T} \coth \left(\frac{\hbar\omega}{2k_B T} \right) - \ln \left(2 \sinh \left(\frac{\hbar\omega}{2k_B T} \right) \right) \right) g(\omega) d\omega \quad (3)$$

where the integral is carried out over all the phonon frequencies, k_B is the Boltzmann constant, n is the number of atoms per unit cell, N is the number of unit cells, and T is the temperature.

Entropy provides a good example of quantity integrated over the whole Brillouin zone, that should have errors comparable to those observed for MAE and MARE discussed in Section 4. The rate of convergence with respect to q -points is indeed quite fast and for a density of 1500 q pr the maximum relative error observed for the entropy at room temperature with respect to the reference q -point grid is below 0.7% for all the compounds.

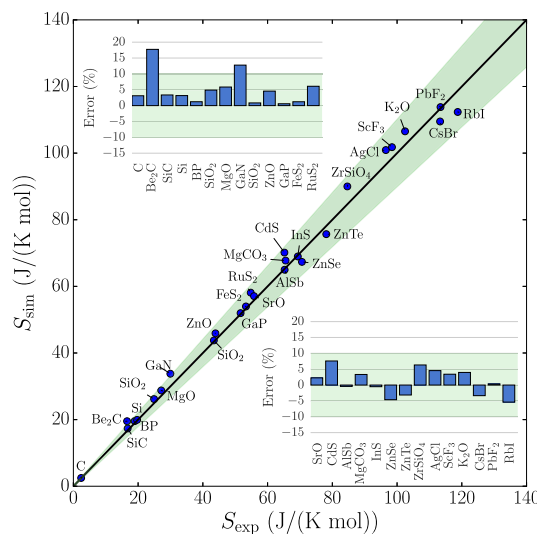


Fig. 8. Experimental vs theoretical data for entropy S at room temperature. The green area delimits a region with relative error below 10%. The insets represent the relative error with respect to the experimental value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

These values of the entropy were compared with the experimental data for the 27 compounds for which these are available [50–55]. This comparison is shown in Fig. 8, where the relative errors are reported as well.

The agreement is generally good, considering that other sources of entropy can enter in the experimental values, with the exception of Be₂C and GaN. The MARE between the theoretical and experimental values is 4.2%, in line with data present in literature [39,56–58], while the MAE is 2.23 J/(K mol) (9.6×10^{-3} meV/(K atom)). Interestingly, at 300 K this is much smaller than the estimated error of other energy contributions for the estimation of phase stability [59]. It should be noted that at room temperature the anharmonic contribution to the entropy, that were not taken into account in our calculations, could partially explain the discrepancies and the systematic overestimation of the theoretical results.

These results allows us to confirm the reliability of our calculations, with a good estimate of the precision in terms of integrated quantities.

6. Conclusion

We have studied the convergence of the phonon frequencies obtained in the framework of DFPT as a function of the sampling of the Brillouin zone, in order to provide a reliable high-throughput framework for the calculation of phonon band structures.

We have highlighted the common problems emerging from the choices of the sampling densities. In particular, for the k -points, a higher density is needed for obtaining well-converged values of the LO-TO splitting compared to the simple phononic perturbations at Γ . A slightly higher density is also generally needed for q -points other than the origin.

Concerning the frequencies obtained from the interpolation of a regular q grid, the region close to Γ is the most problematic due to numerical inaccuracies (insufficiently converged parameters, long range oscillations), proximity of the system to a phase transition and numerical inaccuracies in the Fourier interpolation.

It has also been observed that imposing a cutoff on the IFCs in real space may help in solving problems related to numerical inaccuracies originating from different sources. However, an appropriate value of the cutoff should be chosen individually for each material and the results should be thoroughly checked.

