TOPICAL REVIEW • OPEN ACCESS

Implementation strategies in phonopy and phono3py

OPEN ACCESS
IOP Publishing

Journal of Physics: Condensed Matter

J. Phys.: Condens. Matter **35** (2023) 353001 (22pp)

https://doi.org/10.1088/1361-648X/acd831

Topical Review

Implementation strategies in phonopy and phono3py

Atsushi Togo[1,2,*], **Laurent Chaput**[3], **Terumasa Tadano**[4] **and Isao Tanaka**[2,5,6]

[1] Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan
[2] Center for Elements Strategy Initiative for Structural Materials, Kyoto University, Sakyo, Kyoto 606-8501, Japan
[3] Université de Lorraine, LEMTA, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7563, BP 70239, 54506 Vandœuvre Cedex, France
[4] Research Center for Magnetic and Spintronic Materials, National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan
[5] Department of Materials Science and Engineering, Kyoto University, Sakyo, Kyoto 606-8501, Japan
[6] Nanostructures Research Laboratory, Japan Fine Ceramics Center, Atsuta, Nagoya 456-8587, Japan

E-mail: togo.atsushi@gmail.com

Received 13 January 2023, revised 17 April 2023
Accepted for publication 23 May 2023
Published 2 June 2023

Abstract

Scientific simulation codes are public property sustained by the community. Modern technology allows anyone to join scientific software projects, from anywhere, remotely via the internet. The phonopy and phono3py codes are widely used open-source phonon calculation codes. This review describes a collection of computational methods and techniques implemented in these codes and shows their implementation strategies as a whole, aiming to be useful for the community. Some of the techniques presented here are not limited to phonon calculations and may therefore be useful in other areas of condensed matter physics.

Keywords: phonon calculation, computer simulation, software development

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

1. Introduction

Supported by the exponential growth of computer power, computer simulations and scientific software development are essential for condensed matter physics and materials science. Results obtained from computer simulations are getting more realistic and scientific software development more robust.

Since many scientific codes are distributed under open-source licenses, it is easy to start using scientific software and perform computations. Participation in scientific software development has also become easier with the availability of open-source compilers and source code editors. Collaborative development by members at different worksites or locations, i.e. *distributed* development, has become quite popular for scientific software these days. Anyone can contribute to open-source software projects, and the contributions are usually united on internet hosting services for software development. Documentation of scientific software is an important part of software development, not only for users but also for scientific software developers. For users, it explains how to use the software. For developers, it describes how to read and write the code.

[*] Author to whom any correspondence should be addressed.

Development of key software can be a project that lasts for many years. Once a key code is developed, it may be used for an unexpectedly long time. However, computer architectures are constantly evolving. The way to achieve high-performance computing may therefore change drastically in a relatively short time, on the order of a decade. For example, it is always required to follow the increase in the number of cores in processors. Indeed, algorithms and data structures must be optimized for a concurrent use of many cores in a multi-core processor. Increasing memory space usually allows more flexible software design. To adapt the code to those changes, the documentation is important to share the meaning of the software design choices among users and developers.

The phonopy and phono3py codes are scientific software developed to perform phonon calculations. A variety of physical properties can be calculated using them: the phonon spectrum, the dynamical structure factor, and the lattice thermal conductivity (LTC), to mention just a few. The computational method is based on the supercell approach. In the computer implementations, various numerical methods and techniques are employed. Some of those implemented in the phonopy and phono3py codes are covered in this review. Our motivation for writing this review is to provide essential information to understand the codes in depth and to invite scientific software developers in the phonopy and phono3py projects.

Although this review is written targeting scientific software developers, it may be also useful for expert users of the phonopy and phono3py codes. The computational methods and techniques underlying the phonon calculations are described as implemented to the phonopy and phono3py codes.

In section 2, the representation of the crystal structure and the crystal symmetry are presented in a crystallographic way. The reciprocal space of the crystal and the crystal symmetry in the reciprocal space are also described. Then, the phonon coordinates are introduced. The concept of Brillouin zone (BZ) required for the phonon calculation is briefly defined.

In section 3, the geometry of the supercell structure model is described. The supercell geometry is associated with the primitive cell by an integer matrix and how to construct the supercell geometry using the integer matrix is explained.

In section 4, the transformation between the supercell force constants and the dynamical matrices is presented. Since the phonopy and phono3py codes employ the supercell approach, the force constants elements are limited to within the supercell. The transformation, therefore, needs to be performed with special care. Such details are given in this section.

In section 5, the treatment of long range dipole–dipole interaction in the dynamical matrix, as implemented in the phonopy code, is explained. The implementation is mainly based on [1], and its application to the supercell approach is reviewed.

In section 6, the traditional and generalized regular grids used to sample reciprocal space discretely are defined. The symmetry treatment of those grid points, used in the phonopy and phono3py codes, is also presented.

In section 7, a linear tetrahedron method is described using the formulation as implemented in the phonopy and phono3py codes. The implementation is mainly based on [2, 3]. The method is then applied to the three-phonon scattering.

In section 8, a scheme to generate random atomic displacements in the supercell at finite temperatures is presented. This is achieved by superpositions of displacements corresponding to randomly sampled harmonic oscillators.

In section 9, a band unfolding technique implemented in the phonopy code following [4] is explained. The formulation specific to the phonon calculation with the supercell approach is provided.

2. Crystal and symmetry

2.1. Crystal structure

As shown in figure 1, a crystal model is defined by a collection of unit cells periodically repeated in the three directions of space. Each unit cell contains atoms. We normally choose a conventional set of basis vectors to span the unit cell, $(\mathbf{a},\mathbf{b},\mathbf{c})$, so that it naturally represents the crystal symmetry along with the choice of the origin of atomic positions [5]. Positions of atoms are represented either in Cartesian coordinates or in crystallographic coordinates. Denoting the crystallographic coordinates of an atom as $\boldsymbol{x} = (x_1, x_2, x_3)^\intercal$, its position $\mathbf{R}$ is given by

$$\begin{aligned}\mathbf{R} &= x_1\mathbf{a} + x_2\mathbf{b} + x_3\mathbf{c}\\ &= (\mathbf{a},\mathbf{b},\mathbf{c})\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}\\ &= (\mathbf{a},\mathbf{b},\mathbf{c})\boldsymbol{x}.\end{aligned} \tag{1}$$

In the above equation, if the basis vectors are represented using column vectors containing their Cartesian components, then $(\mathbf{a},\mathbf{b},\mathbf{c})$ becomes a matrix, and $\mathbf{R}$ a column vector containing the Cartesian coordinates of the atomic position vector,

$$\begin{pmatrix}\mathbf{R}_x\\ \mathbf{R}_y\\ \mathbf{R}_z\end{pmatrix} = \begin{pmatrix}\mathbf{a}_x & \mathbf{b}_x & \mathbf{c}_x\\ \mathbf{a}_y & \mathbf{b}_y & \mathbf{c}_y\\ \mathbf{a}_z & \mathbf{b}_z & \mathbf{c}_z\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}. \tag{2}$$

In the following we will use both representations. Depending on the context, $(\mathbf{a},\mathbf{b},\mathbf{c})$ can therefore be considered as a row vector containing three elements of a vector space, or a 3 by 3 matrix.

It is assumed that elements of $\boldsymbol{x}$ are in the interval $[0,1)$ to describe the structure. However it is often the case that we have to compare two positions of atoms which may be in different unit cells, possibly after a space group operation. For example, to identify two coordinates $\boldsymbol{x}$ and $\boldsymbol{x}'$ which correspond to the same location within the unit cell, but are possibly shifted by a lattice vector, it is convenient to bring each element of the difference $\Delta\boldsymbol{x} = \boldsymbol{x}' - \boldsymbol{x}$ into the interval $[-0.5, 0.5)$ and then confirm that the length of $\Delta\boldsymbol{x}$ is smaller than a tolerance ϵ. This

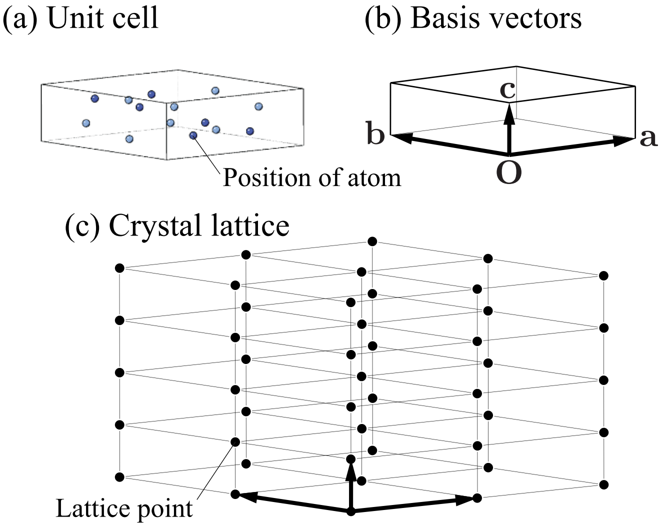

Figure 1. Crystal structure model of β-Si$_3$N$_4$. See also figure 2 for the space group symmetry. (a) Unit cell containing atoms. (b) Basis vectors of the unit cell. (c) Crystal lattice. The filled circle symbols show the lattice points. In this model, the unit cell is assumed to be primitive.

is implemented in the phonopy and phono3py codes everywhere by rounding components of $\Delta\boldsymbol{x}$ to the nearest integer $(\mathrm{nint}(\Delta\boldsymbol{x}))$ and checking $|(\mathbf{a},\mathbf{b},\mathbf{c})[\Delta\boldsymbol{x}-\mathrm{nint}(\Delta\boldsymbol{x})]|<\epsilon$.

2.2. Space group operation

Various symmetries of crystals and phonons are used in the phonopy and phono3py codes. The most important one is the space group symmetry. A space group operation is composed of a (proper or improper) rotation $\mathcal{S}$ and a translation τ. It is often written with a composite notation $(\mathcal{S},\tau)$ which sends position $\mathbf{R}$ to

$$\mathbf{R}'=(\mathcal{S},\tau)\mathbf{R}\equiv\mathcal{S}\mathbf{R}+\tau. \tag{3}$$

Using crystallographic coordinates this is written as

$$(\mathbf{a},\mathbf{b},\mathbf{c})\boldsymbol{x}'=\mathcal{S}(\mathbf{a},\mathbf{b},\mathbf{c})\boldsymbol{x}+(\mathbf{a},\mathbf{b},\mathbf{c})\boldsymbol{w} \tag{4}$$

or

$$\begin{aligned}\boldsymbol{x}'&=(\mathbf{a},\mathbf{b},\mathbf{c})^{-1}\mathcal{S}(\mathbf{a},\mathbf{b},\mathbf{c})\boldsymbol{x}+\boldsymbol{w}\\&\equiv\boldsymbol{S}\boldsymbol{x}+\boldsymbol{w},\end{aligned} \tag{5}$$

where $\boldsymbol{S}$ can be shown to be a unimodular matrix. $\boldsymbol{w}$ contains the components of the translation vector in the crystallographic basis. If Cartesian coordinates were used instead, the matrix representing the rotation would be orthogonal. Representing the space group by $\mathbb{S}=\{(\mathcal{S},\tau)\}$, the crystallographic point group is given by $\mathbb{P}=\{\mathcal{S}|(\mathcal{S},\tau)\in\mathbb{S}\}$. An example of space group operation is shown in figure 2. The crystal structure overlaps with itself after a rotation of 60° along the z axis and a $+\frac{1}{2}\mathbf{c}$ shift. Therefore

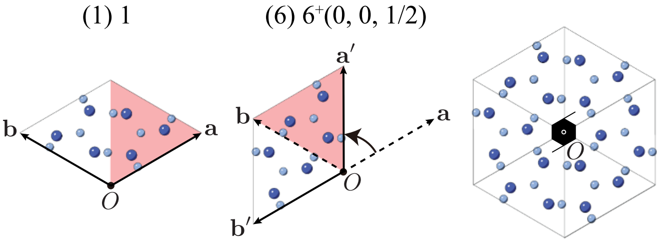

Figure 2. Symmetry operations applied on the unit cell of β-Si$_3$N$_4$ whose space-group type is $P6_3/m$ (No. 176). The left and middle figures illustrate identity and six-fold screw axis operations. Their notations follow the international tables for crystallography volume A [5]. The right figure shows the crystal structure viewed from the top (z-axis) that depicts the six-fold screw axis passing through the origin O.

$$\boldsymbol{S}=\begin{pmatrix}1&\bar{1}&0\\1&0&0\\0&0&1\end{pmatrix},\boldsymbol{w}=\begin{pmatrix}0\\0\\1/2\end{pmatrix}, \tag{6}$$

must be a space group operation. It means that any atom in the unit cell at point $\boldsymbol{x}$ is sent by the space group operation to a new point

$$\boldsymbol{x}'=\begin{pmatrix}1&\bar{1}&0\\1&0&0\\0&0&1\end{pmatrix}\boldsymbol{x}+\begin{pmatrix}0\\0\\1/2\end{pmatrix}. \tag{7}$$

The new point $\boldsymbol{x}'$ can be located out of the original unit cell, but an atom with the same atomic type must be found at this new location.

2.3. Primitive cell

In most cases, a unit cell as described in section 2.1 is used as the input crystal structure model of a phonon calculation. The most plausible choice of the unit cell is either a primitive cell or a conventional unit cell. The former represents the minimum unit of periodicity of a crystal lattice. It is therefore the one used for Fourier expansions and is necessary for a reciprocal space representation of phonon properties. The Bloch theorem evidenced this lattice translational symmetry. The latter follows the crystallographic convention. Although it provides intuitive shapes of the unit cells (cubic, hexagonal, etc), it may contain several lattice points. Figure 3 shows the conventional unit cell and a primitive cell for silicon crystallized within the face-centred-cubic structure.

The conventional unit cell structure is uniquely defined to follow the crystallographic convention, [5] within the freedom due to Euclidean normalizer [5]. On the other hand, the choice of the primitive cell is not unique. This is inconvenient for a systematic handling of crystal structure models in phonon calculations. Therefore, phonopy suggests a predefined transformation matrix $\boldsymbol{P}_{\mathrm{prim}}$ from the conventional unit cell to the primitive cell. $\boldsymbol{P}_{\mathrm{prim}}$ is used as

$$(\mathbf{a}_{\mathrm{p}},\mathbf{b}_{\mathrm{p}},\mathbf{c}_{\mathrm{p}})=(\mathbf{a}_{\mathrm{c}},\mathbf{b}_{\mathrm{c}},\mathbf{c}_{\mathrm{c}})\boldsymbol{P}_{\mathrm{prim}}, \tag{8}$$

Table 1. Choices of transformation matrices $\boldsymbol{P}_{\text{prim}}$ of equation (8) used in the phonopy and phono3py codes. The subscripts X of the matrices $\boldsymbol{P}_X$ indicate the centring types: A, B, C for the base centring types, I and F for the body and face centring types, respectively, and R for the (obverse) rhombohedral centring type.

$$\boldsymbol{P}_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{\bar{1}}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \boldsymbol{P}_B = \begin{pmatrix} \frac{1}{2} & 0 & \frac{\bar{1}}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \quad \boldsymbol{P}_C = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{\bar{1}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{P}_I = \begin{pmatrix} \frac{\bar{1}}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{\bar{1}}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{\bar{1}}{2} \end{pmatrix}, \quad \boldsymbol{P}_F = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}, \quad \boldsymbol{P}_R = \begin{pmatrix} \frac{2}{3} & \frac{\bar{1}}{3} & \frac{\bar{1}}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{\bar{2}}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

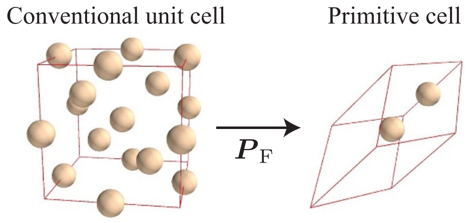

Figure 3. Conventional unit cell and primitive cell of silicon. The space group type is $Fd\bar{3}m$ (No. 227). Since the conventional unit cell contains four lattice points with eight atoms, the primitive cell is chosen to have two atoms by $\det(\boldsymbol{P}_F) = 1/4$.

where $(\mathbf{a}_{\text{p}}, \mathbf{b}_{\text{p}}, \mathbf{c}_{\text{p}})$ are the basis vectors of the primitive cell and $(\mathbf{a}_{\text{c}}, \mathbf{b}_{\text{c}}, \mathbf{c}_{\text{c}})$ those of the conventional unit cell. The choices of the transformation matrices used in the phonopy and phono3py codes are shown in table 1. For example, the transformation matrix used for silicon in figure 3 is $\boldsymbol{P}_{\text{prim}} = \boldsymbol{P}_F$. These matrices are equivalent to those presented in table 2 of [6] although those are given for the reciprocal space basis vectors.

The crystal lattice is defined as the set of integer linear combinations of primitive lattice vectors. A lattice vector $\mathbf{R}_l$ can therefore be written as

$$\mathbf{R}_l = l_1\mathbf{a}_{\text{p}} + l_2\mathbf{b}_{\text{p}} + l_3\mathbf{c}_{\text{p}} = (\mathbf{a}_{\text{p}}, \mathbf{b}_{\text{p}}, \mathbf{c}_{\text{p}})\boldsymbol{l} \;\; (l_i \in \mathbb{Z}). \tag{9}$$

In figure 1, if the unit cell is assumed to be primitive, then figure 1(c) shows the crystal lattice. The lattice point may then be chosen to coincide with the origin of each unit cell. Sometimes it is also useful to regard the crystal lattice as generated using the lattice translations of the space group, $(\boldsymbol{I}, \boldsymbol{t})$, where $\boldsymbol{I}$ and $\boldsymbol{t}$ denote the identity operation and lattice translation, respectively.

2.4. Reciprocal space

The crystal structure, and the symmetry explained above, are defined in direct (real) space. In this section, crystallography is briefly summarized in reciprocal space. This is quite useful for our purpose since phonons are described in reciprocal space using wave vectors. It is convenient to introduce a reciprocal lattice whose basis vectors $(\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}})$ are solutions of the equation

$$(\mathbf{a}_{\text{p}}, \mathbf{b}_{\text{p}}, \mathbf{c}_{\text{p}})^{\intercal}(\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}}) = 2\pi\mathbf{1}. \tag{10}$$

This equation can be solved to give

$$\mathbf{a}^*_{\text{p}} = 2\pi\frac{\mathbf{b}_{\text{p}} \times \mathbf{c}_{\text{p}}}{V_{\text{c}}}, \;\; \mathbf{b}^*_{\text{p}} = 2\pi\frac{\mathbf{c}_{\text{p}} \times \mathbf{a}_{\text{p}}}{V_{\text{c}}}, \;\; \mathbf{c}^*_{\text{p}} = 2\pi\frac{\mathbf{a}_{\text{p}} \times \mathbf{b}_{\text{p}}}{V_{\text{c}}}, \tag{11}$$

where $V_{\text{c}} = \mathbf{a}_{\text{p}} \cdot (\mathbf{b}_{\text{p}} \times \mathbf{c}_{\text{p}}) = \mathbf{b}_{\text{p}} \cdot (\mathbf{c}_{\text{p}} \times \mathbf{a}_{\text{p}}) = \mathbf{c}_{\text{p}} \cdot (\mathbf{a}_{\text{p}} \times \mathbf{b}_{\text{p}})$. The reciprocal lattice points are then obtained from the integer linear combinations of reciprocal basis vectors,

$$\mathbf{G} = G_1\mathbf{a}^*_{\text{p}} + G_2\mathbf{b}^*_{\text{p}} + G_3\mathbf{c}^*_{\text{p}} = (\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}})\boldsymbol{G} \;\; (G_i \in \mathbb{Z}). \tag{12}$$

Similarly to the atomic positions $\mathbf{R}$, wave vectors $\mathbf{q}$ are defined from real linear combinations of the reciprocal lattice vectors,

$$\mathbf{q} = q_1\mathbf{a}^*_{\text{p}} + q_2\mathbf{b}^*_{\text{p}} + q_3\mathbf{c}^*_{\text{p}} = (\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}})\boldsymbol{q}. \tag{13}$$

If the q_i are restricted to the interval $[0, 1)$, a primitive cell of reciprocal space is obtained.

