

# The Hiphive Package for the Extraction of High-Order Force Constants by Machine Learning

Fredrik Eriksson, Erik Fransson, and Paul Erhart\*

The efficient extraction of force constants (FCs) is crucial for the analysis of many thermodynamic materials properties. Approaches based on the systematic enumeration of finite differences scale poorly with system size and can rarely extend beyond third order when input data is obtained from first-principles calculations. Methods based on parameter fitting in the spirit of interatomic potentials, on the other hand, can extract FC parameters from semi-random configurations of high information density and advanced regularized regression methods can recover physical solutions from a limited amount of data. Here, the HIPHIVE Python package, that enables the construction of force constant models up to arbitrary order is presented. HIPHIVE exploits crystal symmetries to reduce the number of free parameters and then employs advanced machine learning algorithms to extract the force constants. Depending on the problem at hand, both over and underdetermined systems are handled efficiently. The FCs can be subsequently analyzed directly and or be used to carry out, for example, molecular dynamics simulations. The utility of this approach is demonstrated via several examples including ideal and defective monolayers of MoS<sub>2</sub> as well as bulk nickel.

## 1. Introduction

The vibrational properties of solids are pivotal for a large number of physical phenomena, including phase stability and thermal conduction. In crystalline solids, the vibrational motion of the atoms is periodic and commonly described using phonons—quasi-particles that represent collective excitations of the lattice. At the first level of approximation, phonons can be obtained within the harmonic limit, which implies noninteracting quasi-particles with infinite lifetimes. This approach, along with its so-called quasi-harmonic extension, already provides a wealth of information. There are, however, countless examples where anharmonic effects are crucial and must be accounted for in order to capture the correct physical behavior of a system. Notable examples include the lattice contribution to the thermal conductivity or the vibrational stabilization of metastable phases, for


example, the body-centered cubic phase of the group IV elements (Ti, Zr, Hf) or the cubic phase of zirconia.<sup>[1]</sup> In more general terms, phonon–phonon coupling must be taken into account, which leads to finite phonon lifetimes and temperature-dependent frequencies.

Formally, the analysis of vibrational material properties requires a set of force constants (FCs), which allows the computation of atomic forces solely based on the displacements of atoms from their reference positions. The harmonic approximation requires only knowledge of the second-order FCs, which can be readily extracted using software packages such as PHONOPY.<sup>[2]</sup> Third-order FCs, which are required, for example, for computing the thermal conductivity to the lowest permissible order of permutation theory, can be constructed using, for example, PHONO3PY,<sup>[3]</sup> SHENGBTE,<sup>[4]</sup> ALMABTE,<sup>[5]</sup> and AAPL-AFLOW.<sup>[6]</sup> A systematic enumeration of atomic displacements, while reference forces are usually obtained

using density functional theory (DFT) calculations. It is also possible to obtain FCs from density functional perturbation theory calculations as implemented, for example, in the ABINIT<sup>[7]</sup> and QUANTUM-ESPRESSO packages.<sup>[8]</sup> The latter approach is, however, commonly limited to second- and third-order FCs as the computation of higher-order terms, by power of the  $2n + 1$  theorem (see e.g., ref. [9]), would require knowledge of second and higher-order derivatives of the wave function, which are not commonly available in these codes. FCs beyond third-order are required to describe, for example, metastable systems or the temperature dependence of phonon modes. The number of degrees of freedoms (DOFs) in higher-order FCs increases exponentially with the interaction range and the latter are hence increasingly difficult to extract by enumeration schemes. Alternatively, one can employ regression schemes,<sup>[10,11]</sup> as implemented in the ALAMODE<sup>[12]</sup> and TDEP codes.<sup>[13]</sup> They employ use linear least-square fitting and thus require the number of input forces to exceed the number of parameters, that is, they solve an *overdetermined* problem. More recently, techniques based on compressive sensing<sup>[14]</sup> have been proposed<sup>[15–18]</sup> that can also efficiently solve *underdetermined* systems.

