Fast and accessible first-principles calculations of vibrational properties of materials

Timur Bazhirov and E. X. Abot\textsuperscript{1,2}

\textsuperscript{1}Exabyte Inc., San Francisco, California 94103, USA

We present example applications of an approach to first-principles calculations of vibrational properties of materials implemented within the Exabyte platform\textsuperscript{11}. We deploy models based on the Density Functional Perturbation Theory to extract the phonon dispersion relations and densities of states for an example set of 35 samples and find the results to be in agreement with prior similar calculations. We construct modeling workflows that are both accessible, accurate, and efficient with respect to the human time involved. This is achieved through efficient parallelization of the tasks for the individual vibrational modes. We report achieved speedups in the 10-100 range, approximately, and maximum attainable speedups in the 30-300 range, correspondingly. We analyze the execution times on the current up-to-date computational infrastructure centrally available from a public cloud provider. Results and all associated data, including the materials and simulation workflows, are made available online in an accessible, repeatable and extensible setting.

\section{I. INTRODUCTION}

Properties associated with the dynamics of a crystal lattice are of key importance for the design and discovery of new materials from nanoscale. Such properties can be computationally obtained by studying phonons - collective excitation in a periodic arrangement of atoms characterized by the modes of vibration and the corresponding frequencies. The phonon dispersion relations facilitate the calculation of many transport properties: heat capacity and vibrational contribution to the entropy of states for an example set of 35 samples and find the results to be in agreement with prior similar calculations. We construct modeling workflows that are both accessible, accurate, and efficient with respect to the human time involved. This is achieved through efficient parallelization of the tasks for the individual vibrational modes. We report achieved speedups in the 10-100 range, approximately, and maximum attainable speedups in the 30-300 range, correspondingly. We analyze the execution times on the current up-to-date computational infrastructure centrally available from a public cloud provider. Results and all associated data, including the materials and simulation workflows, are made available online in an accessible, repeatable and extensible setting.

\section{II. METHODOLOGY}

\subsection{A. General logic}

We demonstrate the general execution flow employed in this work in Ref\textsuperscript{2}. We employ the framework explained therein in order to construct the simulation workflows for the calculation of the vibrational properties discussed here. The users of Exabyte platform can clone the associated entities (eg. materials, workflows) - and re-create our calculations in order to reproduce or further
improve the results.

B. Materials

All materials studied in this work constitute a subset of the ESC-71 set from\textsuperscript{2} with the total count of 35. The details about the materials studied are given in Fig I. Our selection is based initially on the widely used semi-conducting compounds that have relatively small crystal unit cells.

| Formula | MP id | $N_{at}$ | Formula | MP id | $N_{at}$ |
|---------|-------|----------|---------|-------|----------|
| Si      | mp-149| 2        | Li$_2$O | mp-1960| 3        |
| Ge      | mp-32 | 2        | BN      | mp-7991| 4        |
| Bi      | mp-23152| 2      | AlN     | mp-661 | 4        |
| Sn      | mp-117| 2        | CaO     | mp-545512| 4    |
| BP      | mp-1479| 2       | MgSe    | mp-1018040| 4   |
| GaP     | mp-2490| 2       | MgTe    | mp-1039 | 4        |
| AlAs    | mp-2172| 2       | SnO$_2$ | mp-856 | 6        |
| GaAs    | mp-2534| 2       | AlGaAs  | N/A    | 8        |
| GaN     | mp-830 | 2        | InGaAs  | N/A    | 8        |
| YN      | mp-2114| 2       | InGaP   | N/A    | 8        |
| ZnS     | mp-10695| 2      | AlInAs  | N/A    | 8        |
| BeSe    | mp-1541| 2       | AlInSb  | N/A    | 8        |
| MgO     | mp-1265| 2       | GaAsP   | N/A    | 8        |
| MgS     | mp-1315| 2       | AlGaN   | mp-1019508| 8 |
| ZnO     | mp-2229| 2       | B2O$_3$ | mp-717 | 10       |
| BeO     | mp-1778| 2       | Al$_2$O$_3$ | mp-1143| 10       |
| BaSe    | mp-1253| 2       | B       | mp-160 | 12       |
| CaSe    | mp-1415| 2       |         |        |          |

Table I. Materials studied in this work with the corresponding identificators from Ref.\textsuperscript{5}. $N_{at}$ - number of sites (atoms) in the crystal unit cell. Consult Ref.\textsuperscript{2} for additional details.

C. Workflows

We implement a grid-parallel workflow for the calculation of the phonon dynamical matrices initially explained in Ref\textsuperscript{10} and demonstrated in Figs. II C and II C with an additional optional preceding step for a variable-cell relaxation. During the phonon calculation part the following happens:

1. First, the irreducible representations for the vibrational modes (irreps) are generated based on the sampling grid in the reciprocal space (q-point grid). Full symmetry analysis is not performed in the current implementation.