If we consider a component of a Fourier expansion, $f_{\mathbf{q}}(\mathbf{R}) = \mathrm{e}^{i\mathbf{q}\cdot\mathbf{R}}$, it is transformed by an operation of the crystallographic point group to the function

$$\begin{aligned}\mathcal{S}f_{\mathbf{q}}(\mathbf{R}) &= f_{\mathbf{q}}(\mathcal{S}^{-1}\mathbf{R}) = \mathrm{e}^{i\mathbf{q}\cdot\mathcal{S}^{-1}\mathbf{R}} = \mathrm{e}^{i\mathcal{S}\mathbf{q}\cdot\mathcal{S}\mathcal{S}^{-1}\mathbf{R}} = \mathrm{e}^{i\mathcal{S}\mathbf{q}\cdot\mathbf{R}} \\ &= f_{\mathcal{S}\mathbf{q}}(\mathbf{R}).\end{aligned} \tag{14}$$

For this reason, the image of a wave vector through an operation of the crystallographic point group is defined to be

$$\mathbf{q}' = \mathcal{S}\mathbf{q} \Longleftrightarrow (\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}})\boldsymbol{q}' = \mathcal{S}(\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}})\boldsymbol{q} \tag{15}$$

or

$$\boldsymbol{q}' = (\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}})^{-1}\mathcal{S}(\mathbf{a}^*_{\text{p}}, \mathbf{b}^*_{\text{p}}, \mathbf{c}^*_{\text{p}})\boldsymbol{q} \tag{16}$$

$$= (\mathbf{a}_{\text{p}}, \mathbf{b}_{\text{p}}, \mathbf{c}_{\text{p}})^{\intercal}\mathcal{S}(\mathbf{a}_{\text{p}}, \mathbf{b}_{\text{p}}, \mathbf{c}_{\text{p}})^{-\intercal}\boldsymbol{q} \tag{17}$$

$$= \boldsymbol{S}^{-\intercal}\boldsymbol{q}. \tag{18}$$

The set $\{\boldsymbol{S}^{-\intercal}\boldsymbol{q}\}$ with $\boldsymbol{S}$ in the crystallographic point group $\{\boldsymbol{S}\}$ is called the star of $\mathbf{q}$.

2.5. Phonon coordinates

Atoms vibrate in the vicinity of their equilibrium positions in crystals. Instantaneous and equilibrium positions of atoms in crystals are denoted by $\mathbf{R}_{l\kappa}$ and $\mathbf{R}^0_{l\kappa}$, respectively, where l and κ are used to label the lattice points and atoms in the primitive cell of l, respectively. Displacements of atoms are written as $\mathbf{u}_{l\kappa} = \mathbf{R}_{l\kappa} - \mathbf{R}^0_{l\kappa}$.

In this review, we define the dynamical matrix as [7]

$$D_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) = \frac{1}{\sqrt{m_\kappa m_{\kappa'}}}\sum_{l'}\Phi_{0\kappa\alpha,l'\kappa'\alpha'}\mathrm{e}^{i\mathbf{q}\cdot(\mathbf{R}^0_{l'\kappa'} - \mathbf{R}^0_{0\kappa})}. \tag{19}$$

α and m_κ denote the index of the Cartesian coordinates and the mass of the atom κ, respectively. $\Phi_{l\kappa\alpha,l'\kappa'\alpha'}$ are the harmonic force constants, defined as the second derivative of the energy with respect to atomic positions, and evaluated at the equilibrium positions [7]. Phonon frequency $\omega_{\mathbf{q}\nu}$ and eigenvector $W_{\kappa\alpha}(\mathbf{q}\nu)$ are obtained as the solution of the eigenvalue equation of the dynamical matrix in equation (19), which is written as

$$\sum_{\kappa'\alpha'} D_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) W_{\kappa'\alpha'}(\mathbf{q}\nu) = \omega_{\mathbf{q}\nu}^2 W_{\kappa\alpha}(\mathbf{q}\nu). \tag{20}$$

ν labels the phonon band index, and the composite index $\mathbf{q}\nu$ is used to consider a phonon mode.

Equation (20) can also be written in matrix form as

$$\mathrm{D}(\mathbf{q})\mathrm{W}(\mathbf{q}) = \mathrm{W}(\mathbf{q})\Omega^2(\mathbf{q}), \tag{21}$$

or, because $\mathrm{D}(\mathbf{q})$ is hermitian,

$$\mathrm{D}(\mathbf{q}) = \mathrm{W}(\mathbf{q})\Omega^2(\mathbf{q})\mathrm{W}^\dagger(\mathbf{q}), \tag{22}$$

where $\Omega^2(\mathbf{q})$ is the diagonal matrix whose diagonal elements are $\omega_{\mathbf{q}\nu}^2$. Each column of $\mathrm{W}(\mathbf{q})$ contains an eigenvector $W_{\kappa\alpha}(\mathbf{q}\nu)$ corresponding to a different band index ν. Elements of each column are ordered as $(W_{\kappa_1 x}, W_{\kappa_1 y}, W_{\kappa_1 z}, \ldots, W_{\kappa_{n_\mathrm{a}} x}, W_{\kappa_{n_\mathrm{a}} y}, W_{\kappa_{n_\mathrm{a}} z})$, where n_a is the number of atoms in the primitive cell.

Because $D_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) = D^*_{\kappa\alpha,\kappa'\alpha'}(-\mathbf{q})$, we have

$$\omega_{\mathbf{q}\nu} = \omega_{-\mathbf{q}\nu}, \tag{23}$$

and we can choose

$$W_{\kappa\alpha}(\mathbf{q}\nu) = W^*_{\kappa\alpha}(-\mathbf{q}\nu). \tag{24}$$

Moreover eigenvalues and eigenvectors inherit symmetry properties of the force constants. It can be shown [8] that under a space group operation $(\mathcal{S}, \boldsymbol{\tau})$ the eigenvalues and eigenvectors transform as

$$\omega_{\mathcal{S}\mathbf{q}\nu} = \omega_{\mathbf{q}\nu}, \tag{25}$$

$$\mathbf{W}_{\kappa'}(\mathcal{S}\mathbf{q}\nu) = \mathcal{S}\mathbf{W}_{\kappa}(\mathbf{q}\nu)\mathrm{e}^{-i\mathcal{S}\mathbf{q}\cdot\boldsymbol{\tau}}, \tag{26}$$

where $\mathbf{W}_\kappa(\mathbf{q}\nu)$ represents the phonon eigenvector of atom κ in Cartesian coordinates, and κ' is the image of atom κ under the space group operation.

An actual displacement can be described from a linear combination of phonon eigenvectors. This defines the phonon coordinates $Q(\mathbf{q}\nu)$ as

$$u_{l\kappa\alpha} = \sum_{\mathbf{q}\nu} Q(\mathbf{q}\nu) \frac{1}{\sqrt{N m_\kappa}} W_{\kappa\alpha}(\mathbf{q}\nu)\mathrm{e}^{i\mathbf{q}\cdot\mathbf{R}^0_{l\kappa}} \tag{27}$$

$$= \sum_{\mathbf{q}\nu} Q(\mathbf{q}\nu) u_{l\kappa\alpha}(\mathbf{q}\nu), \tag{28}$$

where N is the number of lattice points in crystal. $u_{l\kappa\alpha}(\mathbf{q}\nu)$ is defined for the purpose of convenience in the following sections. The previous equation can be inverted to give

$$Q(\mathbf{q}\nu) = \sum_{l\kappa\alpha} u_{l\kappa\alpha} \sqrt{\frac{m_\kappa}{N}} W^*_{\kappa\alpha}(\mathbf{q}\nu)\mathrm{e}^{-i\mathbf{q}\cdot\mathbf{R}^0_{l\kappa}}. \tag{29}$$

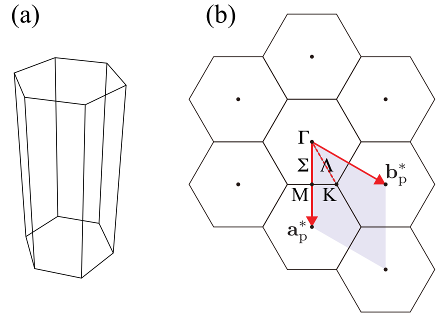

Figure 4. Brillouin zone (BZ) of β-$\mathrm{Si_3N_4}$. See also figures 1 and 2 for the cell shape in direct space. (a) The first BZ. (b) BZs viewed from $\mathbf{c}_\mathrm{p}^*$ axis direction. Symbols of the special points and paths follow the Bilbao crystallographic server [6].

2.6. BZ

Symmetry property of phonons in reciprocal space is best represented in the BZ [6, 8–12]. In the phonopy and phono3py codes, the BZ is defined as a Wigner–Seitz cell of the reciprocal lattice. To check if a $\mathbf{q}$ point belongs to the BZ we proceed as follows. First the basis vectors of the Niggli cell [13–16] are determined. The Niggli cell is a cell with the shortest possible reciprocal basis vectors. Then using integer linear combinations of those Niggli basis vectors, we search for the shortest $|\mathbf{q}+\mathbf{G}|$. Finally $\mathbf{q}+\mathbf{G}$ is used as the $\mathbf{q}$ point in the BZ.

Three $\mathbf{q}$ points are needed for the three-phonon scattering considered in the phono3py code. In case one or more of those three points is on the BZ surface, we choose the translationally equivalent points on the BZ surface which minimize $|\mathbf{q}+\mathbf{q}'+\mathbf{q}''|$.

As an example, the BZ of β-$\mathrm{Si_3N_4}$ is presented in figure 4. The basal plane has the hexagonal shape and $\mathbf{c}_\mathrm{p}^*$ is longer than $\mathbf{a}_\mathrm{p}^*$ and $\mathbf{b}_\mathrm{p}^*$ because $\mathbf{c}_\mathrm{p}$ is shorter than $\mathbf{a}_\mathrm{p}$ and $\mathbf{b}_\mathrm{p}$. By definition, $\mathbf{a}_\mathrm{p}^*/2$, $\mathbf{b}_\mathrm{p}^*/2$, and $\mathbf{c}_\mathrm{p}^*/2$ are located on the BZ surface. The high symmetry points and paths in the BZ have special symbols as shown in figure 4(b) [6].

3. Geometry of supercell model

3.1. Supercell construction

In the phonopy and phono3py codes, the supercell approach is employed. The supercell is defined by multiple primitive cells, so that the basis vectors of the supercell $(\mathbf{a}_\mathrm{s}, \mathbf{b}_\mathrm{s}, \mathbf{c}_\mathrm{s})$ can be represented as the image of the primitive cell basis vectors through an integer matrix $\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}$,

$$(\mathbf{a}_\mathrm{s}, \mathbf{b}_\mathrm{s}, \mathbf{c}_\mathrm{s}) = (\mathbf{a}_\mathrm{p}, \mathbf{b}_\mathrm{p}, \mathbf{c}_\mathrm{p})\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}. \tag{30}$$

Like the unit cell, the supercells are arranged on the supercell lattice. The supercell lattice points are labeled by L, and located from the vectors $\mathbf{R}_L$.

The supercell model is constructed using $\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}$. The lattice points within the supercell are used to perform the summation found in equation (19). These lattice points can be elegantly obtained using the approach reported by Hart and Forcade

[17]. The integer matrix $\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}$ is first reduced to a diagonal integer matrix $\boldsymbol{D} = \boldsymbol{P}\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}\boldsymbol{Q}$ by the unimodular matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, where we choose $\det(\boldsymbol{P}) = 1$ and $\det(\boldsymbol{Q}) = 1$. This matrix decomposition can be the Smith normal form (SNF), however, it is unnecessary to strictly follow the definition of the SNF.

Using this transformation, equation (30) can be written

$$(\tilde{\mathbf{a}}_{\mathrm{s}}, \tilde{\mathbf{b}}_{\mathrm{s}}, \tilde{\mathbf{c}}_{\mathrm{s}}) = (\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}}, \tilde{\mathbf{c}}_{\mathrm{p}})\boldsymbol{D}, \tag{31}$$

with

$$(\tilde{\mathbf{a}}_{\mathrm{s}}, \tilde{\mathbf{b}}_{\mathrm{s}}, \tilde{\mathbf{c}}_{\mathrm{s}}) = (\mathbf{a}_{\mathrm{s}}, \mathbf{b}_{\mathrm{s}}, \mathbf{c}_{\mathrm{s}})\boldsymbol{Q}, \tag{32}$$

$$(\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}}, \tilde{\mathbf{c}}_{\mathrm{p}}) = (\mathbf{a}_{\mathrm{p}}, \mathbf{b}_{\mathrm{p}}, \mathbf{c}_{\mathrm{p}})\boldsymbol{P}^{-1}. \tag{33}$$

$(\tilde{\mathbf{a}}_{\mathrm{s}}, \tilde{\mathbf{b}}_{\mathrm{s}}, \tilde{\mathbf{c}}_{\mathrm{s}})$ and $(\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}}, \tilde{\mathbf{c}}_{\mathrm{p}})$ define a new supercell and a new primitive cell, respectively. Because $\det(\boldsymbol{P}) = \det(\boldsymbol{Q}) = 1$, they have the same volumes as the former ones and generate the same lattices. The benefit of using those new primitive and supercell comes from the fact that $\boldsymbol{D}$ is diagonal. The new primitive and supercell lattice vectors are therefore collinear.

We can easily find the lattice vectors located within the new supercell. If we write $\boldsymbol{D} = \mathrm{diag}(n_1, n_2, n_3)$, they are given by the vectors

$$\mathbf{R}_m = (\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}}, \tilde{\mathbf{c}}_{\mathrm{p}}) \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = (\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}}, \tilde{\mathbf{c}}_{\mathrm{p}})\boldsymbol{m}, \tag{34}$$

with $m_i \in \{0, 1, \ldots, n_i - 1\}$. This can also be written

$$\mathbf{R}_m = (\tilde{\mathbf{a}}_{\mathrm{s}}, \tilde{\mathbf{b}}_{\mathrm{s}}, \tilde{\mathbf{c}}_{\mathrm{s}})\boldsymbol{D}^{-1}\boldsymbol{m} \tag{35}$$

$$= (\mathbf{a}_{\mathrm{s}}, \mathbf{b}_{\mathrm{s}}, \mathbf{c}_{\mathrm{s}})\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{m}. \tag{36}$$

Therefore, the lattice points within the original supercell have coordinates

$$\boldsymbol{x}_{\mathrm{s}} = \boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{m} \pmod 1 \tag{37}$$

along the supercell lattice vectors $(\mathbf{a}_{\mathrm{s}}, \mathbf{b}_{\mathrm{s}}, \mathbf{c}_{\mathrm{s}})$. The $\pmod 1$ operation is used to shift the lattice vectors from within the new supercell to within the original supercell.

In figure 5, an example of a supercell construction for a two-dimensional lattice is presented. $\boldsymbol{D} = \boldsymbol{P}\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}\boldsymbol{Q}$ is computed as

$$\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}. \tag{38}$$

This gives

$$(\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}}) = (\mathbf{a}_{\mathrm{p}}, \mathbf{b}_{\mathrm{p}})\boldsymbol{P}^{-1} = (\mathbf{a}_{\mathrm{p}} + \mathbf{b}_{\mathrm{p}}, -\mathbf{a}_{\mathrm{p}}), \tag{39}$$

$$(\tilde{\mathbf{a}}_{\mathrm{s}}, \tilde{\mathbf{b}}_{\mathrm{s}}) = (\mathbf{a}_{\mathrm{s}}, \mathbf{b}_{\mathrm{s}})\boldsymbol{Q} = (\mathbf{b}_{\mathrm{s}}, -\mathbf{a}_{\mathrm{s}} - \mathbf{b}_{\mathrm{s}}). \tag{40}$$

As can be seen in figure 5, $(\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}})$ and $(\tilde{\mathbf{a}}_{\mathrm{s}}, \tilde{\mathbf{b}}_{\mathrm{s}})$ are mutually parallel, and the new supercell is simply built by 2 × 4 units of the new primitive cell. Since the lattices generated by the original or new basis vectors are the same, the lattice points inside the new supercell are simply brought into the original supercell by supercell lattice translations.

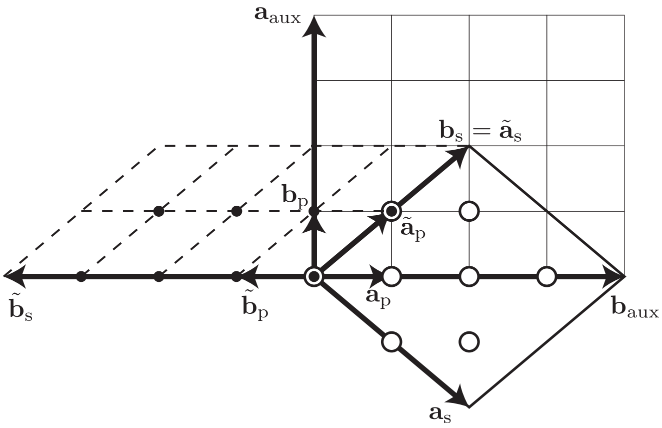
$\mathbf{a}_{\mathrm{aux}}$
$\mathbf{b}_{\mathrm{s}} = \tilde{\mathbf{a}}_{\mathrm{s}}$
$\mathbf{b}_{\mathrm{p}}$
$\tilde{\mathbf{a}}_{\mathrm{p}}$
$\tilde{\mathbf{b}}_{\mathrm{s}}$
$\tilde{\mathbf{b}}_{\mathrm{p}}$
$\mathbf{a}_{\mathrm{p}}$
$\mathbf{b}_{\mathrm{aux}}$
$\mathbf{a}_{\mathrm{s}}$

Figure 5. Lattices and basis vectors of primitive cell $(\mathbf{a}_{\mathrm{p}}, \mathbf{b}_{\mathrm{p}})$, supercell $(\mathbf{a}_{\mathrm{s}}, \mathbf{b}_{\mathrm{s}})$, new primitive cell $(\tilde{\mathbf{a}}_{\mathrm{p}}, \tilde{\mathbf{b}}_{\mathrm{p}})$, new supercell $(\tilde{\mathbf{a}}_{\mathrm{s}}, \tilde{\mathbf{b}}_{\mathrm{s}})$, and auxiliary supercell $(\mathbf{a}_{\mathrm{aux}}, \mathbf{b}_{\mathrm{aux}})$. Circle and small filled circle symbols depict lattice points in the supercell and in the new supercell, respectively. The latter points can be brought by the supercell lattice translations to the former points uniquely.

Another way of constructing the supercell is to employ an auxiliary supercell,

$$(\mathbf{a}_{\mathrm{aux}}, \mathbf{b}_{\mathrm{aux}}, \mathbf{c}_{\mathrm{aux}}) = (\mathbf{a}_{\mathrm{p}}, \mathbf{b}_{\mathrm{p}}, \mathbf{c}_{\mathrm{p}})\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}}, \tag{41}$$

which is defined by a diagonal integer matrix $\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}} = \mathrm{diag}(n_1, n_2, n_3)$. This gives

$$(\mathbf{a}_{\mathrm{aux}}, \mathbf{b}_{\mathrm{aux}}, \mathbf{c}_{\mathrm{aux}}) = (\mathbf{a}_{\mathrm{s}}, \mathbf{b}_{\mathrm{s}}, \mathbf{c}_{\mathrm{s}})\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}^{-1}\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}}. \tag{42}$$

We choose $\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}^{-1}\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}}$ to be an integer matrix. To satisfy it, in the implementation, the diagonal elements of $\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}}$ are determined by making the smallest parallelepiped of $(\mathbf{a}_{\mathrm{aux}}, \mathbf{b}_{\mathrm{aux}}, \mathbf{c}_{\mathrm{aux}})$ that includes the parallelepiped of $(\mathbf{a}_{\mathrm{s}}, \mathbf{b}_{\mathrm{s}}, \mathbf{c}_{\mathrm{s}})$. The lattice points within this auxiliary supercell are then

$$\mathbf{R}_m = (\mathbf{a}_{\mathrm{p}}, \mathbf{b}_{\mathrm{p}}, \mathbf{c}_{\mathrm{p}}) \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = (\mathbf{a}_{\mathrm{p}}, \mathbf{b}_{\mathrm{p}}, \mathbf{c}_{\mathrm{p}})\boldsymbol{m}, \tag{43}$$

with $m_i \in \{0, 1, \ldots, n_i - 1\}$. Their coordinates along the basis vectors of the auxiliary supercell are

$$\boldsymbol{x}_{\mathrm{aux}} = \boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}}^{-1}\boldsymbol{m} \tag{44}$$

while their coordinates along the basis vectors of the original supercell are

$$\boldsymbol{x}_{\mathrm{s}} = \boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}^{-1}\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}}\boldsymbol{x}_{\mathrm{aux}} \pmod 1. \tag{45}$$

As before, we used the $\pmod 1$ operation to shift the lattice points from within the auxiliary supercell to within the original supercell. Notice that every lattice point within the original supercell is obtained from $\det(\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}^{-1}\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}})$ lattice points within the auxiliary supercell. Only one of them should be conserved.

An example is shown in figure 5. Lattice points in the auxiliary supercell are brought into the supercell by the supercell lattice translations. But since $\det(\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}^{-1}\boldsymbol{M}_{\mathrm{p}\to\mathrm{aux}}) = 2$, each lattice point within the original supercell is obtained from two lattice points in the auxiliary supercell.