Here, we introduce the HIPHIVE Python package, which allows one to efficiently obtain high-order FCs both in large systems and systems with low-symmetry. HIPHIVE can take advantage of various powerful machine learning algorithms for FC extraction

F. Eriksson, E. Fransson, Prof. P. Erhart  
Department of Physics  
Chalmers University of Technology  
Gothenburg 412 96, Sweden  
E-mail: erhart@chalmers.se

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adts.201800184>

DOI: 10.1002/adts.201800184

via SCIKIT-LEARN<sup>[19]</sup> and it can be readily interfaced with a large number of electronic structure codes via the atomic simulation environment (ASE) package.<sup>[20]</sup> This enables a flexible workflow with easy access to various advanced optimization techniques some of which are designed to find sparse solutions<sup>[15,21]</sup> that reflect the short-range nature of the FCs. If the input configurations are constructed sensibly, this approach requires a *much* smaller number of input configurations and thus considerably reduces the computational effort, the overwhelming part of which is usually associated with DFT calculations. This approach becomes genuinely advantageous for obtaining second-order FCs in large and/or low symmetry systems (defects, interfaces, surfaces, large unit cells, etc.) and high-order FCs, for which a strict enumeration scheme quickly leads to a dramatic increase in the number of force calculations.

FCs can be post-processed in a number of ways including analysis via PHONOPY and PHONO3PY as well as molecular dynamics (MD) simulations via ASE. An extensive user guide is available online,<sup>[22]</sup> which includes a basic tutorial as well as a number of advanced examples. The package is maintained under an open source license on GITLAB and can be installed from the PYPPI index.

The remainder of this paper is organized as follows. The next section provides a concise summary of FC expansions, which sets up a description of algorithmic and methodological aspects in Section 3 as well as the HIPHIVE workflow in Section 4. Possible applications are finally illustrated by several examples in Section 5.

## 2. Force Constants

This section provides a brief introduction to FC expansions. The similarities between an FC and a cluster expansion are described and the effect of crystal symmetries is demonstrated. Finally, the constraints due to translational and rotational symmetry are summarized.

### 2.1. Basics

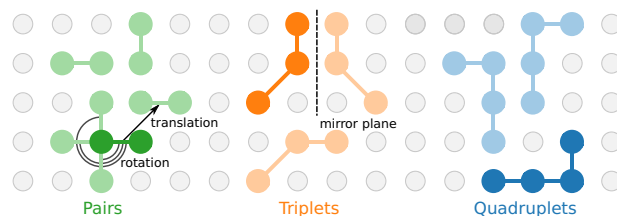
The potential energy  $V$  of a solid can be represented by a Taylor expansion of the potential energy surface (PES) in ionic displacements  $\mathbf{u} = \mathbf{R} - \mathbf{R}_0$  away from the equilibrium positions  $\mathbf{R}_0$

$$V = V_0 + \Phi_i^\alpha u_i^\alpha + \frac{1}{2} \Phi_{ij}^{\alpha\beta} u_i^\alpha u_j^\beta + \frac{1}{3!} \Phi_{ijk}^{\alpha\beta\gamma} u_i^\alpha u_j^\beta u_k^\gamma + \dots$$

where the Einstein summation convention applies and

$$\Phi_i^\alpha = \frac{\partial V}{\partial u_i^\alpha}, \quad \Phi_{ij}^{\alpha\beta} = \frac{\partial^2 V}{\partial u_i^\alpha \partial u_j^\beta} \quad \text{etc.}$$

$\Phi_i, \Phi_{ij}, \dots$  are the FCs corresponding to increasing orders of the expansion. Latin indices  $i$  indicate the atomic labels, where the summation is over an infinite crystal lattice, while Greek indices  $\alpha$  run over the Cartesian coordinates  $x, y, z$ .  $V_0$  is a constant term, which is commonly ignored when dealing with lattice dynamics. The first-order FC is also often dropped since the expansion in



**Figure 1.** Examples for one-body (singlet; red), two-body (pair; green), and three-body (triplet; blue) clusters. Dark colors indicate representative cluster whereas lighter colors represent symmetry equivalent clusters that belong to the associated orbit. While the two-body clusters shown here are related by a simple fourfold rotation, the three-body clusters can be mapped onto each by the application of a fourfold rotation combined with a mirror operation.