In general, a smaller density of 1000 points per reciprocal atoms seems to be necessary for the k -points compared to the q -points. However, having observed some pathological behavior close to the origin of the Brillouin zone when using q grids denser than k grids, we suggest that the best approach would be to use k grids with an appropriate single shift that preserves the symmetries of the crystal along with equivalent q grids, both with a density of approximately 1500 points per reciprocal atom. This should provide well converged results for the majority of semiconducting materials.

While these hints were extracted to tune the calculations in an high-throughput regime, they can also be used as indications for standard phonon calculations.

Acknowledgement

We gratefully acknowledge discussions with B. Van Troeye and M. Giantomassi. G.P., X.G. and G.-M.R. acknowledge support from the Communauté française de Belgique through the BATTAB project (ARC 14/19-057). G.-M.R. is also grateful to the F.R.S.-FNRS for the financial support. Computational resources have been provided by the supercomputing facilities of the Université catholique de Louvain (CISM/UCL) and the Consortium des Equipements de

Calcul Intensif en Fédération Wallonie Bruxelles (CECI) funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.commatsci.2017.12.040>.

References

- [1] S. Curtarolo, G.L. Hart, M. Buongiorno Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, *Nat. Mater.* 12 (3) (2013) 191–201, <https://doi.org/10.1038/nmat3568>.
- [2] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: the materials project: a materials genome approach to accelerating materials innovation, *APL Mater.* 1 (1) (2013) 011002, <https://doi.org/10.1063/1.4812323>.
- [3] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L. Hart, S. Sanvito, M. Buongiorno Nardelli, N. Mingo, O. Levy, Aflowlib.org: a distributed materials properties repository from high-throughput ab initio calculations, *Comput. Mater. Sci.* 58 (2012) 227–235, <https://doi.org/10.1016/j.commatsci.2012.02.002>, URL <<http://www.sciencedirect.com/science/article/pii/S0927025612000687>> .
- [4] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD), *JOM* 65 (11) (2013) 1501–1509, <https://doi.org/10.1007/s11837-013-0755-4>.
- [5] Nomad Repository. <<http://nomad-repository.eu>>.
- [6] Open Materials Database. <<http://openmaterialsdb.se>>.
- [7] A. Jain, S.P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, K.A. Persson, Fireworks: a dynamic workflow system designed for high-throughput applications, *Concurr. Comput.: Pract. Exper.* 27 (17) (2015) 5037–5059, <https://doi.org/10.1002/cpe.3505>.
- [8] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python materials genomics (pymatgen): a robust, open-source python library for materials analysis, *Comput. Mater. Sci.* 68 (2013) 314–319, <https://doi.org/10.1016/j.commatsci.2012.10.028>, URL <<http://www.sciencedirect.com/science/article/pii/S0927025612006295>> .
- [9] S. Curtarolo, W. Setyawan, G.L. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, AFLOW: an automatic framework for high-throughput materials discovery, *Comput. Mater. Sci.* 58 (2012) 218–226, <https://doi.org/10.1016/j.commatsci.2012.02.005>, URL <<http://www.sciencedirect.com/science/article/pii/S0927025612000717>> .
- [10] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky, AiiDa: automated interactive infrastructure and database for computational science, *Comput. Mater. Sci.* 111 (2016) 218–230, <https://doi.org/10.1016/j.commatsci.2015.09.013>, URL <<http://www.sciencedirect.com/science/article/pii/S0927025615005820>> .
- [11] S.R. Bahn, K.W. Jacobsen, An object-oriented scripting interface to a legacy electronic structure code, *Comput. Sci. Eng.* 4 (3) (2002) 56–66, <https://doi.org/10.1109/5992.998641>.
- [12] A. Jain, G. Hautier, S.P. Ong, K. Persson, New opportunities for materials informatics: resources and data mining techniques for uncovering hidden relationships, *J. Mater. Res.* 31 (2016) 977–994, <https://doi.org/10.1557/jmr.2016.80>, URL <http://journals.cambridge.org/article_S0884291416000807> .
- [13] J.B. Varley, A. Miglio, V.-A. Ha, M.J. van Setten, G.-M. Rignanese, G. Hautier, High-throughput design of non-oxide p-type transparent conducting materials: data mining, search strategy, and identification of boron phosphide, *Chem. Mater.* <https://doi.org/10.1021/acs.chemmater.6b04663>.
- [14] W. Chen, J.-H. Pohls, G. Hautier, D. Broberg, S. Bajaj, U. Aydemir, Z.M. Gibbs, H. Zhu, M. Asta, G.J. Snyder, B. Meredig, M.A. White, K. Persson, A. Jain, Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment, *J. Mater. Chem. C* (2016), <https://doi.org/10.1039/C5TC04339E>.
- [15] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, Universal fragment descriptors for predicting properties of inorganic crystals, *Nat. Commun.* 8 (2017) 15679, <https://doi.org/10.1038/ncomms15679>, URL <<https://www.nature.com/articles/ncomms15679>> .
- [16] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *NPJ Comput. Mater.* 2 (2016) 16028, <https://doi.org/10.1038/npjcompumats.2016.28>.
- [17] P. Nath, J.J. Plata, D. Usanmaz, R.A.R.A. Orabi, M. Fornari, M. Buongiorno Nardelli, C. Toher, S. Curtarolo, High-throughput prediction of finite-temperature properties using the quasi-harmonic approximation, *Comput. Mater. Sci.* 125 (2016) 82–91, <https://doi.org/10.1016/j.commatsci.2016.07.043>, URL <<http://www.sciencedirect.com/science/article/pii/S0927025616303718>> .
- [18] F. Legrain, J. Carrete, A. van Roekeghem, S. Curtarolo, N. Mingo, How chemical composition alone can predict vibrational free energies and entropies of solids, *Chem. Mater.* 29 (15) (2017) 6220–6227, <https://doi.org/10.1021/acs.chemmater.7b00789>.