For backward compatibility, to preserve indices of atoms in the supercell, it is the latter way of the supercell construction which is used as default in the phonopy and phono3py codes. However, the latter algorithm is much slower than the former, and therefore the former should be used for very larger supercells.

3.2. Commensurate points

From equations (30) and (10), we have

$$(\mathbf{a}_{\mathrm{p}}^{*}, \mathbf{b}_{\mathrm{p}}^{*}, \mathbf{c}_{\mathrm{p}}^{*}) = (\mathbf{a}_{\mathrm{s}}^{*}, \mathbf{b}_{\mathrm{s}}^{*}, \mathbf{c}_{\mathrm{s}}^{*})\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}^{\intercal}. \tag{46}$$

$(\mathbf{a}_{\mathrm{s}}^{*}, \mathbf{b}_{\mathrm{s}}^{*}, \mathbf{c}_{\mathrm{s}}^{*})$ are the reciprocal basis vectors associated with the supercell. $\mathbf{q}$ points given by integer linear combinations of the reciprocal supercell basis vectors are called commensurate with the supercell because they fulfil $\mathrm{e}^{i\mathbf{q}\cdot\mathbf{R}_L} = 1$ for any supercell lattice vector $\mathbf{R}_L$.

For practical purposes, we are interested in the commensurate points located within the reciprocal primitive cell. Since equation (46) has the same form as equation (30), the same approaches as those explained in section 3.1 can be used for the generation of the commensurate points within the reciprocal primitive cell.

4. Transformation between force constants and dynamical matrices

In equations (19) and (20), force constants are transformed to phonon eigenvalues and eigenvectors. In principle, in equation (19) the summation over l' is performed for all lattice vectors. However, in the supercell approach, it is not $\Phi_{0\kappa\alpha,l'\kappa'\alpha'}$ which is obtained from numerical simulations, but $\Phi^{\mathrm{SC}}_{0\kappa\alpha,l'\kappa'\alpha'} = \sum_{L'} \Phi_{0\kappa\alpha,L'+l'\kappa'\alpha'}$, [18] with l' restricted to the supercell. It means that the force constants are only known within a restricted region of the crystal, the supercell, and within this region, they are contaminated by the periodic repetitions of the supercell. This also means that the supercell force constants $\Phi^{\mathrm{SC}}_{0\kappa\alpha,l'\kappa'\alpha'}$ are periodic under a supercell lattice translation while the phase factor $\mathrm{e}^{i\mathbf{q}\cdot(\mathbf{R}^0_{l'\kappa'}-\mathbf{R}^0_{0\kappa})}$ is not, except when $\mathbf{q}$ is a wave vector commensurate with the supercell. There is therefore an ambiguity about where to choose the atom $(l'\kappa')$. For commensurate wave vectors, it could be chosen in any periodic repetition of the supercell containing atom (0κ) without changing the value of the phase factor. However, for other values of $\mathbf{q}$, this value would be changed, and therefore this choice matters.

In the phonopy and phono3py codes, following [18], the phase factors given by the shortest vectors of $\mathbf{R}^0_{l'\kappa'} + \mathbf{R}_L - \mathbf{R}^0_{0\kappa}$ are chosen, where $\mathbf{R}_L$ are the positions of the supercell lattice points L. This is implemented by searching L by

$$\{L\}_{\kappa\kappa'l'} = \underset{L}{\arg\min}(|\mathbf{R}^0_{l'\kappa'} + \mathbf{R}_L - \mathbf{R}^0_{0\kappa}|), \tag{47}$$

and the results are stored and used many times in the calculation. Since multiple L can be found for each pair of atoms (0κ) in the primitive cell and $(l'\kappa')$ in the supercell, equation (47) gives a set of L. The dynamical matrix of the supercell is computed from the supercell force constants $\Phi^{\mathrm{SC}}_{0\kappa\alpha,l'\kappa'\alpha'}$ averaging the phase factors of $\{L\}$ as follows, [18]

$$\begin{aligned} D^{\mathrm{SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) = \frac{1}{\sqrt{m_\kappa m_{\kappa'}}} \sum_{l'} \Phi^{\mathrm{SC}}_{0\kappa\alpha,l'\kappa'\alpha'} \\ \times \frac{1}{|\{L\}_{\kappa\kappa'l'}|} \sum_{\{L\}_{\kappa\kappa'l'}} \mathrm{e}^{i\mathbf{q}\cdot(\mathbf{R}^0_{l'\kappa'}+\mathbf{R}_L-\mathbf{R}^0_{0\kappa})}. \end{aligned} \tag{48}$$

Equation (19) can be inverted to compute force constants from N dynamical matrices, [19]

$$\Phi_{0\kappa\alpha,l'\kappa'\alpha'} = \frac{\sqrt{m_\kappa m_{\kappa'}}}{N} \sum_{\mathbf{q}} D_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q})\mathrm{e}^{-i\mathbf{q}\cdot(\mathbf{R}^0_{l'\kappa'}-\mathbf{R}^0_{0\kappa})}. \tag{49}$$

$D_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q})$ can be obtained from known eigenvectors and eigenvalues, as in equation (22), or calculated by equation (48). $\Phi^{\mathrm{SC}}_{0\kappa\alpha,l'\kappa'\alpha'}$ can be obtained from $D^{\mathrm{SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q})$ by restricting the above $\mathbf{q}$ sum to the wavevectors $\mathbf{q}^\star$ commensurate with the supercell. Equation (49) can also be used to transform supercell force constants in a different supercell shape by oversampling $\mathbf{q}^\star$ points. Indeed, we obtain

$$\Phi^{\mathrm{SC}_{\mathrm{L}\leftarrow\mathrm{S}}}_{0\kappa\alpha,l'_{\mathrm{L}}\kappa'\alpha'} = \frac{1}{N} \sum_{\mathbf{q}^\star_{\mathrm{L}}} \mathrm{e}^{-i\mathbf{q}^\star_{\mathrm{L}}\cdot\left(\mathbf{R}^0_{l'_{\mathrm{L}}\kappa'}-\mathbf{R}^0_{0\kappa}\right)} \left[\sum_{l'_{\mathrm{S}}} \Phi^{\mathrm{SC}_{\mathrm{S}}}_{0\kappa\alpha,l'_{\mathrm{S}}\kappa'\alpha'} \frac{1}{|\{L_{\mathrm{S}}\}_{\kappa\kappa'l'_{\mathrm{S}}}|} \sum_{\{L_{\mathrm{S}}\}_{\kappa\kappa'l'_{\mathrm{S}}}} \mathrm{e}^{i\mathbf{q}^\star_{\mathrm{L}}\cdot\left(\mathbf{R}^0_{l'_{\mathrm{S}}\kappa'}+\mathbf{R}_{L_{\mathrm{S}}}-\mathbf{R}^0_{0\kappa}\right)}\right], \tag{50}$$

where subscripts of variables X_{L} and X_{S} mean those defined in different supercell shapes, e.g. larger (L) and smaller (S) supercells, respectively. The original supercell force constants $\Phi^{\mathrm{SC}_{\mathrm{S}}}_{0\kappa\alpha,l'_{\mathrm{S}}\kappa'\alpha'}$ are given in the smaller supercell. Commensurate points $\mathbf{q}^\star_{\mathrm{L}}$ are sampled with respect to the larger supercell. The transformed supercell force constants $\Phi^{\mathrm{SC}_{\mathrm{L}\leftarrow\mathrm{S}}}_{0\kappa\alpha,l'_{\mathrm{L}}\kappa'\alpha'}$ are given in the larger supercell. A possible application is to embed anharmonic contribution obtained through self-consistent harmonic approximation using a smaller supercell into the harmonic force constants of a larger supercell [20].

5. Non-analytical term correction

Long-range dipole–dipole interactions are difficult to capture in a supercell approach. Therefore it is treated with the help of a model, the so-called non-analytical term correction (NAC) [1, 21–23].

At the commensurate points, this contribution is already included in equation (48) via the supercell force constants. However, at general $\mathbf{q}$ points it is not. Gonze and Lee [1] have formulated this contribution to the dynamical matrix, and in the phonopy and phono3py codes only the reciprocal space term of the dipole–dipole interaction contribution to the dynamical matrix is calculated. It is given by

$$\begin{aligned} D^{\mathrm{DD}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) = {} & \frac{4\pi}{V_{\mathrm{c}}}\frac{e^2}{4\pi\epsilon_0}\frac{1}{\sqrt{m_\kappa m_{\kappa'}}}\sum_{\beta\beta'} Z^*_{\kappa,\beta\alpha}Z^*_{\kappa',\beta'\alpha'} \\ & \times \sum_{\mathbf{Q}=\mathbf{G}+\mathbf{q}} \frac{Q_\beta Q_{\beta'}}{\sum_{\gamma\gamma'} Q_\gamma \epsilon_{\gamma\gamma'} Q_{\gamma'}} \\ & \times \mathrm{e}^{i\mathbf{G}\cdot(\mathbf{R}^0_{0\kappa}-\mathbf{R}^0_{0\kappa'})} \exp\left(-\sum_{\gamma\gamma'}\frac{Q_\gamma\epsilon_{\gamma\gamma'}Q_{\gamma'}}{4\Lambda^2}\right), \end{aligned} \tag{51}$$

where β and γ indicate the Cartesian coordinates. If the polarization is called $\mathbf{P}$, then $Z^*_{\kappa,\beta\alpha} = \frac{V_{\mathrm{c}}}{e}\frac{\partial \mathbf{P}_\beta}{\partial \mathbf{R}_{0\kappa\alpha}}|_{\mathbf{E}=0}$ are the Born effective charges at zero electric field, $\mathbf{E}=0$. $\epsilon_{\gamma\gamma'}$ is the high-frequency static dielectric constant tensor. Λ is a parameter adjusted together with the cutoff radius used for the summation over $\mathbf{G}$. A translational invariance condition can be applied to equation (51) by [1]

$$D^{\mathrm{DD}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) \leftarrow D^{\mathrm{DD}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) - \delta_{\kappa\kappa'}\sum_{\kappa''}\sqrt{\frac{m_{\kappa''}}{m_\kappa}} D^{\mathrm{DD}}_{\kappa\alpha,\kappa''\alpha'}(\mathbf{q}=0). \tag{52}$$

As said already, at the commensurate points, this dipole–dipole contribution is already included in equation (48) via the supercell force constants. Therefore, equation (51) is used for the interpolation of the dynamical matrix at general $\mathbf{q}$ points. The procedure is reported in [1], which is described shortly as follows. At the commensurate points $\mathbf{q}^\star$, the short-range dynamical matrix is calculated as

$$D^{\mathrm{SR,SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}^\star) = D^{\mathrm{SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}^\star) - D^{\mathrm{DD}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}^\star). \tag{53}$$

Next, the short-range supercell force constants $\Phi^{\mathrm{SR,SC}}_{0\kappa\alpha,l'\kappa'\alpha'}$ are obtained from $\{D^{\mathrm{SR,SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}^\star)\}$ using equation (49), then the short-range dynamical matrix $D^{\mathrm{SR,SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q})$ is calculated at the general $\mathbf{q}$ point by equation (48), i.e.

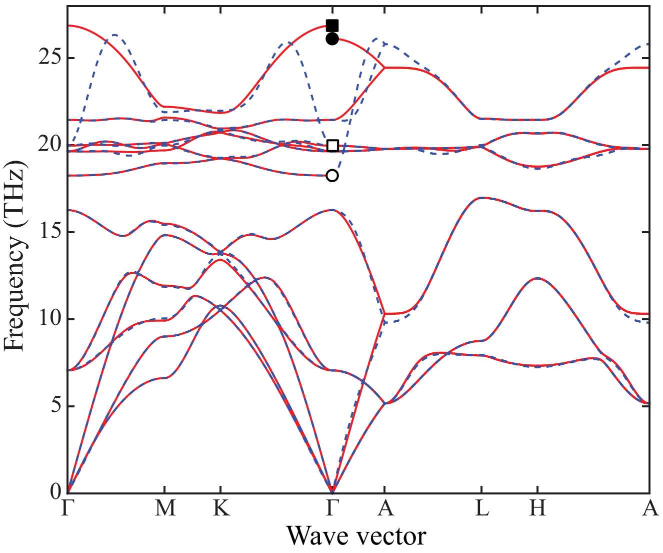

Figure 6. Calculated phonon band structure of wurtzite-type AlN using the 5 × 5 × 3 supercell with (solid curves) and without (dashed curves) non-analytical term correction (NAC). Circle and square symbols depict phonon frequencies at $\mathbf{q}\to\mathbf{0}$, where filled and open symbols indicate phonon modes calculated with and without NAC, respectively. Wave vector path was selected by Seek-path [24, 25].

$$\left\{D^{\mathrm{SR,SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}^\star)\right\} \xrightarrow[\text{Eq. (49)}]{} \Phi^{\mathrm{SR,SC}}_{0\kappa\alpha,l'\kappa'\alpha'} \xrightarrow[\text{Eq. (48)}]{} D^{\mathrm{SR,SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}). \tag{54}$$

Finally, the dynamical matrix at the general $\mathbf{q}$ point with NAC is obtained by

$$D^{\mathrm{NAC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) = D^{\mathrm{SR,SC}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}) + D^{\mathrm{DD}}_{\kappa\alpha,\kappa'\alpha'}(\mathbf{q}). \tag{55}$$

An example of the application of NAC to wurtzite-type AlN is shown in figure 6. As can be seen, the correction is significant near the Γ point. It also shows the directional dependence near the Γ, in the direction of K and A, due to the non-spherical symmetry of $\epsilon_{\gamma\gamma'}$ and $Z^*_{\kappa,\beta\alpha}$ of AlN.

6. Regular grid in reciprocal primitive cell

When phonon properties at $\mathbf{q}$ points need to be integrated over the BZ, a technique often used is to discretize the reciprocal space. Usually, reciprocal primitive cells are uniformly sampled by a regular grid such as [26], even if other choices are possible [27]. Traditionally, the regular grid is defined by evenly dividing the reciprocal basis vectors, as shown in figure 7(a). However, other regular grids may be chosen, as shown in figure 7(b). For example, the grid shown in figure 7(b) is defined by dividing evenly the reciprocal basis vectors of the conventional unit cell instead. This is a type of generalized regular grid, [28–31] as will be explained later.

6.1. Traditional regular grid

Traditionally, the volume of the reciprocal primitive cell is divided into uniform microzones so that the basis vectors of

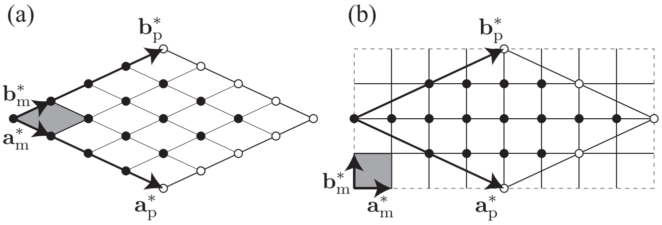
(a)
$\mathbf{b}_\mathrm{p}^*$
$\mathbf{b}_\mathrm{m}^*$
$\mathbf{a}_\mathrm{m}^*$
$\mathbf{a}_\mathrm{p}^*$
(b)
$\mathbf{b}_\mathrm{p}^*$
$\mathbf{b}_\mathrm{m}^*$
$\mathbf{a}_\mathrm{m}^*$
$\mathbf{a}_\mathrm{p}^*$

Figure 7. Different Γ-centred regular grids in plane reciprocal primitive cell (*c*2*mm*). Filled circle symbols depict the grid points. Open circle symbols show the grid points equivalent to some other grid points by periodicity. Shaded area indicates a microzone [2] and $\mathbf{a}_\mathrm{m}^*$ and $\mathbf{b}_\mathrm{m}^*$ give the basis vectors of the microzone. (a) Traditional regular grid. (b) A type of generalized regular grid [28–31].

the reciprocal primitive cell are simply integer multiples of the basis vectors of each microzone $(\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)$. The equation which defines the microzone basis vectors is then

$$(\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*) = (\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)\boldsymbol{D} \tag{56}$$

with $\boldsymbol{D} = \mathrm{diag}(n_1, n_2, n_3)$, $n_i \in \mathbb{N}$.

A 2D example of this microzone is shown in figure 7(a). The $\mathbf{q}$ points of the grid points are represented by integer linear combinations of the microzone basis vectors plus a possible rigid shift $(s_1, s_2, s_3)^\intercal$. Therefore their coordinates in the $(\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*)$ basis are

$$\begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} (m_1 + s_1)/n_1 \\ (m_2 + s_2)/n_2 \\ (m_3 + s_3)/n_3 \end{pmatrix}, \tag{57}$$
$$m_i \in \{0, 1, \ldots, n_i - 1\},\ 0 \leqslant s_i < 1.$$

To conserve symmetry, n_i and s_i are chosen so that the microzone lattice with the shift is invariant under the crystallographic point group $\mathbb{P}$.

6.2. Generalized regular grid

A generalized regular grid is defined using a conventional unit cell related to the primitive cell by equation (8). The reciprocal conventional unit cell is therefore

$$(\mathbf{a}_\mathrm{c}^*, \mathbf{b}_\mathrm{c}^*, \mathbf{c}_\mathrm{c}^*) = (\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*)\boldsymbol{P}_\mathrm{prim}^\intercal. \tag{58}$$

Microzones can be defined considering that their basis vectors are integer divisions of the reciprocal basis vectors of the conventional unit cell. This is written as

$$(\mathbf{a}_\mathrm{c}^*, \mathbf{b}_\mathrm{c}^*, \mathbf{c}_\mathrm{c}^*) = (\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)\mathrm{diag}(N_1, N_2, N_3),\ N_i \in \mathbb{N}. \tag{59}$$

A 2D example of this microzone is shown in figure 7(b). From equations (58) and (59), we have

$$\begin{aligned}(\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*) &= (\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)\mathrm{diag}(N_1, N_2, N_3)\boldsymbol{P}_\mathrm{prim}^{-\intercal} \\ &= (\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)\boldsymbol{M}_\mathrm{g}. \end{aligned} \tag{60}$$

The grid matrix $\boldsymbol{M}_\mathrm{g} = \mathrm{diag}(N_1, N_2, N_3)\boldsymbol{P}_\mathrm{prim}^{-\intercal}$ is an integer matrix. $\det(\boldsymbol{M}_\mathrm{g})$ is the number of translationally nonequivalent grid points in the reciprocal primitive cell. When $\boldsymbol{M}_\mathrm{g}$ is not a diagonal matrix, the grid generated by equation (60) is a generalized regular grid, otherwise we obtain the traditional regular grid as described in section 6.1. Notice however that the generalized regular grid may be transformed into the traditional regular grid of some reciprocal cell using an SNF kind of decomposition, as explained in section 3.1. This is shown in the next section.

6.3. Indexing of grid points

The integer matrix $\boldsymbol{M}_\mathrm{g}$ can be reduced to a diagonal integer matrix such as $\boldsymbol{D} = \boldsymbol{P}\boldsymbol{M}_\mathrm{g}\boldsymbol{Q}$ as has been employed in section 3.1. This property of the matrix decomposition is similarly used to index grid points by integer numbers of $\{0, 1, \ldots, \det(\boldsymbol{D}) - 1\}$ [17] in the phono3py code. Note that $\det(\boldsymbol{D}) = \det(\boldsymbol{M}_\mathrm{g})$. Equation (60) is rewritten as

$$(\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*)\boldsymbol{Q} = (\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)\boldsymbol{P}^{-1}\boldsymbol{D}. \tag{61}$$

Denoting

$$(\tilde{\mathbf{a}}_\mathrm{p}^*, \tilde{\mathbf{b}}_\mathrm{p}^*, \tilde{\mathbf{c}}_\mathrm{p}^*) = (\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*)\boldsymbol{Q}, \tag{62}$$

$$(\tilde{\mathbf{a}}_\mathrm{m}^*, \tilde{\mathbf{b}}_\mathrm{m}^*, \tilde{\mathbf{c}}_\mathrm{m}^*) = (\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)\boldsymbol{P}^{-1}, \tag{63}$$

we have

$$(\tilde{\mathbf{a}}_\mathrm{p}^*, \tilde{\mathbf{b}}_\mathrm{p}^*, \tilde{\mathbf{c}}_\mathrm{p}^*) = (\tilde{\mathbf{a}}_\mathrm{m}^*, \tilde{\mathbf{b}}_\mathrm{m}^*, \tilde{\mathbf{c}}_\mathrm{m}^*)\boldsymbol{D}. \tag{64}$$

Since $\boldsymbol{Q}$ is a unimodular matrix, $(\tilde{\mathbf{a}}_\mathrm{p}^*, \tilde{\mathbf{b}}_\mathrm{p}^*, \tilde{\mathbf{c}}_\mathrm{p}^*)$ and $(\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*)$ generate the same reciprocal primitive lattice. Similarly $(\tilde{\mathbf{a}}_\mathrm{m}^*, \tilde{\mathbf{b}}_\mathrm{m}^*, \tilde{\mathbf{c}}_\mathrm{m}^*)$ and $(\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)$ generate the same microzone lattice due to the unimodular matrix $\boldsymbol{P}^{-1}$.