2. Second, a separate calculation is prepared and submitted for execution to the cloud infrastructure manager per each irreducible representation ("map" stage).

3. Next, the computational infrastructure is provisioned on-demand at a cloud provider with a cap on the total number of nodes as explained further in this section.

4. Finally, after the calculations for all irreps are finished, the dynamical matrices are collected and phonon dispersions and density of states are calculated ("reduce" stage).

Thus, we employ a "map-reduce" type embarrassingly parallel scenario, and couple the calculations of the individual phonon dispersion modes with the allocation of computational resources on the cloud. This allows for the improved efficiency and speedup, such that the limiting phase in the total calculation is the longest run per individual irreducible representation. Unlike the previously considered workflows categorization based on the inclusion of the semi-core states, spin-orbit coupling and magnetism\textsuperscript{2}, in this work we omit the considerations of the latter two and only include semi-cores as it is implemented in the GBRV pseudopotential set\textsuperscript{19}.

D. Computational setup

We use Density Functional Theory\textsuperscript{20} in the planewave pseudopotential formalism\textsuperscript{21} as implemented in Quantum ESPRESSO (QE) package\textsuperscript{11}. Within the generalized gradient approximation the exchange-correlation effects were modeled using the Perdew-Berke-Ernzerhof (PBE)\textsuperscript{21} functional. The ultrasoft GBRV pseudopotentials at version 1.5\textsuperscript{19} with the recommended cutoff values of 40 Ry and 200 Ry are used for the electronic wavefunctions and electronic densities correspondingly. We implemented sampling in the reciprocal cell based on k-points per reciprocal atom (KPPRA) with a uniform unshifted grid. A minimum KPPRA of 1,600 was used for the electronic structure calculations. The phonon properties are calculated on an grid that corresponds to a minimum QPPRA of 1,800 was used for the electronic structure calculations. The phonon properties are calculated on an grid that corresponds to a minimum QPPRA of 200. The Fourier transform and subsequent interpolation to an effective IPPRA (interpolated points per reciprocal atom) of 12,800 was used, as it is implemented in QE through "q2r" and "matdyn" modules.

All calculations were performed using the hardware available from Microsoft Azure cloud computing service\textsuperscript{22} in the same manner as described in Refs.\textsuperscript{2,3}, except for the use of "F16" instances for the current work. Computational resources were provisioned and assembled on-demand by software implemented and available within the Exabyte platform\textsuperscript{1}. All runs were executed by a single person within a one-week period in June 2018, which emphasizes the power of the underlying general approach to materials modeling implemented in Exabyte platform. The peak size of the computational infrastructure used during this work was administratively limited to 200 nodes or 3,200 total computing cores.
E. Data access and repeatability

The materials, workflows, batch jobs for each material with the associated properties, and files for each step of the simulation workflows are all freely available online at the link in Ref.\textsuperscript{18}. Readers interested in repeating or improving upon our work may create an account, copy materials and/or workflows to their account collection, and recreate the simulation for this materials. An example simulation workflow, as employed in the current work, is presented in Fig. II C. Readers can see the map-reduce type logic included in the workflow, where the individual calculation tasks are performed independently in parallel in order to speed up the execution. Example results for InGaAs and B, in the same exact representation as can be accessed through Ref.\textsuperscript{18}, are shown in Fig. III A and Fig. III A correspondingly.

III. RESULTS AND DISCUSSION

A. Results and comparison with prior calculations

The results for all the materials studied in this work are available online at the link in Ref.\textsuperscript{18}. Fig. 5 shows a comparison of the calculated phonon density of states for a subset of 9 compounds, including boron, GaN, GaAs, ZnS, Al\textsubscript{2}O\textsubscript{3}, MgTe, BeO, SiO\textsubscript{2}, SnO\textsubscript{2}, with the results of
the Materials Project\textsuperscript{2} (further referred to as MP). As it can be seen, the results are in agreement with each other in the overall shape, with a small (1-5\%) shift toward the high frequency range in MP case. We attribute this shift to the use of PBEsol\textsuperscript{23} functional in their work, versus PBE\textsuperscript{21} for our calculations.

### B. Goals

We meant this study as a demonstration of the capabilities of Exabyte platform in deploying the density functional perturbation theory (DFPT) tools to predict the vibrational properties of materials. We also focused our attention on how it can be accelerated and applied in an accessible way with minimal additional computational setup (i.e. no specialized hardware or compilation routines). We elaborate on the results of our prior work\textsuperscript{2} and extend the spectrum of materials properties available.