- [19] A. Togo, Phonon Database. <phonondb.mtl.kyoto-u.ac.jp>.
- [20] I. Petousis, W. Chen, G. Hautier, T. Graf, T.D. Schladt, K.A. Persson, F.B. Prinz, Benchmarking density functional perturbation theory to enable high-throughput screening of materials for dielectric constant and refractive index, *Phys. Rev. B* 93 (2016) 115151, <https://doi.org/10.1103/PhysRevB.93.115151>.
- [21] M. De Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C.K. Ande, S. Van Der Zwaag, J.J. Plata, C. Toher, S. Curtarolo, G. Ceder, K.A. Persson, M. Asta, Charting the complete elastic properties of inorganic crystalline compounds, *Sci. Data* 2 (2015) 150009, <https://doi.org/10.1038/sdata.2015.9>.
- [22] M.J. van Setten, M. Giantomassi, X. Gonze, G.-M. Rignanese, G. Hautier, Automation methodologies and large-scale validation for GW: towards high-throughput GW calculations, *Phys. Rev. B* 96 (2017) 155207, <https://doi.org/10.1103/PhysRevB.96.155207>.
- [23] J.J. Plata, D. Usanmaz, P. Nath, C. Toher, J. Carrete, M. Asta, M. de Jong, M. Buongiorno Nardelli, M. Fornari, S. Curtarolo, Predicting the Lattice Thermal Conductivity of Solids by Solving the Boltzmann Transport Equation: AFLOW-AAPL An Automated, Accurate and Efficient Framework. Available from: <[arXiv:1611.05481](https://arxiv.org/abs/1611.05481)>.
- [24] B.B. Karki, R.M. Wentzcovitch, S. de Gironcoli, S. Baroni, High-pressure lattice dynamics and thermoelasticity of MgO, *Phys. Rev. B* 61 (2000) 8793–8800, <https://doi.org/10.1103/PhysRevB.61.8793>.
- [25] A. Togo, L. Chaput, I. Tanaka, G. Hug, First-principles phonon calculations of thermal expansion in Ti_3SiC_2 , Ti_3AlC_2 , and Ti_3GeC_2 , *Phys. Rev. B* 81 (2010) 174301, <https://doi.org/10.1103/PhysRevB.81.174301>.
- [26] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, I. Tanaka, Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization, *Phys. Rev. Lett.* 115 (2015) 205901, <https://doi.org/10.1103/PhysRevLett.115.205901>.
- [27] L. Lindsay, W. Li, J. Carrete, N. Mingo, D.A. Broido, T.L. Reinecke, Phonon thermal transport in strained and unstrained graphene from first principles, *Phys. Rev. B* 89 (2014) 155426, <https://doi.org/10.1103/PhysRevB.89.155426>.
- [28] A.H. Romero, E.K.U. Gross, M.J. Verstraete, O. Hellman, Thermal conductivity in PbTe from first principles, *Phys. Rev. B* 91 (2015) 214310, <https://doi.org/10.1103/PhysRevB.91.214310>.
- [29] E. Bousquet, N.A. Spaldin, P. Ghosez, Strain-induced ferroelectricity in simple rocksalt binary oxides, *Phys. Rev. Lett.* 104 (2010) 037601, <https://doi.org/10.1103/PhysRevLett.104.037601>.
- [30] A. Togo, F. Oba, I. Tanaka, First-principles calculations of the ferroelastic transition between rutile-type and CaCl_2 -type SiO_2 at high pressures, *Phys. Rev. B* 78 (2008) 134106, <https://doi.org/10.1103/PhysRevB.78.134106>.