Equation (64) has the same form as equation (56) with $\boldsymbol{D} = \mathrm{diag}(n_1, n_2, n_3)$. Therefore the $\mathbf{q}$ points of the grid points are calculated similarly as equation (57) but in the basis $(\tilde{\mathbf{a}}_\mathrm{p}^*, \tilde{\mathbf{b}}_\mathrm{p}^*, \tilde{\mathbf{c}}_\mathrm{p}^*)$,

$$\mathbf{q} = (\tilde{\mathbf{a}}_\mathrm{p}^*, \tilde{\mathbf{b}}_\mathrm{p}^*, \tilde{\mathbf{c}}_\mathrm{p}^*)\begin{pmatrix} (m_1 + \tilde{s}_1)/n_1 \\ (m_2 + \tilde{s}_2)/n_2 \\ (m_3 + \tilde{s}_3)/n_3 \end{pmatrix}. \tag{65}$$

Notice that in the above equation we use $\tilde{s}_i$ rather than s_i. In fact for practical purposes, it may be convenient to define the grid shift $\boldsymbol{s} = (s_1, s_2, s_3)^\intercal$ in the basis $(\mathbf{a}_\mathrm{m}^*, \mathbf{b}_\mathrm{m}^*, \mathbf{c}_\mathrm{m}^*)$. Therefore we have $\tilde{\boldsymbol{s}} = \boldsymbol{P}\boldsymbol{s}$.

The indexing of the grid points in the reciprocal primitive cell $(\tilde{\mathbf{a}}_\mathrm{p}^*, \tilde{\mathbf{b}}_\mathrm{p}^*, \tilde{\mathbf{c}}_\mathrm{p}^*)$ is a trivial task. Indeed, in the phonopy and phono3py codes, each grid point is bijectively mapped to an integer p by

$$p = m_1 + n_1 m_2 + n_1 n_2 m_3,\ m_i \in \{0, 1, \ldots, n_i - 1\}. \tag{66}$$

With this definition, $0 \leqslant p < n_1 n_2 n_3$.

The $\mathbf{q}$ points generated this way are located within the reciprocal cell $(\tilde{\mathbf{a}}_\mathrm{p}^*, \tilde{\mathbf{b}}_\mathrm{p}^*, \tilde{\mathbf{c}}_\mathrm{p}^*)$. They could be shifted to the reciprocal cell $(\mathbf{a}_\mathrm{p}^*, \mathbf{b}_\mathrm{p}^*, \mathbf{c}_\mathrm{p}^*)$ by reciprocal lattice vector translations if needed. Notice however that to obtain the integer p through equation (66) for a general integer triplet $(m_1,m_2,m_3)^\mathsf{T}$, modulo $\boldsymbol{n}=(n_1,n_2,n_3)^\mathsf{T}$ is required to locate the point within the cell $(\tilde{\mathbf{a}}^*_\mathrm{p},\tilde{\mathbf{b}}^*_\mathrm{p},\tilde{\mathbf{c}}^*_\mathrm{p})$. Indeed, $(m_1,m_2,m_3)^\mathsf{T}$ and $(m_1+G_1n_1,m_2+G_2n_2,m_3+G_3n_3)^\mathsf{T}$ indicate different locations in $\mathbf{q}$ space, although they are equivalent points due to periodicity.

6.4. Symmetry of generalized regular grids

For a Γ centred grid ($s_i=0$), the coordinates of a $\mathbf{q}$ point in the basis $(\mathbf{a}^*_\mathrm{p},\mathbf{b}^*_\mathrm{p},\mathbf{c}^*_\mathrm{p})$ are given by

$$\boldsymbol{q}=\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{m}+\boldsymbol{G}, \tag{67}$$

where $\boldsymbol{m}=(m_1,m_2,m_3)^\mathsf{T}$ and the reciprocal lattice vector $\boldsymbol{G}$ is chosen to bring q_i in the interval $[0,1)$. If the regular grid is a traditional one, we simply have $\boldsymbol{Q}=\boldsymbol{P}=\mathbf{1}$. As shown by equation (18), the image of this $\mathbf{q}$ point through an operation of the crystallographic point group is given by

$$\boldsymbol{q}'=\boldsymbol{S}^{-\mathsf{T}}\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{m}+\boldsymbol{S}^{-\mathsf{T}}\boldsymbol{G}+\boldsymbol{G}' \tag{68}$$

where we have added a reciprocal lattice vector $\boldsymbol{G}'$ to shift q_i' of the image point in the interval $[0,1)$. If $\boldsymbol{q}'$ belongs to the grid we have defined, for all $\boldsymbol{S}^{-\mathsf{T}}$ in the crystallographic point group, we will say that the grid is invariant under the crystallographic point group. If it is so, $\boldsymbol{q}'$ can be written as

$$\boldsymbol{q}'=\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{m}'+\boldsymbol{G}'' \tag{69}$$

with $m_i'\in\{0,1,\ldots,n_i-1\}$. Comparing equations (68) and (69), we obtain

$$\boldsymbol{m}'=(\boldsymbol{Q}\boldsymbol{D}^{-1})^{-1}\boldsymbol{S}^{-\mathsf{T}}(\boldsymbol{Q}\boldsymbol{D}^{-1})\boldsymbol{m} \tag{70}$$

$$+(\boldsymbol{Q}\boldsymbol{D}^{-1})^{-1}(\boldsymbol{S}^{-\mathsf{T}}\boldsymbol{G}+\boldsymbol{G}'-\boldsymbol{G}'') \tag{71}$$

Because $\boldsymbol{Q}$ and $\boldsymbol{S}^{-\mathsf{T}}$ are unimodular, and $\boldsymbol{D}$ contains the number of divisions along $(\tilde{\mathbf{a}}^*_\mathrm{p},\tilde{\mathbf{b}}^*_\mathrm{p},\tilde{\mathbf{c}}^*_\mathrm{p})$, the last term is always a reciprocal lattice vector. The matrix

$$\tilde{\boldsymbol{S}}^{-\mathsf{T}}=(\boldsymbol{Q}\boldsymbol{D}^{-1})^{-1}\boldsymbol{S}^{-\mathsf{T}}(\boldsymbol{Q}\boldsymbol{D}^{-1}) \tag{72}$$

is the matrix representation of $\mathcal{S}$ in the $(\tilde{\mathbf{a}}^*_\mathrm{m},\tilde{\mathbf{b}}^*_\mathrm{m},\tilde{\mathbf{c}}^*_\mathrm{m})$ basis. Its determinant is always 1 or -1, but for $\boldsymbol{m}'$ to be integer, it has to have integer entries, and therefore be unimodular. This is checked in the phono3py code, and if it is true for all $\boldsymbol{S}^{-\mathsf{T}}$ in the crystallographic point group, the regular grid follows the crystallographic point group, and we consider it is properly defined. The last equation can also be written as

$$\boldsymbol{P}^{-1}\tilde{\boldsymbol{S}}^{-\mathsf{T}}\boldsymbol{P}=\boldsymbol{M}_\mathrm{g}\boldsymbol{S}^{-\mathsf{T}}\boldsymbol{M}_\mathrm{g}^{-1} \tag{73}$$

and an equivalent strategy is to check that both sides are unimodular.

Introducing the subgrid shift $\boldsymbol{s}$ (see section 6.3) can further break the symmetry of the regular grid. In equation (67) $\boldsymbol{m}$ is replaced by $\boldsymbol{m}+\boldsymbol{P}\boldsymbol{s}$ and $\boldsymbol{m}'$ by $\boldsymbol{m}'+\boldsymbol{P}\boldsymbol{s}$ in equation (69). Requiring the shift to be invariant $\pmod 1$ under an operation of the crystallographic point group, $\tilde{\boldsymbol{S}}^{-\mathsf{T}}\boldsymbol{P}\boldsymbol{s}=\boldsymbol{P}\boldsymbol{s}\pmod 1$, we obtain the condition

$$\boldsymbol{P}^{-1}\tilde{\boldsymbol{S}}^{-\mathsf{T}}\boldsymbol{P}\boldsymbol{s}=\boldsymbol{M}_\mathrm{g}\boldsymbol{S}^{-\mathsf{T}}\boldsymbol{M}_\mathrm{g}^{-1}\boldsymbol{s}=\boldsymbol{s}\pmod 1 \tag{74}$$

against all $\boldsymbol{S}^{-\mathsf{T}}$ in the crystallographic point group.

6.5. Double grid for subgrid shift

To satisfy the crystallographic symmetry, the subgrid shift s_i is normally chosen to be either 0 or $1/2$. It is better to treat grid point arithmetic by integers for the computer implementation and its performance. This is realized by doubling $\boldsymbol{m}$ and $\boldsymbol{s}$. This well-known technique, e.g. presented in [3], is directly usable for the generalized regular grid in section 6.2. As a choice, we define it by

$$\boldsymbol{m}^\mathrm{d}=2(\boldsymbol{m}+\boldsymbol{P}\boldsymbol{s}),\ \ (\mathrm{mod}\ 2\boldsymbol{n}), \tag{75}$$

and the $\mathbf{q}$ points are given like in equation (67) by

$$\boldsymbol{q}=\boldsymbol{Q}\boldsymbol{D}^{-1}\boldsymbol{m}^\mathrm{d}/2+\boldsymbol{G}. \tag{76}$$

The symmetry operation is implemented as

$$\boldsymbol{m}^{\mathrm{d}\prime}=\tilde{\boldsymbol{S}}^{-\mathsf{T}}\boldsymbol{m}^\mathrm{d}\ \ (\mathrm{mod}\ 2\boldsymbol{n}). \tag{77}$$

The index of the grid point $\boldsymbol{m}^{\mathrm{d}\prime}$ is obtained by equation (66) after recovering $\boldsymbol{m}'$ using equation (75),

$$\boldsymbol{m}'=(\boldsymbol{m}^{\mathrm{d}\prime}-2\boldsymbol{P}\boldsymbol{s})/2\ \ (\mathrm{mod}\ )\boldsymbol{n}. \tag{78}$$

In the phono3py code, an integer vector of $2\boldsymbol{s}$ is used for the implementation, instead of $\boldsymbol{s}$, to avoid floating point arithmetic.

6.6. Grid points in BZ

In this section, the BZ grid points are defined to include different $\mathbf{q}$ points on the BZ surface that may be equivalent grid points by reciprocal lattice translations, in addition to the grid points inside the BZ.

For a Γ centred grid, in the $(\tilde{\mathbf{a}}^*_\mathrm{m},\tilde{\mathbf{b}}^*_\mathrm{m},\tilde{\mathbf{c}}^*_\mathrm{m})$ basis, each BZ grid point is represented by the integer triplet

$$\boldsymbol{m}^\mathrm{BZ}=\begin{pmatrix}m_1^\mathrm{BZ}\\ m_2^\mathrm{BZ}\\ m_3^\mathrm{BZ}\end{pmatrix}=\begin{pmatrix}m_1+G_1^{\boldsymbol{m}}n_1\\ m_2+G_2^{\boldsymbol{m}}n_2\\ m_3+G_3^{\boldsymbol{m}}n_3\end{pmatrix},\ G_i^{\boldsymbol{m}}\in\mathbb{Z}, \tag{79}$$

where $\boldsymbol{G}^{\boldsymbol{m}}$ is chosen to minimize the norm of $\mathbf{q}=(\tilde{\mathbf{a}}^*_\mathrm{p},\tilde{\mathbf{b}}^*_\mathrm{p},\tilde{\mathbf{c}}^*_\mathrm{p})\boldsymbol{D}^{-1}\boldsymbol{m}^\mathrm{BZ}=(\tilde{\mathbf{a}}^*_\mathrm{m},\tilde{\mathbf{b}}^*_\mathrm{m},\tilde{\mathbf{c}}^*_\mathrm{m})\boldsymbol{m}^\mathrm{BZ}$. Therefore, $\boldsymbol{m}^\mathrm{BZ}$ is written as

$$\boldsymbol{m}^\mathrm{BZ}=\boldsymbol{m}+\boldsymbol{D}\ \underset{\boldsymbol{G}^{\boldsymbol{m}}}{\arg\min}(|(\tilde{\mathbf{a}}^*_\mathrm{p},\tilde{\mathbf{b}}^*_\mathrm{p},\tilde{\mathbf{c}}^*_\mathrm{p})(\boldsymbol{D}^{-1}\boldsymbol{m}+\boldsymbol{G}^{\boldsymbol{m}})|). \tag{80}$$

Multiple $\boldsymbol{m}^\mathrm{BZ}$ can be found when the grid point is on the BZ surface.

In the phono3py code, equation (80) is implemented as follows. As the first step, reduced basis vectors of the reciprocal primitive cell are obtained by using the Niggli reduction or any reduction scheme. This is written using the change-of-basis matrix $\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}}$ as

$$(\mathbf{a}^*_{\mathrm{r}}, \mathbf{b}^*_{\mathrm{r}}, \mathbf{c}^*_{\mathrm{r}}) = (\mathbf{a}^*_{\mathrm{p}}, \mathbf{b}^*_{\mathrm{p}}, \mathbf{c}^*_{\mathrm{p}})\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}}, \tag{81}$$

where $\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}}$ is unimodular. This gives

$$\mathbf{q} = (\mathbf{a}^*_{\mathrm{r}}, \mathbf{b}^*_{\mathrm{r}}, \mathbf{c}^*_{\mathrm{r}})(\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}})^{-1}\boldsymbol{QD}^{-1}[\boldsymbol{m} + \boldsymbol{DG}^{\boldsymbol{m}}] \tag{82}$$

$$= (\mathbf{a}^*_{\mathrm{r}}, \mathbf{b}^*_{\mathrm{r}}, \mathbf{c}^*_{\mathrm{r}})\left[(\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}})^{-1}\boldsymbol{QD}^{-1}\boldsymbol{m} + \boldsymbol{G}\right]. \tag{83}$$

The second line defines the reciprocal lattice vector $\boldsymbol{G}$. $\boldsymbol{G}$ is divided into two pieces,

$$\boldsymbol{G} = \bar{\boldsymbol{G}} + \Delta\boldsymbol{G}. \tag{84}$$

Nearest integers of $-(\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}})^{-1}\boldsymbol{QD}^{-1}\boldsymbol{m}$ are stored in $\bar{\boldsymbol{G}}$, which is formally written as

$$\bar{\boldsymbol{G}} = -\mathrm{nint}[(\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}})^{-1}\boldsymbol{QD}^{-1}\boldsymbol{m}] \tag{85}$$

to bring the following $\bar{\mathbf{q}}$ closer to the origin,

$$\bar{\mathbf{q}} = (\mathbf{a}^*_{\mathrm{r}}, \mathbf{b}^*_{\mathrm{r}}, \mathbf{c}^*_{\mathrm{r}})\left[(\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}})^{-1}\boldsymbol{QD}^{-1}\boldsymbol{m} + \bar{\boldsymbol{G}}\right]. \tag{86}$$

$\Delta\boldsymbol{G}$ that minimizes $|\bar{\mathbf{q}} + (\mathbf{a}^*_{\mathrm{r}}, \mathbf{b}^*_{\mathrm{r}}, \mathbf{c}^*_{\mathrm{r}})\Delta\boldsymbol{G}|$ is searched in $\Delta G_i \in \{-2,-1,0,1,2\}$. Finally, $\boldsymbol{m}^{\mathrm{BZ}}$ is obtained as

$$\boldsymbol{m}^{\mathrm{BZ}} = \boldsymbol{m} + \boldsymbol{DQ}^{-1}\boldsymbol{M}^*_{\mathrm{p}\to\mathrm{r}}(\bar{\boldsymbol{G}} + \Delta\boldsymbol{G}). \tag{87}$$

As written above, $\boldsymbol{m}$ and p are in one-to-one correspondence and multiple $\boldsymbol{m}^{\mathrm{BZ}}$ can be found for each $\boldsymbol{m}$ or p. In the phono3py code, $\{\{\boldsymbol{m}^{\mathrm{BZ}}(p)\}|0 \leqslant p < \det(\boldsymbol{D})\}$ is calculated and stored at the initial step.

With the subgrid shift, $\boldsymbol{m}$ in the equations above are replaced by $\boldsymbol{m} + \boldsymbol{Ps}$. Either with or without the subgrid shift, $\boldsymbol{m}^{\mathrm{BZ}}$ is determined using the double grid technique presented in section 6.5. From stored $\boldsymbol{m}^{\mathrm{BZ}}$, the BZ $\mathbf{q}$ points are obtained by

$$\mathbf{q} = (\mathbf{a}^*_{\mathrm{p}}, \mathbf{b}^*_{\mathrm{p}}, \mathbf{c}^*_{\mathrm{p}})\boldsymbol{QD}^{-1}(\boldsymbol{m}^{\mathrm{BZ}} + \boldsymbol{Ps}). \tag{88}$$

6.7. Irreducible grid points

Using symmetry properties of phonons, e.g. equation (25), the computation required for the phonon calculations can be reduced. Indeed, many phonon properties result from the BZ integration of functions of a single $\mathbf{q}$ variable. Collecting the values of those functions for the $\mathbf{q}$ points in the irreducible part of the BZ only may speed up the calculation and also save memory space. When the BZ integration is performed on a regular grid, the irreducible $\mathbf{q}$ points are obtained from the irreducible grid points. In this section, we explain how to obtain those irreducible grid points, where it is assumed that the regular grid satisfies the symmetry as described in section 6.4.

A 2D example of BZ is presented in figure 8. A symmetry operation acts on a $\mathbf{q}$ point as $\mathbf{q}' = \mathcal{S}\mathbf{q}$. The star of $\mathbf{q}$ is the set of those points written as $\{\mathcal{S}\mathbf{q}|\mathcal{S} \in \mathbb{P}\}$, where one $\mathbf{q}$ point represents its star. The irreducible BZ is the set of representative $\mathbf{q}$ points of all stars in the BZ. In figure 8(a), the 2D BZ has the eight-fold symmetry of $C_{4\mathrm{v}}$, and the irreducible BZ is depicted by the shaded area.

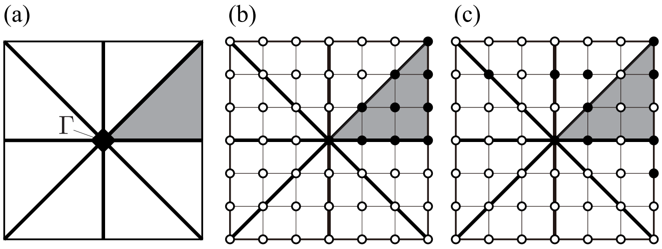

Figure 8. (a) 2D Brillouin zone (BZ). Shaded area is a choice of the irreducible BZ. (b) 6 × 6 regular grid on the BZ including the BZ surface. Filled and open circle symbols show the grid points belonging to the irreducible BZ under the $C_{4\mathrm{v}}$ symmetry and the other grid points, respectively. (c) The irreducible grid points may not be located in a connected space.

In a phonon calculation, $\mathbf{q}$ points are sampled on a regular grid, and the irreducible grid points are defined as the grid points that belong to the irreducible BZ as shown in figure 8(b). In practical calculations, the irreducible grid points considered may however not be located in a connected space, as shown in figure 8(c). This simply means that the selected representative point in the star is located somewhere else.

In the phonopy and phono3py codes, the irreducible grid points are obtained as follows. Each grid point indexed by p is represented in the double grid of equation (75). All symmetry operations $\tilde{\boldsymbol{S}}^{-\intercal}$ of the crystallographic point group are applied to the grid point as given by equation (77). The grid point index of the rotated grid point, $p(\mathcal{S})$, is recovered successively applying equations (78) and (66). The minimum value in $\{p(\mathcal{S})|p(\mathcal{S}) \leqslant p, \mathcal{S} \in \mathbb{P}\}$ is chosen as the irreducible grid point. The mappings of the grid point indices $p \xrightarrow{\{\mathcal{S}\}} \underline{p}$ are stored for all grid points in $0 \leqslant p < \det(\boldsymbol{D})$, and the unique elements of $\{\underline{p}\}$ give the irreducible grid points of the regular grid. It is important to perform these operations only by integers for computational efficiency.