### C. Further improvements

We can consider multiple ways to further improve the results presented in this work. Firstly, as mentioned in\textsuperscript{4}, employing PBEsol instead of the more widely used (at the moment of this writing) PBE functional for the treatment of the exchange-correlation in the materials studied might be beneficial. According to the comparison presented in Fig. 5 this effect is expected to be small - within 1-5\%. Secondly, a rigorous relaxation routine, together with a more extensive convergence study for the reciprocal cell sampling in each of the computed cases might be beneficial. As the magnitude of the phonon frequencies is relatively small (meV range), the artifacts related to convergence can sometimes lead to the artificially present negative frequencies\textsuperscript{14}. We originally attempted to include all materials from the ESC-71 set\textsuperscript{2}, and left some out due to the time constraints related to resolving the presence of negative frequencies. Lastly, an improvement in speed and efficiency of the modeling workflows may be achieved by further optimizing the coupling of the computational infrastructure to the individual calculations per each irreducible representation. As the Table III D demonstrates, the minimum attainable runtime with the "map-reduce" workflow can be as much as 374 shorter than for a sequential, while we practically achieved speedups up to 134.

### D. Computational time and cost

We present the analysis of the runtimes for the different calculation scenarios. First, we list the total "sequential"...
Figure 3. Example results for an 8-atom InGaAs cell, as available within the Exabyte platform\textsuperscript{13}. The top panel lists parameters for the initial self-consistent field calculation, and the bottom two contain the plots for the phonon density of states and the phonon dispersions.

execution time $t_{seq}$ that a phonon calculation would take without parallelizing the tasks for irreducible representations. In practice, this would mean confining the simulation to a single computing node with 16 cores as explained in section II. Next, we present the corresponding speedup ratios for the actual runtime recorded $t_{act}$ and for the minimum attainable runtime $t_{min}$ corresponding

Figure 4. Results page for a 12-atom boron calculation.
Figure 5. Comparative plot of the results for the phonon dispersions calculated in this work and available in MP database.
(a) - boron, (b) - GaN, (c) - GaAs, (d) - ZnS, (f) - sapphire Al₂O₃, (e) - MgTe, (f) - BeO, (g) - SiO₂, (h) - SnO₂. (Left panels for each sub-figure show MP data and contain the projected density of states for each atom. Right sub-figure panels has the total phonon density of states calculated in this work.)
Table II. Average calculation times for a selected subset of materials. $N_{\text{at}}$ - number of sites (atoms) in the crystal unit cell. $t_{\text{seq}}$ - phonon calculation runtime for a sequential mode, where all irreducible representations are calculated one-by-one (leftmost branch in Fig. II C). $t_{\text{act}}$ - actual runtime as recorded in Exabyte platform. $t_{\text{min}}$ - minimum possible runtime for the phonon part, equal to the maximum calculation time per a single irreducible representation. Approximate costs are given as well.

| Material   | $N_{\text{at}}$ | $t_{\text{seq}}$ (hr) | $t_{\text{seq}}/t_{\text{act}}$ | $t_{\text{act}}/t_{\text{min}}$ | Cost ($) |
|------------|-----------------|------------------------|---------------------------------|---------------------------------|---------|
| MgO        | 2               | 44.1                   | 134                             | 174                             | 100     |
| MgSe       | 4               | 66.3                   | 30                              | 248                             | 150     |
| AlGaN      | 8               | 18.6                   | 13                              | 34                              | 50      |
| B          | 12              | 102.3                  | 117                             | 374                             | 250     |

to the longest run among all irreps. As can be seen from the table, we achieved speedups in the 13-134 range in practice, while corresponding maximum speedups possible are 34-374. We also present the associated costs, which, due to the elastic nature of cloud computing, do not depend on the specific calculation scenario.

E. Future outlook

The landscape of computational materials design is rapidly evolving toward a data-driven science, with multiple initiatives contributing toward the automated aggregation and categorization of materials properties. Major improvements in the way computational materials science is used would be possible when the range of materials properties feasible for calculation is extended to include the vibrational spectra and related. The approach described in this work can assist with the above. This work demonstrates that high-fidelity data about the vibrational properties, perhaps only for the electronic materials at this moment, is readily attainable in an accessible and repeatable manner. Our intent is to welcome collaborative contributions in order to, firstly, further grow the online repository of the results; secondly, allow contributions from other modeling techniques beyond studied here; and, finally, facilitate the creation of statistical (machine learning) models based on the available data.

IV. CONCLUSIONS

We present the applications of a novel approach to materials modeling from nanoscale implemented within the Exabyte platform and capable of rapidly delivering results about the vibrational properties of materials in an accessible and data-centric manner. We apply this approach to a set of 35 materials in order to demonstrate how it works. We report the results for the phonon densities of states and phonon dispersions obtained using the Density Functional Perturbation Theory within the Generalized Gradient Approximation (GGA). We compare the results with prior similar calculation attempts and discuss the corresponding computational costs and pathways to further improvements.