- [31] S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory, *Rev. Mod. Phys.* 73 (2001) 515–562, <https://doi.org/10.1103/RevModPhys.73.515>.
- [32] A. Jain, G. Hautier, C.J. Moore, S.P. Ong, C.C. Fischer, T. Mueller, K.A. Persson, G. Ceder, A high-throughput infrastructure for density functional theory calculations, *Comput. Mater. Sci.* 50 (8) (2011) 2295–2310, <https://doi.org/10.1016/j.commatsci.2011.02.023>, URL <<http://www.sciencedirect.com/science/article/pii/S0927025611001133>>.
- [33] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty, D. Allan, First-principles computation of material properties: the ABINIT software project, *Comput. Mater. Sci.* 25 (3) (2002) 478–492, [https://doi.org/10.1016/S0927-0256\(02\)00325-7](https://doi.org/10.1016/S0927-0256(02)00325-7), URL <<http://www.sciencedirect.com/science/article/pii/S0927025602003257>>.
- [34] X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté, T. Deutsch, L. Genovese, P. Ghosez, M. Giantomassi, S. Goedecker, D. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. Verstraete, G. Zerah, J. Zwanziger, ABINIT: first-principles approach to material and nanosystem properties, *Comput. Phys. Commun.* 180 (12) (2009) 2582–2615, <https://doi.org/10.1016/j.cpc.2009.07.007>, URL <<http://www.sciencedirect.com/science/article/pii/S0010465509002276>>.
- [35] X. Gonze, F. Jollet, F. Araujo, D. Adams, B. Amadon, T. Applencourt, C. Audouze, J.-M. Beuken, J. Bieder, A. Bokhanchuk, E. Bousquet, F. Bruneval, D. Caliste, M. Cote, F. Dahm, F. Pieve, M. Delaveau, B. Dorado, C. Espejo, G. Geneste, L. Genovese, A. Gerossier, M. Giantomassi, Y. Gillet, D.R. Hamann, L. He, G. Jomard, J. Janssen, S. Roux, A. Levitt, A. Lherbier, F. Liu, I. Lukacevic, A. Martin, C. Martins, M.J.T. Oliveira, S. Ponce, Y. Pouillon, T. Rangel, G.-M. Rignanese, A.H. Romero, B. Rousseau, O. Rubel, A.A. Shukri, M. Stankovski, M. Torrent, M. Setten, B. Troeye, M.J. Verstraete, D. Waroquier, J. Wiktor, B. Xue, A. Zhou, J.W. Zwanziger, Recent developments in the ABINIT software package, *Comput. Phys. Commun.* 205 (2016) 106, <https://doi.org/10.1016/j.cpc.2016.04.003>, M.G. URL <<http://www.sciencedirect.com/science/article/pii/S0010465516300923>>.
- [36] X. Gonze, First-principles responses of solids to atomic displacements and homogeneous electric fields: implementation of a conjugate-gradient algorithm, *Phys. Rev. B* 55 (1997) 10337–10354, <https://doi.org/10.1103/PhysRevB.55.10337>.
- [37] X. Gonze, C. Lee, Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory, *Phys. Rev. B* 55 (1997) 10355–10368, <https://doi.org/10.1103/PhysRevB.55.10355>.
- [38] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.* 77 (1996) 3865–3868, <https://doi.org/10.1103/PhysRevLett.77.3865>.