6.8. q point triplets

In the phono3py code, triplets of $\mathbf{q}$ points, $(\mathbf{q}, \mathbf{q}', \mathbf{q}'')$, are considered for the computation of the phonon–phonon interaction. This interaction is therefore not a function of a single $\mathbf{q}$ variable. The number of grid points $\det(\boldsymbol{D})$ becomes large when dense sampling is used, e.g. $\det(\boldsymbol{D}) = 300^3 = 27 \times 10^6$ in the study of [20]. Then, the number of triplets of $\mathbf{q}$ points becomes $300^9 \sim 10^{22}$, which is a really huge number for numerical computations and also storing in the memory of computers.

Due to lattice translational symmetry, elements of the three phonon interaction strength can be non-zero only if $\mathbf{q} + \mathbf{q}' + \mathbf{q}'' = \mathbf{G}$ [32, 33]. This is used to reduce the computation of the interactions for all triplets $(\mathbf{q}, \mathbf{q}', \mathbf{q}'')$ to all pairs of points $(\mathbf{q},\mathbf{q}')$. To satisfy this condition, the Γ centred regular grid is used in the phonon–phonon interaction calculation of the phono3py code. Moreover, the crystallographic point group operations, time-reversal symmetry, and permutation symmetry for each pair of $\mathbf{q}$ and $\mathbf{q}'$ are used to reduce further the number of interactions to be computed. However, even doing so, the number of the combination of two $\mathbf{q}$ points is still large. Therefore, in this section, we describe the strategy used to make those computations practical.

As mentioned in section 6.7, many phonon related properties are written as sum of functions of a single variable $\mathbf{q}$, such as $\sum_{\mathbf{q}} f(\mathbf{q})$. For example, under the relaxation time approximation, the LTC computed in the phono3py code can be written as

$$\kappa = \frac{1}{NV_{\mathrm{c}}} \sum_{\mathbf{q}\nu} \tau_{\mathbf{q}\nu} C_{\mathbf{q}\nu} \mathbf{v}_{\mathbf{q}\nu} \otimes \mathbf{v}_{\mathbf{q}\nu}, \tag{89}$$

where $C_{\mathbf{q}\nu}$ and $\mathbf{v}_{\mathbf{q}\nu}$ denote the phonon mode heat capacity and group velocity, respectively,

$$C_{\mathbf{q}\nu} = k_{\mathrm{B}} \left(\frac{\hbar\omega_{\mathbf{q}\nu}}{k_{\mathrm{B}}T}\right)^2 \frac{\exp(\hbar\omega_{\mathbf{q}\nu}/k_{\mathrm{B}}T)}{[\exp(\hbar\omega_{\mathbf{q}\nu}/k_{\mathrm{B}}T)-1]^2}, \tag{90}$$

$$\mathbf{v}_{\mathbf{q}\nu} = \frac{\partial\omega_{\mathbf{q}\nu}}{\partial\mathbf{q}}. \tag{91}$$

In equation (90), $\hbar$ and k_{B} denote the reduced Planck constant and the Boltzmann constant, respectively, and T is the temperature. $\tau_{\mathbf{q}\nu}$ in equation (89) is the phonon lifetime, and the reciprocal of $\tau_{\mathbf{q}\nu}$ is calculated from the phonon–phonon interaction strength in the phono3py code. The explicit equation used for $\tau_{\mathbf{q}\nu} = 1/2\Gamma_{\mathbf{q}\nu}(\omega_{\mathbf{q}\nu})$ is given at equation (111).

The LTC of equation (89) is conveniently computed iterating over irreducible $\mathbf{q}$ points. Indeed, the mode heat capacity and the lifetime have the same symmetry as the phonon band structure, equation (25),

$$C_{\mathcal{S}\mathbf{q}\nu} = C_{\mathbf{q}\nu}, \tag{92}$$

$$\tau_{\mathcal{S}\mathbf{q}\nu} = \tau_{\mathbf{q}\nu}, \tag{93}$$

and the derivative of equation (25) gives

$$\mathbf{v}_{\mathcal{S}\mathbf{q}\nu} = \mathcal{S}\mathbf{v}_{\mathbf{q}\nu}. \tag{94}$$

Therefore, the summation in equation (89) can be reduced to a summation over the irreducible grid points. Denoting $\underline{\mathbf{q}}$ the irreducible points, we obtain

$$\kappa = \frac{1}{NV_{\mathrm{c}}} \sum_{\underline{\mathbf{q}}\nu} \tau_{\underline{\mathbf{q}}\nu} C_{\underline{\mathbf{q}}\nu} \left(\sum_{\mathcal{S}} \mathcal{S}\mathbf{v}_{\underline{\mathbf{q}}\nu} \otimes \mathcal{S}\mathbf{v}_{\underline{\mathbf{q}}\nu}\right) \frac{m(\underline{\mathbf{q}})}{|\{\mathcal{S}\}|} \tag{95}$$

where the last factor is the number of branches in the star of $\underline{\mathbf{q}}$, $m(\underline{\mathbf{q}})$, divided by the cardinality of the crystallographic point group, $|\{\mathcal{S}\}|$.

When each calculation of $\tau_{\underline{\mathbf{q}}\nu}$ is computationally demanding, the computation of $\tau_{\underline{\mathbf{q}}\nu}$ at different $\underline{\mathbf{q}}$ points may be distributed over multiple or many computer nodes. Therefore, it is convenient to have a set of $\mathbf{q}$ point triplets at fixed $\underline{\mathbf{q}}$ in $\{(\underline{\mathbf{q}},\mathbf{q}',\mathbf{q}'')\}$. As shown in equation (95), in this strategy, $\underline{\mathbf{q}}$ is chosen from the irreducible BZ. Now, because $\underline{\mathbf{q}}$ must be kept fixed, it is not the full set of operations of the crystallographic point group which should be used to reduce the summation over $\mathbf{q}'$, but a subgroup which lets $\underline{\mathbf{q}}$ invariant. This subgroup is known as the point group of $\underline{\mathbf{q}}$,

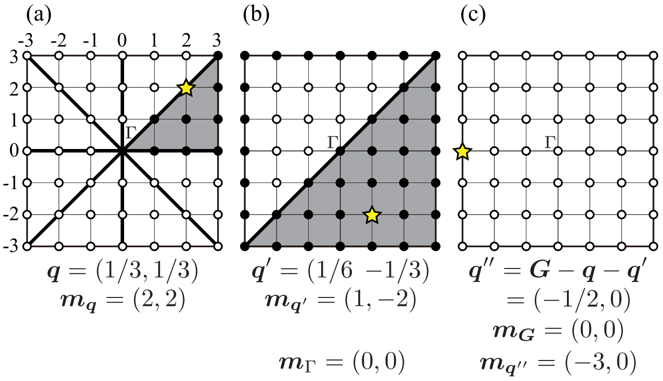

Figure 9. A strategy to choose a $\mathbf{q}$ point triplet. The same 2D Brillouin zone (BZ) as figure 8 is used. The star symbols depict three $\mathbf{q}$ points in the BZ under the constraint of $\boldsymbol{q}+\boldsymbol{q}'+\boldsymbol{q}''=\boldsymbol{G}$. (a) $\boldsymbol{q}$ is chosen from the irreducible grid points of $C_{4\mathrm{v}}$ symmetry. $\boldsymbol{q}$ works as the stabilizer of the subgroup of $C_{4\mathrm{v}}$ symmetry. (b) $\boldsymbol{q}'$ is chosen in the irreducible grid points of the subgroup. (c) $\boldsymbol{q}''=\boldsymbol{G}-\boldsymbol{q}-\boldsymbol{q}'$ where $\boldsymbol{G}$ of the shortest $|(\mathbf{a}^*,\mathbf{b}^*)\boldsymbol{G}|$, i.e. $\boldsymbol{G}=(0,0)$ in this example, is chosen.

$$\mathbb{P}_{\underline{\mathbf{q}}} = \{\mathcal{S}|\mathcal{S}\underline{\mathbf{q}} = \underline{\mathbf{q}} \pmod{\boldsymbol{G}},\ \mathcal{S}\in\mathbb{P}\}. \tag{96}$$

Consequently, $\mathbf{q}'$ is chosen in the irreducible part of the BZ defined by $\mathbb{P}_{\underline{\mathbf{q}}}$. Finally, $\mathbf{q}''$ is computed as $\mathbf{q}''=\mathbf{G}-\underline{\mathbf{q}}-\mathbf{q}'$ where $\mathbf{G}$ is a reciprocal lattice vector chosen to shift $\mathbf{q}''$ within the BZ of the same origin. For phonons, one of $(\underline{\mathbf{q}},\mathbf{q}',\mathbf{q}'')$ and $(\underline{\mathbf{q}},\mathbf{q}'',\mathbf{q}')$ is chosen due to the permutation symmetry.

In the phono3py code, $\mathbf{q}$, $\mathbf{q}'$, and $\mathbf{q}''$ are taken from the BZ of the same origin. When some of them are on the BZ surface, they are chosen among their translationally equivalent points on the BZ surface so that the triplet minimizes $|\mathbf{G}|$ of $\underline{\mathbf{q}}+\mathbf{q}'+\mathbf{q}''=\mathbf{G}$.

This triplet search is implemented using the irreducible grid points and the BZ grid points described in sections 6.7 and 6.6, respectively. The $\mathbf{q}$ points $\underline{\boldsymbol{q}}$, $\boldsymbol{q}'$, and $\boldsymbol{q}''$ are represented by the BZ grid points $\boldsymbol{m}^{\mathrm{BZ}}_{\underline{\boldsymbol{q}}}$, $\boldsymbol{m}^{\mathrm{BZ}}_{\boldsymbol{q}'}$, and $\boldsymbol{m}^{\mathrm{BZ}}_{\boldsymbol{q}''}$, respectively. As shown in figure 9(a), $\boldsymbol{m}^{\mathrm{BZ}}_{\underline{\boldsymbol{q}}}$ is chosen in the irreducible grid points. The point $\boldsymbol{m}^{\mathrm{BZ}}_{\underline{\boldsymbol{q}}}$ breaks the crystallographic point-group $\mathbb{P}$. Using $\tilde{\boldsymbol{S}}^{-\mathsf{T}}$ given by equation (72), the point group of $\underline{\boldsymbol{q}}$ is obtained as

$$\mathbb{P}_{\underline{\boldsymbol{q}}} = \left\{\mathcal{S}|\boldsymbol{m}^{\mathrm{BZ}}_{\underline{\boldsymbol{q}}} = \tilde{\boldsymbol{S}}^{-\mathsf{T}}\boldsymbol{m}^{\mathrm{BZ}}_{\underline{\boldsymbol{q}}} \pmod{\boldsymbol{n}},\ \ \mathcal{S}\in\mathbb{P}\right\}. \tag{97}$$

As shown in figure 9(b), $\boldsymbol{m}^{\mathrm{BZ}}_{\boldsymbol{q}'}$ is sampled from the irreducible grid points of $\mathbb{P}_{\underline{\boldsymbol{q}}}$. The third grid point is given by $\boldsymbol{m}^{\mathrm{BZ}}_{\boldsymbol{q}''} = \boldsymbol{DG}-\boldsymbol{m}^{\mathrm{BZ}}_{\underline{\boldsymbol{q}}}-\boldsymbol{m}^{\mathrm{BZ}}_{\boldsymbol{q}'}$. Finally, $\boldsymbol{m}^{\mathrm{BZ}}_{\underline{\boldsymbol{q}}}$, $\boldsymbol{m}^{\mathrm{BZ}}_{\boldsymbol{q}'}$, or $\boldsymbol{m}^{\mathrm{BZ}}_{\boldsymbol{q}''}$ may be shifted to minimize $|\mathbf{G}|$ if some of them are on the BZ surface as explained for $(\underline{\mathbf{q}},\mathbf{q}',\mathbf{q}'')$ above.

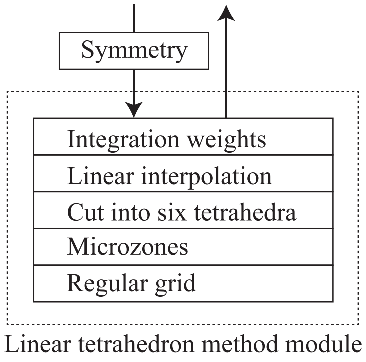

Figure 10. Implementation of the linear tetrahedron method in the phonopy and phono3py codes. This routine is provided as a module that returns integration weights. In the module, each layer depends on the lower layers. Crystal symmetry is treated outside of this module.

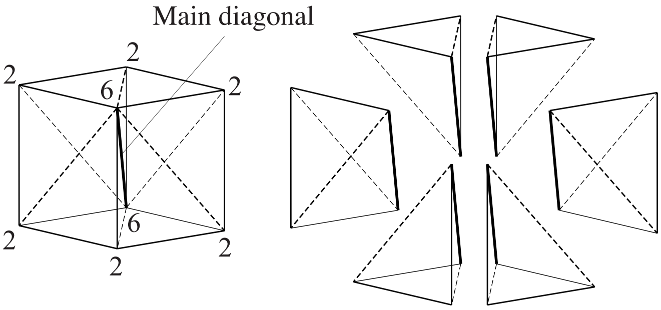

Figure 11. A scheme to divide a parallelepiped microzone into six tetrahedra with the same volume [3]. The eight vertices of the parallelepiped are shared by 6, 6, 2, 2, 2, 2, 2, and 2 tetrahedra, respectively.

7. Tetrahedron method

Once implemented, tetrahedron methods are easy to use and robust. Moreover, when the code is written in a modular way, it can be reusable for different kinds of BZ integrations. The phonopy and phono3py codes employ a linear tetrahedron method where several techniques are picked and mixed from the various reports, notably, by MacDonald *et al* [2] and Blöchl *et al* [3]

The implementation in the phonopy and phono3py codes is structured as schematically illustrated in figure 10. The bottom of the structure is the regular grid. The parallelepiped formed by the basis vectors of the reciprocal primitive cell is divided into many smaller parallelepiped microzones on the regular grid. Every microzone has the same shape, defined by the basis vectors $(\mathbf{a}^*_\mathrm{m}, \mathbf{b}^*_\mathrm{m}, \mathbf{c}^*_\mathrm{m})$, as presented in section 6. In the case of the traditional regular grid, equation (56) defines the microzones to be considered, while for the generalized regular grid equation (59) is used. Each of the microzones is divided into six tetrahedra in the same way as shown in figure 11.

The integral over the BZ to be performed is then regarded as a sum of contributions from those tetrahedra. The function to be integrated is linearly interpolated within each tetrahedron (see section 7.2) and therefore the integration within each tetrahedron can be done analytically. To interpolate linearly a tridimensional function one needs four values. Each tetrahedron has four vertices. Therefore the values of the function to be integrated on the vertices of the tetrahedra are used to build the interpolations.

Different recipes can be used to implement this program. We follow [2] as well as [3] to obtain integration weights for grid points rather than for the tetrahedra themselves. Indeed, as it will be shown, it is possible to rearrange the contributions from tetrahedra to contributions from the grid points. This is useful to make the tetrahedron method easy use, and to make it similar to smearing methods. In addition, this way, the symmetry of the regular grid described in section 6.4 is applied straightforwardly to the integration weights.

7.1. Division into six tetrahedra

The scheme to divide the microzone into six tetrahedra follows [3] reported by Blöchl *et al* We choose a shortest main diagonal of the parallelepiped $(\mathbf{a}^*_\mathrm{m}, \mathbf{b}^*_\mathrm{m}, \mathbf{c}^*_\mathrm{m})$. Then the six tetrahedra are selected sharing this main diagonal as shown in figure 11.

7.2. Functions

In this section we introduce the functions $g(\omega)$, $n(\omega)$, $I(\omega)$, and $J(\omega)$ to be computed using the linear tetrahedron method. The notation roughly follows [2]. Moreover, the band index ν is not written explicitly, since each band can be treated independently. It is therefore understood that a sum over the bands should be performed at the end of the calculation.

The density of states $g(\omega)$ is written as

$$g(\omega) \equiv \int_{BZ} \mathrm{d}^3 q \delta(\omega - \omega_\mathbf{q}) \tag{98}$$

$$\approx V_\mathrm{t} \sum_{i=1}^{6N} g(\omega, \omega_1^i, \omega_2^i, \omega_3^i, \omega_4^i)$$

$$\equiv V_\mathrm{t} \sum_{i=1}^{6N} g^i. \tag{99}$$

The definition is written at the first line, the approximation obtained from the linear tetrahedron method at the second line and an obvious definition of g^i, the contribution of tetrahedron i, at the third line. Here i is an index running throughout the $6N$ tetrahedra, and V_t is the volume of a tetrahedron. The values of frequency at the vertices are assumed to be arranged in ascending order, $\omega_1^i < \omega_2^i < \omega_3^i < \omega_4^i$.

The integrated density of states, or number of states function, $n(\omega)$, is written following the same conventions,

$$n(\omega) \equiv \int_{-\infty}^{\omega} \mathrm{d}\omega' g(\omega') \tag{100}$$

$$\approx V_\mathrm{t} \sum_{i=1}^{6N} n(\omega, \omega_1^i, \omega_2^i, \omega_3^i, \omega_4^i)$$

$$\equiv V_\mathrm{t} \sum_{i=1}^{6N} n^i. \tag{101}$$

Weighted density of states frequently appears in phonon calculations. They are defined by

$$\begin{aligned} I(\omega) &= \int_{\mathrm{BZ}} \mathrm{d}^3 q F(\mathbf{q}) \delta(\omega - \omega_{\mathbf{q}}) \\ &\approx V_{\mathrm{t}} \sum_{i=1}^{6N} g^i \sum_{k=1}^{4} I_k(\omega, \omega_1^i, \omega_2^i, \omega_3^i, \omega_4^i) F_k^i \\ &\equiv V_{\mathrm{t}} \sum_{i=1}^{6N} g^i \sum_{k=1}^{4} I_k^i F_k^i, \end{aligned} \tag{102}$$

where $F(\mathbf{q})$ is a function of $\mathbf{q}$ and we used the notation $F_k^i = F(\mathbf{q}_k^i)$ at the vertex k of the tetrahedron labelled by i. $I_k^i = I_k(\omega, \omega_1^i, \omega_2^i, \omega_3^i, \omega_4^i)$ are given in appendix A. We notice in this equation an additional summation over the vertices of each tetrahedron. This comes from the function F, which is linearly interpolated within each tetrahedron. When $F = 1$, this equation reduces to the density of states.

Finally, the integral of $I(\omega)$ over frequency, $J(\omega)$, is written as

$$\begin{aligned} J(\omega) &= \int_{-\infty}^{\omega} \mathrm{d}\omega' I(\omega') = \int_{\mathrm{BZ}} \mathrm{d}^3 q F(\mathbf{q}) \theta(\omega - \omega_{\mathbf{q}}) \\ &\approx V_{\mathrm{t}} \sum_{i=1}^{6N} n^i \sum_{k=1}^{4} J_k(\omega, \omega_1^i, \omega_2^i, \omega_3^i, \omega_4^i) F_k^i \\ &\equiv V_{\mathrm{t}} \sum_{i=1}^{6N} n^i \sum_{k=1}^{4} J_k^i F_k^i, \end{aligned} \tag{103}$$

where $J_k^i = J_k(\omega, \omega_1^i, \omega_2^i, \omega_3^i, \omega_4^i)$ are given in appendix A.

In appendix A, the formulae for the coefficients appearing for the quantities, g^i, n^i, I_k^i, J_k^i, are stated from [2]. To derive those formulae, the calculation of $J(\omega)$ is considered first, and $I(\omega)$ is obtained through differentiation. $n(\omega)$ and $g(\omega)$ are just special cases for $F = 1$. The $\theta(\omega - \omega_{\mathbf{q}})$ function in $J(\omega)$ shows that computing the contribution from one tetrahedron is related to the estimation of the volume $\omega_{\mathbf{q}} < \omega$, which can be described from an intersection of a tetrahedron with a plane $\omega_{\mathbf{q}} = \omega$, since $\omega_{\mathbf{q}}$ is now assumed to be a linear function of $\mathbf{q}$. As shown in figure 12 the volume will assume different shapes depending on the value of ω. Different formulae are therefore obtained for the different cases to be considered.

7.3. Integration weights

In equation (102), $I(\omega)$ is expressed as a summation over $6N$ tetrahedra and four vertices. Following Blöchl *et al* in [3], this can be rearranged as a sum over grid points. This scheme hides the cumbersome handling of tetrahedra from outside the linear tetrahedron method module as shown in figure 10. Once integration weights, that are described in this section, are computed, the data structure is shared with smearing methods and the symmetry handling is made easy.