We demonstrate how computationally demanding task of calculating the phonon frequencies, that would otherwise take from 18 to 102 hours on an up-to-date high-performance computing server, can be accelerated by a factor of 13-134 such that the resulting runtime fits within one hour. We present not only the results and the associated data, but also an easy-to-access way to reproduce and extend the results by means of the Exabyte platform. Our work provides an accessible and repeatable practical recipe for performing high-fidelity first-principles calculations of the vibrational properties of materials in a high-throughput manner.

---

1 Exabyte.io: materials discovery cloud (2018).
2 Protik Das, Mohammad Mohammad, and Timur Bazhirov, “Accessible computational materials design with high fidelity and high throughput,” arxiv.org/abs/1807.05623 (2018).
3 Protik Das and Timur Bazhirov, “Electronic properties of binary compounds with high throughput and high fidelity,” arxiv.org/abs/1808.05325 (2018).
4 Guido Petretto, Shyam Dwaraknath, Henrique P.C. Miranda, Donald Winston, Matteo Gianomassi, Michiel J. van Setten, Xavier Gonze, Kristin A. Persson, Geoffrey Hautier, and Gian-Marco Rignanese, “High-throughput density-functional perturbation theory phonons for inorganic materials,” Scientific Data 5, 180065 EP – (2018).
5 Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al., “Commentary: The materials project: A materials genome approach to accelerating materials innovation,” Apl Materials 1, 011002 (2013).
6 Stefano Curtarolo, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kesong Yang, Richard H Taylor, Lance J Nelson, Gus IW Hart, Stefano Sanvito, Marco Buongiorno-Nardelli, et al., “Aflowlib. org: A distributed materials properties repository from high-throughput ab initio calculations,” Computational Materials Science 58, 227–235 (2012).
7 James E Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and Christopher Wollerton, “Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd),” Jom 65, 1501–1509 (2013).
8 Giovanni Pizzi, Andrea Cepellotti, Riccardo Sabatini, Nicola Marzari, and Boris Kozinsky, “Aiida: automated
interactive infrastructure and database for computational science,” Computational Materials Science **111**, 218–230 (2016).

9 - **NOMAD laboratory: A European Centre of Excellence** (2018).

10 - P. Giannozzi R. di Meo, A. Dal Corso and S. Cozzini, “Calculation of phonon dispersions on the grid using quantum espresso,” ICTP lecture notes 24, 163 24, 163 (2009).

11 - Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiocchio, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Cristos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch, “Quantum espresso: a modular and open-source software project for quantum simulations of materials,” *Journal of Physics: Condensed Matter* **21**, 395502 (2009).

12 - A Togo and I Tanaka, “First principles phonon calculations in materials science,” Scr. Mater. **108**, 1–5 (2015).

13 - Jose J. Plata, Pinku Nath, Demet Usanmaz, Jesús Carrete, Cormac Toher, Maarten de Jong, Mark Asta, Marco Fornari, Marco Buongiorno Nardelli, and Stefano Curtarolo, “An efficient and accurate framework for calculating lattice thermal conductivity of solids: Aflow-aapl automatic anharmonic phonon library,” npj Computational Materials **3**, 45 (2017).

14 - Geoffroy Hautier Gian-Marco Rignanese Guido Petretto, Xavier Gonze, “Convergence and pitfalls of density functional perturbation theory phonon calculations from a high-throughput perspective,” arxiv.org/abs/1710.06028 (2017).

15 - Xavier Gonze, “Adiabatic density-functional perturbation theory,” *Phys. Rev. A* **52**, 1096–1114 (1995).

16 - P. Hohenberg and W. Kohn, “Inhomogeneous electron gas,” *Phys. Rev.* **136**, B864–B871 (1964).

17 - J Ihm, A Zunger, and M.L. Cohen, “Momentum-space formalism for the total energy of solids,” *Journal of Physics C: Solid State Physics* **12**, 4409 (1979).

18 - Exabyte platform: project URL with data about simulations (2018).

19 - Kevin F. Garrity, Joseph W. Bennett, Karin M. Rabe, and David Vanderbilt, “Pseudopotentials for high-throughput dft calculations,” *Computational Materials Science* **81**, 446–452 (2014).

20 - Walter Kohn and Lu Jeu Sham, “Self-consistent equations including exchange and correlation effects,” *Physical review* **140**, A1133 (1965).

21 - John P Perdew, Kieron Burke, and Matthias Ernzerhof, “Generalized gradient approximation made simple,” *Physical review letters* **77**, 3865 (1996).

22 - Microsoft Azure Cloud Computing platform: web page.

23 - John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, Oleg A. Vydrov, Gustavo E. Scuseria, Lucian A. Constantin, Xiaolun Zhou, and Kieron Burke, “Restoring the density-gradient expansion for exchange in solids and surfaces,” *Phys. Rev. Lett.* **100**, 136406 (2008).