- [39] L. He, F. Liu, G. Hautier, M.J.T. Oliveira, M.A.L. Marques, F.D. Vila, J.J. Rehr, G.-M. Rignanese, A. Zhou, Accuracy of generalized gradient approximation functionals for density-functional perturbation theory calculations, *Phys. Rev. B* 89 (2014) 064305, <https://doi.org/10.1103/PhysRevB.89.064305>.
- [40] D.R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, *Phys. Rev. B* 88 (2013) 085117, <https://doi.org/10.1103/PhysRevB.88.085117>.
- [41] M. van Setten, M. Giantomassi, E. Bousquet, M. Verstraete, D. Hamann, X. Gonze, G.-M. Rignanese, The Pseudodojo: Training and Grading a 85 Element Optimized Norm-conserving Pseudopotential Table. Available from: <[arXiv:1710.10138](https://arxiv.org/abs/1710.10138)>.
- [42] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I.E. Castelli, S.J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J.K. Dewhurst, I. Di Marco, C. Draxl, M. Dulak, O. Eriksson, J.A. Flores-Livas, K.F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E.K. U. Gross, A. Gulans, F. Gygi, D.R. Hamann, P.J. Hasnip, N.A.W. Holzwarth, D. Iuşan, D.B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Küçükbenli, Y. O. Kvashnin, I.L.M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C.J. Pickard, W. Poelmans, M.I.J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M.J. van Setten, V. Van Speybroeck, J.M. Wills, J.R. Yates, G.-X. Zhang, S. Cottenier, Reproducibility in density functional theory calculations of solids, *Science* 351 (6280), <https://doi.org/10.1126/science.aad3000>. URL <<http://science.sciencemag.org/content/351/6280/aad3000>>.
- [43] W. Setyawan, S. Curtarolo, High-throughput electronic band structure calculations: challenges and tools, *Comput. Mater. Sci.* 49 (2) (2010) 299–312, <https://doi.org/10.1016/j.commatsci.2010.05.010>, URL <<http://www.sciencedirect.com/science/article/pii/S0927025610002697>>.
- [44] Abipy. <github.com/abinit/abipy>.
- [45] Abiflows. <github.com/abinit/abiflows>.
- [46] C. Lee, X. Gonze, Ab initio calculation of the thermodynamic properties and atomic temperature factors of SiO_2 α -quartz and stishovite, *Phys. Rev. B* 51 (1995) 8610–8613, <https://doi.org/10.1103/PhysRevB.51.8610>.
- [47] Y. Li, L. Zhang, T. Cui, Y. Ma, G. Zou, D.D. Klug, Phonon instabilities in rocksalt AgCl and AgBr under pressure studied within density functional theory, *Phys. Rev. B* 74 (2006) 054102, <https://doi.org/10.1103/PhysRevB.74.054102>.
- [48] G. Kresse, J. Furthmüller, J. Hafner, Ab initio force constant approach to phonon dispersion relations of diamond and graphite, *EPL (Europhys. Lett.)* 32 (9) (1995) 729, URL <<http://stacks.iop.org/0295-5075/32/i=9/a=005>>.
- [49] K. Kunc, R.M. Martin, Ab initio force constants of GaAs: a new approach to calculation of phonons and dielectric properties, *Phys. Rev. Lett.* 48 (1982) 406–409, <https://doi.org/10.1103/PhysRevLett.48.406>.