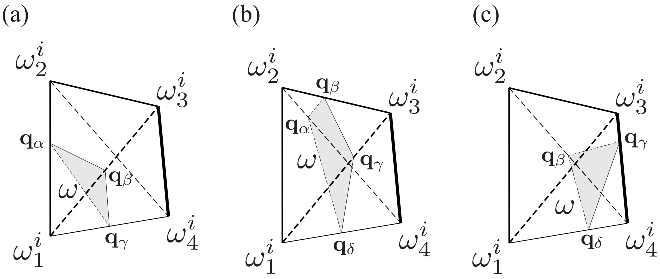

Figure 12. Different linear cuts by ω plane of a tetrahedron. (a) $\omega_1^i < \omega < \omega_2^i < \omega_3^i < \omega_4^i$, (b) $\omega_1^i < \omega_2^i < \omega < \omega_3^i < \omega_4^i$, and (c) $\omega_1^i < \omega_2^i < \omega_3^i < \omega < \omega_4^i$ [2].

$I(\omega)$ can be rewritten as

$$I(\omega) \approx V_{\mathrm{t}} \sum_{p=0}^{N-1} \sum_{i_p=1}^{24} g^{i_p} \sum_{k=1}^{4} I_k^{i_p} F_k^{i_p} \delta_{(i_p,k),p}, \tag{104}$$

where p denotes the grid points (see section 6.3), and i_p is the index of a tetrahedron in the 24 tetrahedra that are sharing the grid point p (see figure 11 for $24 = 6 + 6 + 2 + 2 + 2 + 2 + 2 + 2$). The composite index (i_p, k) indicates a grid point, and $\delta_{(i_p,k),p}$ selects the vertex which is identical to the grid point p in the four vertices of the tetrahedron i_p.

Since $F_k^{i_p} \delta_{(i_p,k),p}$ is only non-zero when (i_p, k) is the point p, we write $F_k^{i_p} \delta_{(i_p,k),p} = F_p \delta_{(i_p,k),p}$. The summation (104) becomes

$$I(\omega) \approx V_{\mathrm{t}} \sum_{p=0}^{N-1} F_p w_p, \tag{105}$$

where

$$w_p = \sum_{i_p=1}^{24} g^{i_p} \sum_{k=1}^{4} I_k^{i_p} \delta_{(i_p,k),p}. \tag{106}$$

$I(\omega)$ has therefore been expressed as a summation over grid points. This integral is given by the weighted sum of the values of the function F at points p, F_p, with weights w_p given by equation (106).

A 2D illustration of the summation (105) is presented in figure 13. Figure 13(a) shows that each parallelogram is cut in two triangles, and the four vertices of each parallelogram are shared by 2, 2, 1, and 1 of those triangles. Six triangles share the grid point marked by the circle symbol as shown in figure 13(b). This shows that only six terms are summed up in the summation (106) because of $\delta_{(i_p,k),p}$. For the tetrahedra in 3D, 24 terms are summed up at each grid point in the summation (106).

To compute equation (106), it is convenient to represent the positions of grid points using shifts from the grid point p, $\boldsymbol{m}_p + \Delta\boldsymbol{m}$. The 2D example is depicted in figure 14. The main diagonal $(0,0)$–$(1,1)$ is chosen. For this main diagonal, the six triangles that share the central point are uniquely determined.

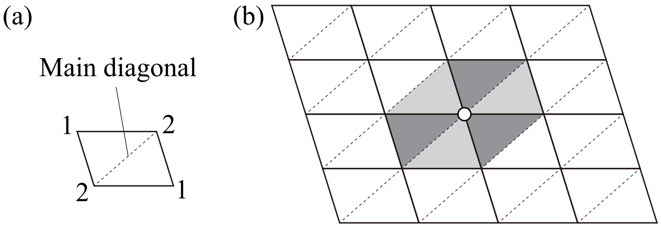

Figure 13. 2D schematic illustration of rearrangement of summation (104). (a) Each parallelogram is cut into two triangles. Four vertices of the parallelogram are shared by 2, 2, 1, and 1 triangles, respectively. (b) Six triangles (shaded) that share a grid point (circle symbol) contribute to the integration weight of the grid point.

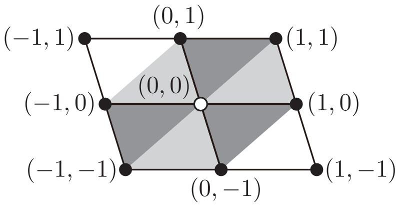

Figure 14. 2D schematic illustration of grid point shifts $\Delta\boldsymbol{m}$ from the focused grid point (centre) to the neighbours.

They are represented by the following $\Delta\boldsymbol{m}$ of the apices: $(0,0)$–$(0,1)$–$(1,1)$, $(0,0)$–$(1,1)$–$(1,0)$, ...

For the linear tetrahedron method, there are four choices for the main diagonal, and for each choice of the main diagonal, the $\Delta\boldsymbol{m}$ of the vertices of the 24 tetrahedra shared by a grid point are predetermined. Therefore, the datasets of the 24 tetrahedra of the four main diagonals are predetermined and hard-corded in the phonopy and phono3py codes as $4\times 24\times 4\times 3$ integer array whose elements are in $\{-1,0,1\}$ for the regular grid.

For the generalized regular grid, the shifts -1 or 1 correspond to the coordinates of the neighbouring points in the $(\mathbf{a}^*_\mathrm{m}, \mathbf{b}^*_\mathrm{m}, \mathbf{c}^*_\mathrm{m})$ basis. The coordinates of those shifts are however easily obtained in the $(\tilde{\mathbf{a}}^*_\mathrm{m}, \tilde{\mathbf{b}}^*_\mathrm{m}, \tilde{\mathbf{c}}^*_\mathrm{m})$ basis used in the calculation, $\Delta\boldsymbol{m}^\mathrm{G} = \boldsymbol{P}\Delta\boldsymbol{m}$, since

$$\begin{aligned}(\mathbf{a}^*_\mathrm{m}, \mathbf{b}^*_\mathrm{m}, \mathbf{c}^*_\mathrm{m})\Delta\boldsymbol{m} &= (\mathbf{a}^*_\mathrm{m}, \mathbf{b}^*_\mathrm{m}, \mathbf{c}^*_\mathrm{m})\boldsymbol{P}^{-1}\boldsymbol{P}\Delta\boldsymbol{m} \\ &= (\tilde{\mathbf{a}}^*_\mathrm{m}, \tilde{\mathbf{b}}^*_\mathrm{m}, \tilde{\mathbf{c}}^*_\mathrm{m})\Delta\boldsymbol{m}^\mathrm{G}, \end{aligned} \tag{107}$$

where equation (63) is used in the last equation.

7.4. Use of symmetry of grid points

The symmetry of grid points can be applied to the linear tetrahedron method. The symmetry is used to prepare the input dataset of a grid point required by the linear tetrahedron method module as illustrated in figure 10. Once the input dataset is made, symmetry information is unnecessary for the linear tetrahedron method module.

In this section, we denote $\omega_p \equiv \omega^i_k$ and the mapping of the grid point $p \xrightarrow{\{\mathcal{S}\}} \underline{p}$ defined in section 6.7 is written as $\underline{p} = f(p)$. Under the condition $\omega_p = \omega_{f(p)}$, which is satisfied by the phonon frequencies, the integration weights fulfil the relation

$$w_p \approx w_{f(p)}. \tag{108}$$

This is an approximation since the set of 24 tetrahedra may not follow the crystallographic point group, i.e. the rotated tetrahedron may not be mapped to any of the 24 tetrahedra. However, equation (108) is a very good approximation.

When F_p has the symmetry $F_p = F_{f(p)}$, equation (105) can be written as

$$I(\omega) \approx V_\mathrm{t}\sum_{p=0}^{N-1} F_p w_p \approx V_\mathrm{t}\sum_{p=0}^{N-1} F_{f(p)} w_{f(p)}. \tag{109}$$

The right hand side of Approx. (109) can therefore be computed from the values at irreducible grid points only. Defining the multiplicity of the irreducible points as $m(\underline{p}) = |\{p|\underline{p} = f(p),\ \forall p\}|$, the right hand side of Approx. (109) is written as

$$I(\omega) \approx V_\mathrm{t}\sum_{\{\underline{p}\}} F_{\underline{p}} w_{\underline{p}} m(\underline{p}). \tag{110}$$

Here $m(\underline{p})$ is equivalent to $m(\underline{\mathbf{q}})$ in equation (95), i.e. the number of branches in the star of $\underline{\mathbf{q}}$.

7.5. Triplets integration weights

As shown in [33], the phonon linewidth computed by the phono3py code, due to phonon–phonon interactions, is given by

$$\Gamma_{\mathbf{q}\nu}(\omega) = \frac{18\pi}{\hbar^2}\sum_{\mathbf{q}'\mathbf{q}''}\sum_{\nu'\nu''}\left|\Phi_{-\mathbf{q}\nu,\mathbf{q}'\nu',\mathbf{q}''\nu''}\right|^2\{(n_{\mathbf{q}'\nu'} + n_{\mathbf{q}''\nu''} + 1)\delta(\omega - \omega_{\mathbf{q}'\nu'} - \omega_{\mathbf{q}''\nu''}) \tag{111}$$

$$+ (n_{\mathbf{q}'\nu'} - n_{\mathbf{q}''\nu''})\left[\delta(\omega + \omega_{\mathbf{q}'\nu'} - \omega_{\mathbf{q}''\nu''}) - \delta(\omega - \omega_{\mathbf{q}'\nu'} + \omega_{\mathbf{q}''\nu''})\right]\}. \tag{112}$$

If we consider $\mathbf{q}$ to be fixed, the summations to be performed over the BZ have the form

$$K_{\mathbf{q}}(\omega) = \frac{1}{N}\sum_{\mathbf{q}'\mathbf{q}''} F_{\mathbf{q},\mathbf{q}',\mathbf{q}''}\Delta(\mathbf{q}+\mathbf{q}'+\mathbf{q}'')\delta(\omega-\omega_{\mathbf{q}',\mathbf{q}''}) \tag{113}$$

where $F_{\mathbf{q},\mathbf{q}',\mathbf{q}''}$ is an arbitrary function of $\mathbf{q}$, $\mathbf{q}'$ and $\mathbf{q}''$. $\omega_{\mathbf{q}',\mathbf{q}''}$ is a function of $\mathbf{q}'$ and $\mathbf{q}''$. For example, in equation (111), to compute the contribution from the first delta function we need to consider, $\omega_{\mathbf{q}',\mathbf{q}''} = \omega_{\mathbf{q}'} + \omega_{\mathbf{q}''}$.

In the above equation, $\Delta(\mathbf{q}+\mathbf{q}'+\mathbf{q}'')$ means 1 when $\mathbf{q}+\mathbf{q}'+\mathbf{q}'' = \mathbf{G}$ otherwise 0. In equation (111) this factor is included in $\Phi_{\mathbf{q}\nu,\mathbf{q}'\nu',\mathbf{q}''\nu''}$ [33]. For given $\mathbf{q}$, $\mathbf{q}'$, and $\mathbf{q}''$ in the BZ, $\mathbf{q}''$ is chosen so that $|\mathbf{G}|$ is smallest. At fixed $\mathbf{q}$, for given $\mathbf{q}'$, only a single $\mathbf{q}''$ contributes. This allows to eliminate the summation over $\mathbf{q}''$, therefore we obtain

$$K_{\mathbf{q}}(\omega) = \frac{1}{N}\sum_{\mathbf{q}'} F_{\mathbf{q},\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'}\delta(\omega-\omega_{\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'}) \tag{114}$$

$$= \frac{V_{\mathrm{c}}}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}^3 q' F_{\mathbf{q},\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'}\delta(\omega-\omega_{\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'}). \tag{115}$$

This quantity has therefore the form of the function $I(\omega)$ we considered previously. To apply the linear tetrahedron method to $K_{\mathbf{q}}(\omega)$ we need to know the value of the functions $F_{\mathbf{q},\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'}$ and $\omega_{\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'}$ at the vertices $\mathbf{q}'+\Delta\mathbf{q}'$ around the point $\mathbf{q}'$. They are given by

$$F_{\mathbf{q},\mathbf{q}'+\Delta\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'-\Delta\mathbf{q}'} = F_{\mathbf{q},\mathbf{q}'+\Delta\mathbf{q}',\mathbf{q}''-\Delta\mathbf{q}'}, \tag{116}$$

$$\omega_{\mathbf{q}'+\Delta\mathbf{q}',\mathbf{G}-\mathbf{q}-\mathbf{q}'-\Delta\mathbf{q}'} = \omega_{\mathbf{q}'+\Delta\mathbf{q}',\mathbf{q}''-\Delta\mathbf{q}'}. \tag{117}$$

It means that $\Delta\mathbf{q}'' = -\Delta\mathbf{q}'$ for the corresponding neighbouring point of $\mathbf{q}''$. Using the values $\omega_{\mathbf{q}'+\Delta\mathbf{q}',\mathbf{q}''-\Delta\mathbf{q}'}$ as the input dataset of the linear tetrahedron method, the triplets integration weights are computed in the same way as written in section 7.3.

8. Random displacement generation

Once phonon frequency $\omega_{\mathbf{q}\nu}$ and eigenvector $W_{\kappa\alpha}(\mathbf{q}\nu)$ are obtained in a supercell, they can be used to generate atomic configurations relevant at finite temperatures. This can be achieved using equation (27) and writing

$$Q(\mathbf{q}\nu) = \frac{1}{\sqrt{2}}(Q^R(\mathbf{q}\nu) + \mathrm{i}Q^I(\mathbf{q}\nu)), \tag{118}$$

where both $Q^R(\mathbf{q}\nu)$ and $Q^I(\mathbf{q}\nu)$ are real variables. It is shown in [34] that they are real normal coordinates, fulfilling the equation of the harmonic oscillator. Their probability density is therefore a Gaussian,

$$P[Q^{R/I}(\mathbf{q}\nu)] = \frac{1}{\sqrt{2\pi\sigma^2_{\mathbf{q}\nu}}}\exp\left[-\frac{Q^{R/I}(\mathbf{q}\nu)^2}{2\sigma^2_{\mathbf{q}\nu}}\right], \tag{119}$$

with

$$\sigma^2_{\mathbf{q}\nu} = \frac{\hbar}{2\omega_{\mathbf{q}\nu}}\coth\frac{\hbar\omega_{\mathbf{q}\nu}}{2k_{\mathrm{B}}T} = \frac{\hbar}{2\omega_{\mathbf{q}\nu}}(1+2n_{\mathbf{q}\nu}), \tag{120}$$

where $n_{\mathbf{q}\nu} = (\mathrm{e}^{\frac{\hbar\omega_{\mathbf{q}\nu}}{k_{\mathrm{B}}T}} - 1)^{-1}$ is the Bose–Einstein distribution function.

Substituting equation (118) into equation (27) and remembering $Q(-\mathbf{q}\nu) = Q^*(\mathbf{q}\nu)$, we obtain

$$u_{l\kappa\alpha} = \frac{1}{\sqrt{Nm_\kappa}}\Bigg\{\sum_{\mathbf{q}\in A,\nu} Q(\mathbf{q}\nu)\mathrm{Re}[u_{l\kappa\alpha}(\mathbf{q}\nu)] + \frac{1}{\sqrt{2}}\sum_{\mathbf{q}\in B,\nu}\Big[(Q^R(\mathbf{q}\nu)+\mathrm{i}Q^I(\mathbf{q}\nu))u_{l\kappa\alpha}(\mathbf{q}\nu) + (Q^R(\mathbf{q}\nu)-iQ^I(\mathbf{q}\nu))u^*_{l\kappa\alpha}(\mathbf{q}\nu)\Big]\Bigg\} \tag{121}$$

$$= \frac{1}{\sqrt{Nm_\kappa}}\Bigg\{\sum_{\mathbf{q}\in A,\nu} Q(\mathbf{q}\nu)\mathrm{Re}[u_{l\kappa\alpha}(\mathbf{q}\nu)] + \sqrt{2}\sum_{\mathbf{q}\in B,\nu}\Big[Q^R(\mathbf{q}\nu)\mathrm{Re}[u_{l\kappa\alpha}(\mathbf{q}\nu)] - Q^I(\mathbf{q}\nu)\mathrm{Im}[u_{l\kappa\alpha}(\mathbf{q}\nu)]\Big]\Bigg\}. \tag{122}$$

Here, A is a set of $\mathbf{q}$ points commensurate with the supercell, where $Q(\mathbf{q}\nu)$ becomes real (e.g. Γ point) or $\mathbf{q} = -\mathbf{q}+\mathbf{G}$, and B includes other commensurate $\mathbf{q}$ points in the one side of the BZ and $\mathbf{q} \neq -\mathbf{q}+\mathbf{G}$. Namely, only one of $\mathbf{q}$ or $-\mathbf{q}$ should be included in the summation over $\mathbf{q}\in B$ [34].

Finally, generating random numbers following equation (119) and using them for $Q^R(\mathbf{q}\nu)$ and $Q^I(\mathbf{q}\nu)$ in equation (122), random displacements $\{u_{l\kappa\alpha}\}$ in a supercell can be generated systematically at a given temperature.

The random structures generated in this way can be used to investigate the impact of phonon excitation on various physical properties, such as electronic structure and magnetism, at finite temperatures. Indeed, the investigated quantity is computed for each random structure generated, and the average of values obtained for those random structures is computed to mimic the effect of temperature.

We note that this approach is physically valid only when all phonon modes are dynamically stable ($\omega^2_{\mathbf{q}\nu} > 0$) at all commensurate $\mathbf{q}$ points. This condition is usually satisfied in many compounds where lattice anharmonicity is not very strong. However, in some anharmonic systems, including high-temperature metastable phases, the harmonic approximation may yield unstable phonon modes.

Even when all phonons are stable, their frequencies may deviate from experimental values significantly. In such cases, it is recommended to use anharmonic phonon frequencies and eigenvectors at finite temperatures for equations (120) and (122). The anharmonic frequencies and eigenvectors renormalized at finite temperatures can be obtained, for example, by performing a self-consistent phonon calculation using either the real-space stochastic approaches, [35–40] which are often termed the stochastic self-consistent harmonic approximation (SSCHA), or the momentum-space implementation which uses anharmonic force constants [41, 42]. An application of SSCHA performed with random displacements generated by the phonopy code coupled with the ALM force constants calculation code [43] is found in [20].

9. Phonon band unfolding

To draw a phonon band structure for a system with defects computed in a supercell approach, band unfolding technique may be useful. Since even one point defect breaks the periodicity of a crystal, a certain approximation is necessary to achieve this representation. In this section, an implementation of the method proposed by Allen *et al* [4] is explained.

In figure 15, a 2D schematic illustration of a supercell with a vacancy is presented. It is analyzed by presuming the corresponding perfect supercell with the same lattice vectors. Equation (30) gives the relationship between the basis vectors of the primitive cell and those of the supercell, which indicates that the perfect supercell contains $N = \det(\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}})$ primitive cells. The reciprocal basis vectors are given by equation (46). The BZ volume of the primitive cell is N times larger than that of the supercell as shown in figure 16.

For this analysis to be possible, we assume there exists a one-to-one correspondence between the atoms and vacancy sites in the defective supercell and the atoms in the perfect supercell. The atoms and vacancies in the defective supercell can therefore be labelled by the composite index $l\kappa$ in the same way as in the perfect supercell. It means that vacancy sites in the defective supercell must have respective atoms in the perfect supercell. The atoms in the defective supercell can nevertheless have slightly different positions than in the perfect supercell. The 2D schematic illustration we obtain is shown in figure 15.

We write the coordinates of lattice points in the supercell as

$$\mathbf{R}_l = (\mathbf{a}_\mathrm{p}, \mathbf{b}_\mathrm{p}, \mathbf{c}_\mathrm{p})\boldsymbol{l} = (\mathbf{a}_\mathrm{s}, \mathbf{b}_\mathrm{s}, \mathbf{c}_\mathrm{s})\tilde{\boldsymbol{l}}, \tag{123}$$

where $\boldsymbol{l}$ are integers, while $\tilde{\boldsymbol{l}}$ may not be.

Let $\hat{T}_j$ be the jth lattice translation operator which is defined on a function as follows

$$\hat{T}_j f(\mathbf{R}) = f(\mathbf{R} - \mathbf{R}_{l_j}). \tag{124}$$

In this section, $\boldsymbol{l}$ correspond to lattice vectors within the perfect supercell. There are N such lattice vectors and therefore N lattice translation operations. $\hat{T}_j$ maps a lattice vector with index l_i in the perfect supercell to a lattice vector with index l_k. The mapping between indices is determined from

$$\tilde{\boldsymbol{l}}_i \xrightarrow{\hat{T}_j} \tilde{\boldsymbol{l}}_k = \tilde{\boldsymbol{l}}_i + \tilde{\boldsymbol{l}}_j \pmod 1, \tag{125}$$

where $\pmod 1$ means that the lattice points translated outside the supercell are brought back inside the supercell by a supercell lattice translation. The mapping obtained this way is stored in an $N \times N$ permutation matrix. In figure 15, the 2D supercell model is made of 2 × 2 primitive cells. For example, the translation by $\boldsymbol{a}_\mathrm{p}$ brings l_1 to l_2. One more translation by $\boldsymbol{a}_\mathrm{p}$ brings l_2 to l_1 due to the periodicity of the supercell. In the following, the mapping $l_i \xrightarrow{\hat{T}_j} l_k$ is written compactly as $l_k = T_j(l_i)$.

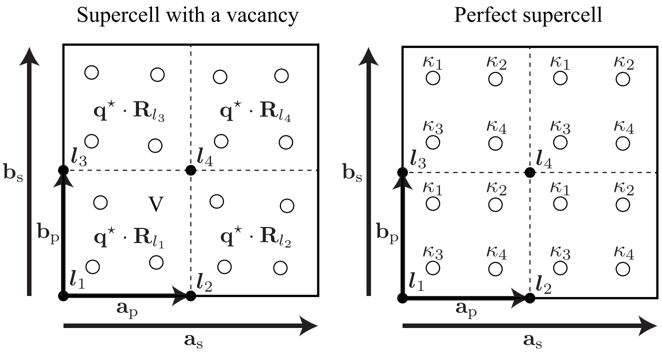

Figure 15. 2D schematic illustration of a supercell with a vacancy (left) and the corresponding perfect supercell (right). The atoms are depicted by open circle symbols. The vacancy is marked by "V". The perfect supercell contains 2 × 2 primitive lattice points ($\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3, \boldsymbol{l}_4$) are indicated by filled circle symbols. Positions of all atoms in the defective supercell can deviate slightly from the equilibrium positions of the perfect supercell. At unfolding, the phase shift of $\mathbf{q}^\star \cdot \mathbf{R}_{l_j}$ is multiplied to the phonon eigenvector of the defective supercell at each primitive lattice translation as given in equation (135).

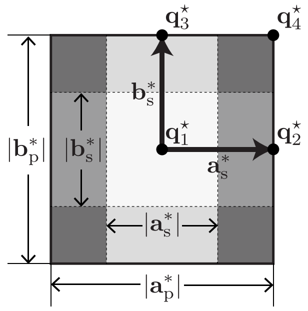
$\mathbf{q}_3^\star$
$\mathbf{q}_4^\star$
$\mathbf{b}_\mathrm{s}^*$
$\mathbf{q}_1^\star$
$\mathbf{q}_2^\star$
$\mathbf{a}_\mathrm{s}^*$
$|\mathbf{b}_\mathrm{p}^*|$
$|\mathbf{b}_\mathrm{s}^*|$
$|\mathbf{a}_\mathrm{s}^*|$
$|\mathbf{a}_\mathrm{p}^*|$

Figure 16. Unfolding of 2D Brillouin zone (BZ) of the supercell ($\mathbf{q}_1^\star$) to 2 × 2 BZs of the supercell ($\mathbf{q}_1^\star, \mathbf{q}_2^\star, \mathbf{q}_3^\star, \mathbf{q}_4^\star$). Gray levels of shaded backgrounds indicate BZs of the supercell which belong to different $\mathbf{q}^\star$ points.

The first step of this method is to compute the phonon eigenvectors of the defective supercell. The Bloch wave of an eigenvector given by equations (19) and (20) is written as

$$\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) = \mathrm{e}^{i\tilde{\mathbf{q}}\cdot\mathbf{R}_{l\kappa}}\mathbf{W}_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}), \tag{126}$$

where $\mathbf{W}_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu})$ is the phonon eigenvector of the defective supercell, with band index $\tilde{\nu}$, and at the wave vector $\tilde{\mathbf{q}}$ in the BZ of the supercell. Eigenvector elements corresponding to vacancies are treated as zeros.

We consider that $\hat{T}_j$ is applied to the Bloch wave as

$$\begin{aligned}\hat{T}_j \Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) &\approx \Psi_{T_j^{-1}(l)\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) \\ &= \mathrm{e}^{i\tilde{\mathbf{q}}\cdot(\mathbf{R}_{l\kappa}-\mathbf{R}_{l_j})}\mathbf{W}_{T_j^{-1}(l)\kappa}(\tilde{\mathbf{q}}\tilde{\nu}).\end{aligned} \tag{127}$$

The first line of equation (127) may not be consistent with equation (124) since positions of atoms in the defective supercell can be displaced slightly from those in the perfect supercell.

The next step is to unfold the $\tilde{\mathbf{q}}$ point from the BZ of the supercell to the N $\mathbf{q}$ points in the BZ of the primitive cell which are equivalent up to vectors of the reciprocal supercell lattice. This is written as

$$\mathbf{q} = \mathbf{q}^\star + \tilde{\mathbf{q}}, \tag{128}$$

where $\mathbf{q}^\star$ is an integer linear combination of $(\mathbf{a}_\mathrm{s}^*, \mathbf{b}_\mathrm{s}^*, \mathbf{c}_\mathrm{s}^*)$. The schematic illustration for the BZ of the 2D supercell is shown in figure 16. The $\tilde{\mathbf{q}}$ points in the BZ of the supercell are shifted by $\mathbf{q}^\star$ inside the BZ of the primitive cell.

Applying periodic boundary conditions on the supercell, the translation operations $\{\hat{T}_j\}$ form an Abelian group of order N with irreducible representations $\mathrm{e}^{-i\mathbf{q}^\star\cdot\mathbf{R}_{l_j}}$. The great orthogonality theorem then gives the relation

$$\frac{1}{N}\sum_{j=1}^{N} \mathrm{e}^{i\mathbf{q}^\star\cdot\mathbf{R}_{l_j}} = \delta_{\mathbf{q}^\star,0}. \tag{129}$$

This result is also proved algebraically in appendix B, together with the dual relation

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{e}^{-i\mathbf{q}_i^\star\cdot(\mathbf{R}_{l_j}-\mathbf{R}_{l_{j'}})} = \delta_{j,j'}. \tag{130}$$

where $\mathbf{q}_i^\star$ are the N vectors $\mathbf{q}^\star$ within the BZ of the primitive cell. We define the operator

$$\hat{P}(\tilde{\mathbf{q}},\mathbf{q}^\star) = \frac{1}{N}\sum_{j=1}^{N} \mathrm{e}^{i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_j}}\hat{T}_j. \tag{131}$$

It can easily be shown that it is a projection operator, which selects the part of $\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu})$ that is also a Bloch function of the primitive lattice with wave vector $\mathbf{q} = \mathbf{q}^\star + \tilde{\mathbf{q}}$. Indeed, in appendix C we obtain

$$\hat{T}_i\hat{P}(\tilde{\mathbf{q}},\mathbf{q}^\star)\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) = \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}}\hat{P}(\tilde{\mathbf{q}},\mathbf{q}^\star)\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}). \tag{132}$$

Moreover, because from equation (130), $\sum_{i=1}^{N}\hat{P}(\tilde{\mathbf{q}},\mathbf{q}_i^\star) = 1$, the decomposition of $\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu})$ into the N functions $\hat{P}(\tilde{\mathbf{q}},\mathbf{q}_i^\star)\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}), i = 1,\ldots,N$ is unique,

$$\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) = \sum_{i=1}^{N}\left[\hat{P}(\tilde{\mathbf{q}},\mathbf{q}_i^\star)\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu})\right]. \tag{133}$$

The projection by equation (131) and the decomposition by equation (133) indicate that the weight of state having the Bloch symmetry $\mathbf{q} = \mathbf{q}^\star + \tilde{\mathbf{q}}$ of the primitive lattice in the supercell Bloch state $\Psi_{l_i\kappa}(\tilde{\mathbf{q}}\tilde{\nu})$ is therefore defined as

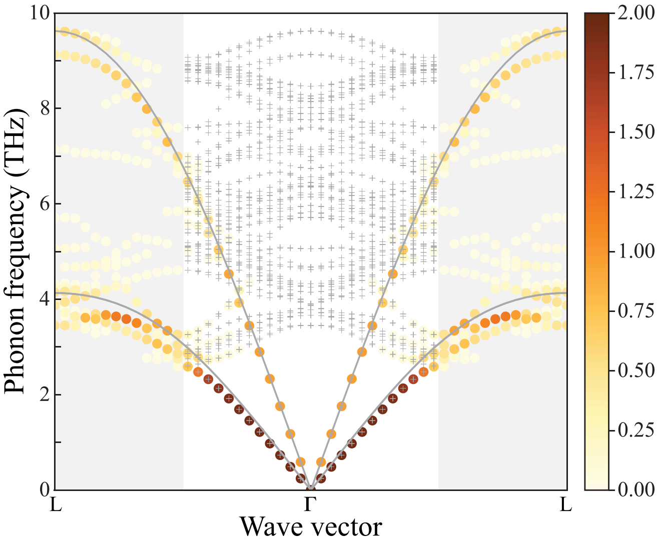

Figure 17. Phonon band unfolding of Al with a vacancy. Color scale indicates unfolding weights of phonon modes depicted by filled circle symbols, where those of the degenerated modes are summed. Phonon modes calculated for the defective supercell are presented by + (plus) symbols. Unshaded and shaded backgrounds show Brillouin zones of the supercell that belong to $\mathbf{q}^\star$ of Γ and L, respectively. Only a part of the weights are unfolded into these two commensurate points among 32 commensurate points. Although this unfolding technique requires no reference states, for the comparison, the phonon band structure of the perfect supercell is shown by solid curves.

$$\begin{aligned} w(\tilde{\mathbf{q}}\tilde{\nu},\mathbf{q}^\star) &= \sum_{i\kappa}|\hat{P}(\tilde{\mathbf{q}},\mathbf{q}^\star)\Psi_{l_i\kappa}(\tilde{\mathbf{q}}\tilde{\nu})|^2 \\ &= \frac{1}{N^2}\sum_{i\kappa}\left|\sum_{j=1}^{N}\mathbf{W}_{T_j^{-1}(l_i)\kappa}(\tilde{\mathbf{q}}\tilde{\nu})\mathrm{e}^{i\mathbf{q}^\star\cdot\mathbf{R}_{l_j}}\right|^2 \qquad (134) \\ &= \frac{1}{N}\sum_{i\kappa}\sum_{j=1}^{N}\mathbf{W}^*_{l_i\kappa}(\tilde{\mathbf{q}}\nu)\mathbf{W}_{T_j^{-1}(l_i)\kappa}(\tilde{\mathbf{q}}\tilde{\nu})\mathrm{e}^{i\mathbf{q}^\star\cdot\mathbf{R}_{l_j}}. \qquad (135) \end{aligned}$$

It can easily be checked that $\sum_{i=1}^{N} w(\tilde{\mathbf{q}}\tilde{\nu},\mathbf{q}_i^\star) = 1$ from equation (130) and the unit normalization of the phonon eigenvectors. Interestingly, this result does not require phonon eigenvectors of the perfect crystal. The second equation is implemented in the phonopy code though the third equation may be more intuitive. $\mathbf{W}_{T_j^{-1}(l_i)\kappa}(\tilde{\mathbf{q}}\tilde{\nu})$ is the phonon eigenvector elements belonging to the lattice point $T_j^{-1}(l_i)$. The unfolding weight indicates its correlation with $\mathbf{W}_{l_i\kappa}(\tilde{\mathbf{q}}\nu)$ after the phase shift $\mathrm{e}^{i\mathbf{q}^\star\cdot\mathbf{R}_{l_j}}$. In the example of figure 15, the phase shifts are integer multiples of π, since the 2D supercell model is made of 2×2 primitive cells.

An example of the phonon band unfolding technique applied to Al with a vacancy is shown in figure 17. The corresponding perfect supercell (32 atoms) is constructed by $2\times 2\times 2$ of the conventional unit cell (4 atoms). The primitive cell contains one atom and

$$\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}} = \begin{pmatrix} -2 & 2 & 2 \\ 2 & -2 & 2 \\ 2 & 2 & -2 \end{pmatrix}. \tag{136}$$

Under fixed basis vectors of the supercell with a vacancy, internal atomic positions are relaxed. For the unfolding, $\mathbf{q}$ points are sampled along the BZ path L–Γ–L of the primitive cell (see figure 17). Those points can be uniquely decomposed as $\mathbf{q} = \tilde{\mathbf{q}} + \mathbf{q}^\star$, where $\tilde{\mathbf{q}}$ belongs to the BZ of the supercell, and $\mathbf{q}^\star$ is a reciprocal lattice vector of the supercell. For the above example, $\mathbf{q}^\star$ is found to be Γ or L, depending on the position of $\mathbf{q}$ along the path L–Γ–L. Those two possibilities are represented using unshaded and shaded backgrounds in figure 17.

The phonon modes of the defective supercell are calculated at those $\tilde{\mathbf{q}}$ points. The projection operator of equation (131) is $\mathbf{q}^\star$ dependent and only a fraction of the phonon modes are unfolded along the L–Γ–L path, because the selected $\mathbf{q}^\star$ points are only a subset of the commensurate $\mathbf{q}^\star$ points in the summation of equation (133). In figure 17, though we can see phonon frequencies are perturbed by the vacancy, the unfolded phonon band structure roughly follows the phonon band structure of the perfect supercell.

Substitutional defects can be managed straightforwardly. Interstitial defects can be handled in the same way as the vacancy case by defining the corresponding supercell differently [44]. Suppose we have one interstitial defect in the supercell model. It is treated as a usual atom in the primitive cell of the corresponding perfect supercell. In the defective supercell, the interstitial atom is located in one primitive cell, and in the other primitive cells, the interstitial sites are treated as vacancies.

10. Conclusion

A scientific simulation code implements mathematical models that are described with formulae. In those formulae, the information is often not detailed enough for direct computer implementation, and the detailed knowledge can be implicit. Scientific software developers may have to learn it from skilled people around, otherwise find it independently. To ease the latter situation in the modern era of distributed software development, we aimed to provide details of the computational methods employed in the phonopy and phono3py codes, keeping the formalism as close as possible to the implementations.

Not all of the computational methods described in this review are specific to the phonon calculation or the supercell approach. For example, the handling of the regular grids and the linear tetrahedron method presented are applicable to electronic structure calculations. We therefore hope that this review will be a useful source of information for software developers in condensed matter science.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

This work was supported by MEXT Japan through ESISM (Elements Strategy Initiative for Structural Materials) of Kyoto University and JSPS KAKENHI Grant Numbers JP21K04632 and JP21K03424. L C thanks the Institut Universitaire de France and the EXPLOR center hosted by the University de Lorraine.

Appendix A. Functions used in the linear tetrahedron method

In this section we give explicit formulae for g^i, n^i, I^i_k, and J^i_k [2]. They assume different forms depending on the position of ω in the sequence $\omega^i_1 < \omega^i_2 < \omega^i_3 < \omega^i_4$. Also, they are expressed using f^i_{nm} and Δ^i_{nm} which are defined as

$$f^i_{nm} \equiv (\omega - \omega^i_m)/(\omega^i_n - \omega^i_m),\ f^i_{nm} + f^i_{mn} = 1, \tag{A1}$$

$$\Delta^i_{nm} \equiv \omega^i_n - \omega^i_m. \tag{A2}$$

g^i, n^i, I^i, and J^i satisfy the following relations:

$$\frac{\mathrm{d}n^i}{\mathrm{d}\omega} = g^i, \tag{A3}$$

$$\frac{\mathrm{d}(n^i J^i_k)}{\mathrm{d}\omega} = g^i I^i_k. \tag{A4}$$

It may be useful to remember

$$f^i_{n'm}/\Delta^i_{nm} = f^i_{nm}/\Delta^i_{n'm}, \tag{A5}$$

$$\frac{\mathrm{d}f^i_{nm}}{\mathrm{d}\omega} = \frac{1}{\Delta^i_{nm}}. \tag{A6}$$

A.1. $\omega < \omega^i_1$

$$g^i = 0. \tag{A7}$$

$$n^i = 0. \tag{A8}$$

$$I^i_k = 0. \tag{A9}$$

$$J^i_k = 0. \tag{A10}$$

A.2. $\omega_1^i < \omega < \omega_2^i$

As shown in figure 12(a), the volume to consider is a tetrahedron with vertices located at $\mathbf{q}_1$, $\mathbf{q}_\alpha = f_{21}^i \mathbf{q}_2 + f_{12}^i \mathbf{q}_1$, $\mathbf{q}_\beta = f_{31}^i \mathbf{q}_3 + f_{13}^i \mathbf{q}_1$, $\mathbf{q}_\gamma = f_{41}^i \mathbf{q}_4 + f_{14}^i \mathbf{q}_1$, and

$$n^i = f_{21}^i f_{31}^i f_{41}^i. \tag{A11}$$

Inside this volume, $F(\mathbf{q})$ is linearly interpolated, i.e. $F(\mathbf{q}_\alpha) \approx f_{21}^i F(\mathbf{q}_2) + f_{12}^i F(\mathbf{q}_1)$, …, and the contribution of ith tetrahedron to $J(\omega)$ is approximated as $V_\mathrm{t} n^i [F(\mathbf{q}_1) + F(\mathbf{q}_\alpha) + F(\mathbf{q}_\beta) + F(\mathbf{q}_\gamma)]/4$. By substituting it in equation (103), we obtain

$$J_1^i = (1 + f_{12}^i + f_{13}^i + f_{14}^i)/4. \tag{A12}$$

$$J_k^i = f_{k1}^i/4, \quad k = 2,3,4. \tag{A13}$$

Using equations (A3) and (A4),

$$\begin{aligned} g^i &= 3n^i/(\omega - \omega_1^i) = 3 f_{21}^i f_{31}^i f_{41}^i/(\omega - \omega_1^i) \\ &= 3 f_{21}^i f_{31}^i/\Delta_{41}^i. \end{aligned} \tag{A14}$$

$$I_1^i = (f_{12}^i + f_{13}^i + f_{14}^i)/3. \tag{A15}$$

$$I_k^i = f_{k1}^i/3, \quad k = 2,3,4. \tag{A16}$$

A.3. $\omega_2^i < \omega < \omega_3^i$

In this case, the occupied part of the tetrahedron is the sum of the following three tetrahedra. Vertices of the first tetrahedron are $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_\alpha = f_{42}^i \mathbf{q}_4 + f_{24}^i \mathbf{q}_2$, $\mathbf{q}_\beta = f_{32}^i \mathbf{q}_3 + f_{23}^i \mathbf{q}_2$, and its volume is $V_{12\alpha\beta}^i = V_\mathrm{t} f_{42}^i f_{32}^i$. The vertices of the second tetrahedron are $\mathbf{q}_1$, $\mathbf{q}_\alpha$, $\mathbf{q}_\beta$, $\mathbf{q}_\delta = f_{41}^i \mathbf{q}_4 + f_{14}^i \mathbf{q}_1$, and its volume is $V_{1\alpha\beta\delta}^i = V_\mathrm{t} f_{41}^i f_{24}^i f_{32}^i$. The vertices of the third tetrahedron are $\mathbf{q}_1$, $\mathbf{q}_\beta$, $\mathbf{q}_\delta$, $\mathbf{q}_\gamma = f_{31}^i \mathbf{q}_3 + f_{13}^i \mathbf{q}_1$, and its volume is $V_{1\beta\delta\gamma}^i = V_\mathrm{t} f_{41}^i f_{31}^i f_{23}^i$. The locations of the vertices are depicted in figure 12(b). We obtain

$$\begin{aligned} n^i &= (V_{12\alpha\beta}^i + V_{1\alpha\beta\delta}^i + V_{1\beta\delta\gamma}^i)/V_\mathrm{t} \\ &= f_{42}^i f_{32}^i + f_{41}^i f_{24}^i f_{32}^i + f_{41}^i f_{31}^i f_{23}^i. \end{aligned} \tag{A17}$$

$$\begin{aligned} J_1^i = [&V_{12\alpha\beta}^i + V_{1\alpha\beta\delta}^i (1 + f_{14}^i) \\ &+ V_{1\beta\delta\gamma}^i (1 + f_{14}^i + f_{13}^i)]/(4n^i V_\mathrm{t}). \end{aligned} \tag{A18}$$

$$\begin{aligned} J_2^i = [&V_{12\alpha\beta}^i (1 + f_{24}^i + f_{23}^i) + V_{1\alpha\beta\delta}^i (f_{24}^i + f_{23}^i) \\ &+ V_{1\beta\delta\gamma}^i f_{23}^i]/(4n^i V_\mathrm{t}). \end{aligned} \tag{A19}$$

$$J_3^i = [V_{12\alpha\beta}^i f_{32}^i + V_{1\alpha\beta\delta}^i f_{32}^i + V_{1\beta\delta\gamma}^i (f_{32}^i + f_{31}^i)]/(4n^i V_\mathrm{t}). \tag{A20}$$

$$J_4^i = [V_{12\alpha\beta}^i f_{42}^i + V_{1\alpha\beta\delta}^i (f_{42}^i + f_{41}^i) + V_{1\beta\delta\gamma}^i f_{41}^i]/(4n^i V_\mathrm{t}). \tag{A21}$$

$$g^i = 3(f_{23}^i f_{31}^i + f_{32}^i f_{24}^i)/\Delta_{41}^i. \tag{A22}$$

$$I_1^i = f_{14}^i/3 + f_{13}^i f_{31}^i f_{23}^i/(g^i \Delta_{41}^i). \tag{A23}$$

$$I_2^i = f_{23}^i/3 + (f_{24}^i)^2 f_{32}^i/(g^i \Delta_{41}^i). \tag{A24}$$

$$I_3^i = f_{32}^i/3 + (f_{31}^i)^2 f_{23}^i/(g^i \Delta_{41}^i). \tag{A25}$$

$$I_4^i = f_{41}^i/3 + f_{42}^i f_{24}^i f_{32}^i/(g^i \Delta_{41}^i). \tag{A26}$$