- [50] I. Barin, Thermochemical Data of Pure Substances, Wiley VCH Verlag GmbH, 2008, <https://doi.org/10.1002/9783527619825>.
- [51] B.S. Hemingway, R.A. Robie, J.R. Fisher, W.H. Wilson, Heat capacities of gibbsite, $\text{Al}(\text{OH})_3$, between 13 and 480 K and magnesite, MgCO_3 , between 13 and 380 K and their standard entropies at 298.15 K and the heat capacities of calorimetry conference benzoic acid between 12 and 316 K, *J. Res. USGS* 5 (1977) 797–806.
- [52] R.C. Lord, J.C. Morrow, Calculation of the heat capacity of α quartz and vitreous silica from spectroscopic data, *J. Chem. Phys.* 26 (2) (1957) 230–232, <https://doi.org/10.1063/1.1743274>.
- [53] M. Chase, C. Davies, J. Downey, D. Frurip, R. McDonald, A. Syverud, NIST Standard Reference Database 13, NIST-JANAF Thermochemical Tables Version 1.0. National Institute of Standards and Technology Gaithersburg, MD 20899. <<http://kinetics.nist.gov/janaf/>>.
- [54] L. Huang, X. Wang, B. Ao, Lattice dynamics and equilibrium properties of boron phosphide: a first-principles study, *J. Phys.: Condens. Matter* 16 (43) (2004) 7829, URL <<http://stacks.iop.org/0953-8984/16/i=43/a=022>>.
- [55] J. Holm, O. Kleppa, E.F. Westrum, Thermodynamics of polymorphic transformations in silica. thermal properties from 5 to 1070° K and pressure-temperature stability fields for coesite and stishovite, *Geochim. Cosmochim. Acta* 31 (12) (1967) 2289–2307, [https://doi.org/10.1016/0016-7037\(67\)90003-8](https://doi.org/10.1016/0016-7037(67)90003-8), URL <<http://www.sciencedirect.com/science/article/pii/S0016703767900038>>.
- [56] S.-G. Ma, Y.-H. Shen, T. Gao, P.-H. Chen, First-principles calculation of the structural, electronic, dynamical and thermodynamic properties of γ - LiAlO_2 , *Int. J. Hydrogen Energy* 40 (9) (2015) 3762–3770, <https://doi.org/10.1016/j.ijhydene.2015.01.088>, URL <<http://www.sciencedirect.com/science/article/pii/S0360319915001561>>.
- [57] T. Gürel, R. Eryiit, Ab initio lattice dynamics and thermodynamics of rare-earth hexaborides LaB_6 and CeB_6 , *Phys. Rev. B* 82 (2010) 104302, <https://doi.org/10.1103/PhysRevB.82.104302>.
- [58] H. Zhang, Y. Yu, Y. Zhao, W. Xue, T. Gao, Ab initio electronic, dynamic, and thermodynamic properties of isotopic lithium hydrides (^6LiH , ^6LiD , ^6LiT , ^7LiH , ^7LiD , and ^7LiT), *J. Phys. Chem. Solids* 71 (7) (2010) 976–982, <https://doi.org/10.1016/j.jpccs.2010.03.037>, URL <<http://www.sciencedirect.com/science/article/pii/S0022369710000971>>.
- [59] G. Hautier, S.P. Ong, A. Jain, C.J. Moore, G. Ceder, Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability, *Phys. Rev. B* 85 (2012) 155208, <https://doi.org/10.1103/PhysRevB.85.155208>.