A.4. $\omega_3^i < \omega < \omega_4^i$

As shown in figure 12(c), the occupied part of the tetrahedron is the full tetrahedron minus the tetrahedron with vertices at $\mathbf{q}_4$, $\mathbf{q}_\beta = f_{42}^i \mathbf{q}_4 + f_{24}^i \mathbf{q}_2$, $\mathbf{q}_\delta = f_{41}^i \mathbf{q}_4 + f_{14}^i \mathbf{q}_1$ and $\mathbf{q}_\gamma = f_{43}^i \mathbf{q}_4 + f_{34}^i \mathbf{q}_3$. Its volume is $V_{4\beta\delta\gamma}^i = V_\mathrm{t} f_{14}^i f_{24}^i f_{34}^i$. Therefore, the contribution of i-th tetrahedron to $J(\omega)$ is approximated as $\{V_\mathrm{t}[F(\mathbf{q}_1) + F(\mathbf{q}_2) + F(\mathbf{q}_3) + F(\mathbf{q}_4)] - V_{4\beta\delta\gamma}^i [F(\mathbf{q}_4) + F(\mathbf{q}_\beta) + F(\mathbf{q}_\gamma) + F(\mathbf{q}_\delta)]\}/4$. We obtain

$$n^i = (1 - f_{14}^i f_{24}^i f_{34}^i). \tag{A27}$$

$$J_1^i = [1 - (f_{14}^i)^2 f_{24}^i f_{34}^i]/(4n^i). \tag{A28}$$

$$J_2^i = [1 - f_{14}^i (f_{24}^i)^2 f_{34}^i]/(4n^i). \tag{A29}$$

$$J_3^i = [1 - f_{14}^i f_{24}^i (f_{34}^i)^2]/(4n^i). \tag{A30}$$

$$J_4^i = [1 - f_{14}^i f_{24}^i f_{34}^i (1 + f_{41}^i + f_{42}^i + f_{43}^i)]/(4n^i). \tag{A31}$$

$$g^i = 3(1 - n^i)/(\omega_4^i - \omega) = 3 f_{24}^i f_{34}^i/\Delta_{41}^i. \tag{A32}$$

$$I_k^i = f_{k4}^i/3, \quad k = 1,2,3. \tag{A33}$$

$$I_4^i = (f_{41}^i + f_{42}^i + f_{43}^i)/3. \tag{A34}$$

A.5. $\omega_4^i < \omega$

The tetrahedron is fully occupied, therefore,

$$g^i = 0. \tag{A35}$$

$$n^i = 1. \tag{A36}$$

$$I_k^i = 0. \tag{A37}$$

$$J_k^i = 1/4. \tag{A38}$$

Appendix B. Summation formulae

When the supercell lattice vectors are collinear to those of the primitive cell,

$$(\mathbf{a}_\mathrm{s},\mathbf{b}_\mathrm{s},\mathbf{c}_\mathrm{s}) = (\mathbf{a}_\mathrm{p},\mathbf{b}_\mathrm{p},\mathbf{c}_\mathrm{p})\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}. \tag{B1}$$

With $\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}} = \mathrm{diag}(n_1,n_2,n_3)$, it is proved in most textbooks that the set of lattice vectors

$$\mathbf{R}_l = (\mathbf{a}_\mathrm{p},\mathbf{b}_\mathrm{p},\mathbf{c}_\mathrm{p})\begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix}, \quad l_i = 0,\ldots,n_i-1 \tag{B2}$$

$$= (\mathbf{a}_\mathrm{p},\mathbf{b}_\mathrm{p},\mathbf{c}_\mathrm{p})\boldsymbol{l} \tag{B3}$$

and their conjugated wave vectors

$$\mathbf{q} = (\mathbf{a}^*_\mathrm{p},\mathbf{b}^*_\mathrm{p},\mathbf{c}^*_\mathrm{p})\begin{pmatrix} p_1/n_1 \\ p_2/n_2 \\ p_3/n_3 \end{pmatrix}, \quad p_i = 0,\ldots,n_i-1 \tag{B4}$$

$$= (\mathbf{a}^*_\mathrm{s},\mathbf{b}^*_\mathrm{s},\mathbf{c}^*_\mathrm{s})\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \tag{B5}$$

$$= (\mathbf{a}^*_\mathrm{s},\mathbf{b}^*_\mathrm{s},\mathbf{c}^*_\mathrm{s})\boldsymbol{p}, \tag{B6}$$

fulfil the summation formulae

$$\frac{1}{\det(\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}})}\sum_{\{\mathbf{R}_l\}} \mathrm{e}^{i\mathbf{q}\cdot\mathbf{R}_l} = \delta_{\mathbf{q},\mathbf{G}}, \tag{B7}$$

$$\frac{1}{\det(\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}})}\sum_{\{\mathbf{q}\}} \mathrm{e}^{-i\mathbf{q}\cdot\mathbf{R}_l} = \delta_{\mathbf{R}_l,\mathbf{R}_L}, \tag{B8}$$

where $\mathbf{G}$ is any reciprocal lattice vector, and $\mathbf{R}_L$ any supercell lattice vector.

When the supercell is not collinear to the primitive cell, the set of lattice vectors, $\{\mathbf{R}_l\}$, are those integer linear combinations of $(\mathbf{a}_\mathrm{p},\mathbf{b}_\mathrm{p},\mathbf{c}_\mathrm{p})$ located within the supercell $(\mathbf{a}_\mathrm{s},\mathbf{b}_\mathrm{s},\mathbf{c}_\mathrm{s})$, and the set of wave vectors, $\{\mathbf{q}\}$, are those integer linear combinations of $(\mathbf{a}^*_\mathrm{s},\mathbf{b}^*_\mathrm{s},\mathbf{c}^*_\mathrm{s})$ located within the reciprocal cell $(\mathbf{a}^*_\mathrm{p},\mathbf{b}^*_\mathrm{p},\mathbf{c}^*_\mathrm{p})$. Since the above summations can not easily be split into the product of three geometric series, it is not obvious whether the above results are still valid.

However, we have seen in section 3.1 that SNF like transformations can always be used to obtain primitive cell and supercell which are collinear, and for which the summation formulae (B7) and (B8) are obviously valid. We have

$$(\tilde{\mathbf{a}}_\mathrm{p},\tilde{\mathbf{b}}_\mathrm{p},\tilde{\mathbf{c}}_\mathrm{p}) = (\mathbf{a}_\mathrm{p},\mathbf{b}_\mathrm{p},\mathbf{c}_\mathrm{p})\boldsymbol{P}^{-1}, \tag{B9}$$

$$(\tilde{\mathbf{a}}_\mathrm{s},\tilde{\mathbf{b}}_\mathrm{s},\tilde{\mathbf{c}}_\mathrm{s}) = (\mathbf{a}_\mathrm{s},\mathbf{b}_\mathrm{s},\mathbf{c}_\mathrm{s})\boldsymbol{Q}, \tag{B10}$$

and therefore

$$(\tilde{\mathbf{a}}^*_\mathrm{p},\tilde{\mathbf{b}}^*_\mathrm{p},\tilde{\mathbf{c}}^*_\mathrm{p}) = (\mathbf{a}^*_\mathrm{p},\mathbf{b}^*_\mathrm{p},\mathbf{c}^*_\mathrm{p})\boldsymbol{P}^\mathsf{T}, \tag{B11}$$

$$(\tilde{\mathbf{a}}^*_\mathrm{s},\tilde{\mathbf{b}}^*_\mathrm{s},\tilde{\mathbf{c}}^*_\mathrm{s}) = (\mathbf{a}^*_\mathrm{s},\mathbf{b}^*_\mathrm{s},\mathbf{c}^*_\mathrm{s})\boldsymbol{Q}^{-\mathsf{T}} \tag{B12}$$

$$= (\tilde{\mathbf{a}}^*_\mathrm{p},\tilde{\mathbf{b}}^*_\mathrm{p},\tilde{\mathbf{c}}^*_\mathrm{p})\boldsymbol{D}^{-1}, \tag{B13}$$

while the set of lattice vectors and wave vectors,

$$\tilde{\mathbf{R}}_l = (\tilde{\mathbf{a}}_\mathrm{p},\tilde{\mathbf{b}}_\mathrm{p},\tilde{\mathbf{c}}_\mathrm{p})\tilde{\boldsymbol{l}}, \tag{B14}$$

$$\tilde{\mathbf{q}} = (\tilde{\mathbf{a}}^*_\mathrm{s},\tilde{\mathbf{b}}^*_\mathrm{s},\tilde{\mathbf{c}}^*_\mathrm{s})\tilde{\boldsymbol{p}}, \tag{B15}$$

with $\tilde{l}_i = 0,\ldots,D_i-1$ and $\tilde{p}_i = 0,\ldots,D_i-1$, fulfils,

$$\frac{1}{\det(\boldsymbol{D})}\sum_{\{\tilde{\mathbf{R}}_l\}} \mathrm{e}^{i\tilde{\mathbf{q}}\cdot\tilde{\mathbf{R}}_l} = \delta_{\tilde{\mathbf{q}},\mathbf{G}}, \tag{B16}$$

$$\frac{1}{\det(\boldsymbol{D})}\sum_{\{\tilde{\mathbf{q}}\}} \mathrm{e}^{-i\tilde{\mathbf{q}}\cdot\tilde{\mathbf{R}}_l} = \delta_{\tilde{\mathbf{R}}_l,\mathbf{R}_L}. \tag{B17}$$

Since the same lattices are generated by the unimodular transformations (B11) and (B12), we can write

$$\tilde{\mathbf{R}}_l = \mathbf{R}_l + \mathbf{R}_L, \tag{B18}$$

$$\tilde{\mathbf{q}} = \mathbf{q} + \mathbf{G} \tag{B19}$$

where $\mathbf{R}_L$ is a supercell lattice vector used to bring $\tilde{\mathbf{R}}_l$ within the original supercell $(\mathbf{a}_\mathrm{s},\mathbf{b}_\mathrm{s},\mathbf{c}_\mathrm{s})$, and $\mathbf{G}$ a reciprocal lattice vector used to bring $\tilde{\mathbf{q}}$ to the original reciprocal cell $(\mathbf{a}^*_\mathrm{p},\mathbf{b}^*_\mathrm{p},\mathbf{c}^*_\mathrm{p})$. Since $\mathbf{q}\cdot\mathbf{R}_L$, $\mathbf{G}\cdot\mathbf{R}_L$, $\mathbf{G}\cdot\mathbf{R}_l$ are multiple of 2π, and $\det(\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}) = \det(\boldsymbol{D})$, we obtain equations (B7) and (B8) for arbitrary supercell and primitive cell related by an integer matrix $\boldsymbol{M}_{\mathrm{p}\to\mathrm{s}}$. Notice however that the definition of the set of lattice vectors and wave vectors for which those summation formulae hold is now more general. The wave vectors are along the reciprocal vectors of the supercell and located within the reciprocal cell of the primitive lattice, while the lattice vectors are along the primitive lattice vectors and located within the supercell.

Appendix C. Projection into primitive cell Bloch states

From the definition of the projection operator,

$$\hat{T}_i\hat{P}(\tilde{\mathbf{q}},\mathbf{q}^\star)\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) = \hat{T}_i\frac{1}{N}\sum_{j=1}^{N}\mathrm{e}^{i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_j}}\hat{T}_j\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) \tag{C1}$$

$$= \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}}\frac{1}{N}\sum_{j=1}^{N}\mathrm{e}^{i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot(\mathbf{R}_{l_j}+\mathbf{R}_{l_i})} \times\hat{T}_i\hat{T}_j\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) \tag{C2}$$

$$= \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}} \frac{1}{N}\sum_{j=1}^{N} \mathrm{e}^{i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot(\mathbf{R}_{l_j}+\mathbf{R}_{l_i})} \times \mathrm{e}^{i\tilde{\mathbf{q}}\cdot(\mathbf{R}_{l\kappa}-\mathbf{R}_{l_j}-\mathbf{R}_{l_i})}\mathbf{W}_{(T_iT_j)^{-1}(l)\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) \tag{C3}$$

$$= \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}} \frac{1}{N}\sum_{j=1}^{N} \mathrm{e}^{i\mathbf{q}^\star\cdot(\mathbf{R}_{l_j}+\mathbf{R}_{l_i})} \times \mathrm{e}^{i\tilde{\mathbf{q}}\cdot\mathbf{R}_{l\kappa}}\mathbf{W}_{(T_iT_j)^{-1}(l)\kappa}(\tilde{\mathbf{q}}\tilde{\nu}). \tag{C4}$$

$\mathbf{R}_{l_j}+\mathbf{R}_{l_i}$ may be outside the supercell, therefore we write $\mathbf{R}_{l_j}+\mathbf{R}_{l_i}=\mathbf{R}_{l_k}+\mathbf{R}_L$, where $\mathbf{R}_L$ is a supercell lattice vector. Because $\mathbf{W}_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu})$ has the periodicity of the supercell, and $\mathrm{e}^{i\mathbf{q}^\star\cdot\mathbf{R}_L}=1$, we obtain

$$\hat{T}_i\hat{P}(\tilde{\mathbf{q}},\mathbf{q}^\star)\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu})$$

$$= \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}} \frac{1}{N}\sum_{k=1}^{N} \mathrm{e}^{i\mathbf{q}^\star\cdot\mathbf{R}_{l_k}}\mathrm{e}^{i\tilde{\mathbf{q}}\cdot\mathbf{R}_{l\kappa}}\mathbf{W}_{T_k^{-1}(l)\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) \tag{C5}$$

$$= \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}} \frac{1}{N}\sum_{k=1}^{N} \mathrm{e}^{i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_k}}\mathrm{e}^{i\tilde{\mathbf{q}}\cdot(\mathbf{R}_{l\kappa}-\mathbf{R}_{l_k})}\mathbf{W}_{T_k^{-1}(l)\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) \tag{C6}$$

$$= \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}} \frac{1}{N}\sum_{k=1}^{N} \mathrm{e}^{i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_k}}\hat{T}_k\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}) \tag{C7}$$

$$= \mathrm{e}^{-i(\tilde{\mathbf{q}}+\mathbf{q}^\star)\cdot\mathbf{R}_{l_i}}\hat{P}(\tilde{\mathbf{q}},\mathbf{q}^\star)\Psi_{l\kappa}(\tilde{\mathbf{q}}\tilde{\nu}). \tag{C8}$$

ORCID iDs

Atsushi Togo https://orcid.org/0000-0001-8393-9766
Terumasa Tadano https://orcid.org/0000-0002-8132-2161

References

[1] Gonze X and Lee C 1997 *Phys. Rev.* B **55** 10355

[2] MacDonald A H, Vosko S H and Coleridge P T 1979 *J. Phys. C: Solid State Phys.* **12** 2991

[3] Blöchl P E, Jepsen O and Andersen O K 1994 *Phys. Rev.* B **49** 16223

[4] Allen P B, Berlijn T, Casavant D A and Soler J M 2013 *Phys. Rev.* B **87** 085322

[5] Aroyo M I 2016 *International Tables for Crystallography* vol A (Hoboken: Wiley)

[6] Aroyo M I, Orobengoa D, de la Flor G, Tasci E S, Perez-Mato J M and Wondratschek H 2014 *Acta Crystallogr.* A **70** 126

[7] Togo A 2023 *J. Phys. Soc. Japan* **92** 012001

[8] Maradudin A A and Vosko S H 1968 *Rev. Mod. Phys.* **40** 1

[9] Shmueli U 2008 *International Tables for Crystallography* vol B (Hoboken: Wiley)

[10] Bradley T, Bradley C and Cracknell A 1972 *The Mathematical Theory of Symmetry in Solids: Representation Theory for Point Groups and Space Groups* (Oxford: Clarendon)

[11] Cracknell A P and Davies B L 1979 *General Introduction and Tables of Irreducible Representations of Space Groups* (*Kronecker Product Tables* vol 1) (New York: IFI/Plenum)

[12] Horton G K and Maradudin A A (eds) 1974 *Dynamical Properties of Solids* vol 1 (Amsterdam: North-Holland Publishing Company)

[13] Niggli P 1928 Krystallographische und strukturtheoretische Grundbegriffe *Handbuch der Experimentalphysik* vol 7 (Leipzig: Akademische Verlagsgesellschaft) pp 108–76

[14] Gruber B 1973 *Acta Crystallogr.* A **29** 433

[15] Křivý I and Gruber B 1976 *Acta Crystallogr.* A **32** 297

[16] Grosse-Kunstleve R W, Sauter N K and Adams P D 2004 *Acta Crystallogr.* A **60** 1

[17] Hart G L W and Forcade R W 2008 *Phys. Rev.* B **77** 224115

[18] Parlinski K, Li Z Q and Kawazoe Y 1997 *Phys. Rev. Lett.* **78** 4063

[19] Wallace D C 1998 *Thermodynamics of Crystals* (New York: Dover Publications)

[20] Togo A, Hayashi H, Tadano T, Tsutsui S and Tanaka I 2022 *J. Phys.: Condens. Matter.* **34** 365401

[21] Pick R M, Cohen M H and Martin R M 1970 *Phys. Rev.* B **1** 910

[22] Giannozzi P, de Gironcoli S, Pavone P and Baroni S 1991 *Phys. Rev.* B **43** 7231

[23] Gonze X, Charlier J-C, Allan D C and Teter M P 1994 *Phys. Rev.* B **50** 13035

[24] Hinuma Y, Pizzi G, Kumagai Y, Oba F and Tanaka I 2017 *Comput. Mater. Sci.* **128** 140

[25] Togo A and Tanaka I 2018 `Spglib`: a software library for crystal symmetry search (arXiv:1808.01590)

[26] Monkhorst H J and Pack J D 1976 *Phys. Rev.* B **13** 5188

[27] Chadi D J and Cohen M L 1973 *Phys. Rev.* B **8** 5747

[28] Moreno J and Soler J M 1992 *Phys. Rev.* B **45** 13891

[29] Wisesa P, McGill K A and Mueller T 2016 *Phys. Rev.* B **93** 155109

[30] Morgan W S, Jorgensen J J, Hess B C and Hart G L 2018 *Comput. Mater. Sci.* **153** 424

[31] Morgan W S, Christensen J E, Hamilton P K, Jorgensen J J, Campbell B J, Hart G L and Forcade R W 2020 *Comput. Mater. Sci.* **173** 109340

[32] Chaput L, Togo A, Tanaka I and Hug G 2011 *Phys. Rev.* B **84** 094302

[33] Togo A, Chaput L and Tanaka I 2015 *Phys. Rev.* B **91** 094306

[34] Alexei G H W, Maradudin A and Montroll E W (eds) 1963 *Theory of Lattice Dynamics in the Harmonic Approximation* (*Solid State Physics: Advances in Research and Applications*) (New York: Academic)

[35] Monacelli L, Bianco R, Cherubini M, Calandra M, Errea I and Mauri F 2021 *J. Phys.: Condens. Matter* **33** 363001

[36] Errea I, Calandra M and Mauri F 2013 *Phys. Rev. Lett.* **111** 177002

[37] Errea I, Calandra M and Mauri F 2014 *Phys. Rev.* B **89** 064302

[38] Bianco R, Errea I, Paulatto L, Calandra M and Mauri F 2017 *Phys. Rev.* B **96** 014111

[39] Monacelli L, Errea I, Calandra M and Mauri F 2018 *Phys. Rev.* B **98** 024106

[40] Eriksson F, Fransson E and Erhart P 2019 *Adv. Theory Simul.* **2** 1800184

[41] Tadano T and Tsuneyuki S 2015 *Phys. Rev.* B **92** 054301

[42] Wu Y-N, Saidi W A, Wuenschell J K, Tadano T, Ohodnicki P, Chorpening B and Duan Y 2020 *J. Phys. Chem. Lett.* **11** 2518

[43] Tadano T and Tsuneyuki S 2018 *J. Phys. Soc. Japan* **87** 041015

[44] Moxon S *et al* 2022 *J. Mater. Chem.* A **10** 1861