displacements can be made around an equilibrium lattice configuration with vanishing forces. These two terms are important for some applications but here are considered zero. Truncating the potential after the second-order term results in the conventional harmonic phonon theory, which is analytically solvable and widely used.<sup>[23,24]</sup>

The forces can be written in terms of the FCs as

$$F_i^\alpha = -\Phi_{ij}^{\alpha\beta} u_j^\beta - \frac{1}{2} \Phi_{ijk}^{\alpha\beta\gamma} u_j^\beta u_k^\gamma - \dots \quad (1)$$

Crucially, this expression that takes the functional form of an interatomic many-body potential is linear in the FCs, which will become relevant in Section 3.

The number of DOFs in the FC expansion scales exponentially with  $\mathcal{O}(N^n)$ , where  $N$  is the number of atomic sites of the supercell and  $n$  is the order after which the expansion is truncated. The number of *independent* parameters is, however, much smaller due to the symmetries of the underlying lattice as well as constraints due to the conservation of linear and angular momentum.<sup>[10]</sup> The number of *numerically relevant* parameters is smaller still due to the decay of the FCs with interaction distance, order, and finite many-body interactions.

### 2.2. Clusters

To represent the FCs and their inherent symmetries, it is convenient to consider clusters of sites  $(ij\dots)$ , each of which can be assigned a size, taken, for example, as the largest distance between any two sites in the cluster. The FCs correspond to interactions between the sites forming a cluster; for example, the term  $\Phi_{ijkk}$  describes a fourth-order interaction in the three-body cluster  $(ijk)$ . Clusters can be categorized based on the number of sites they comprise and accordingly there are one-body (singlet), two-body (pair), and  $n$ -body (many-body) clusters (**Figure 1**). In general, a cluster of order  $n$  is a multiset of order (or cardinality)  $n$  consisting of up to  $n$  different atoms. As a result of the locality of the interactions, one commonly truncates the FC expansion both in size, order, and cluster size. It is thus possible to create, for example, a non-central, short ranged, anharmonic pair potential or a long ranged harmonic pair potential with short ranged anharmonic many-body corrections.

### 2.3. Symmetries

The FCs obey a number of symmetries, the most fundamental of which stems from the requirement that the differentiation of the total energy and thus  $\Phi$  must be invariant under a simultaneous permutation  $P$  of atomic and Cartesian indices

$$\Phi_{ij\dots}^{\alpha\beta\dots} = \Phi_{P(ij\dots)}^{P(\alpha\beta\dots)} \quad (2)$$

The FCs must further comply with the symmetry of the underlying lattice as expressed by the associated spacegroup. Each symmetry operation  $S$  consists of a (possibly improper) rotation  $R$  and a translation  $T$ . If application of  $S$  maps a cluster  $(ij\dots)$  onto another cluster  $(i'j'\dots)$ , the FCs corresponding to these two clusters are related to each other according to

$$\Phi_{ij\dots}^{\alpha\beta\dots} = \Phi_{i'j'\dots}^{\alpha'\beta'\dots} R^{\alpha'\alpha} R^{\beta'\beta} \dots \quad (3)$$

Clusters, which can be transformed into each other as a result of the application of such a symmetry operation, are said to belong to the same *orbit* (Figure 1) and the associated FC parameters are related by the respective symmetry operation. The lattice symmetry can thus be used to reduce the number of free parameters by grouping clusters in orbits.

Furthermore, a symmetry operation that maps a cluster onto itself implies a reduction in the number of internal DOFs in the FCs

$$\Phi_{ij\dots}^{\alpha\beta\dots} = \Phi_{ij\dots}^{P^{-1}(\alpha'\beta'\dots)} R^{\alpha'\alpha} R^{\beta'\beta} \dots \quad (4)$$

where  $P$  can be a permutation, for example, a mapping of the indices  $(ij)$  onto  $(ji)$ . Section 3 addresses how these constraints are imposed in practice.

### 2.4. Constraints

The conservation of linear momentum constrains the FCs further leading to a set of translational (acoustic) sum rules.

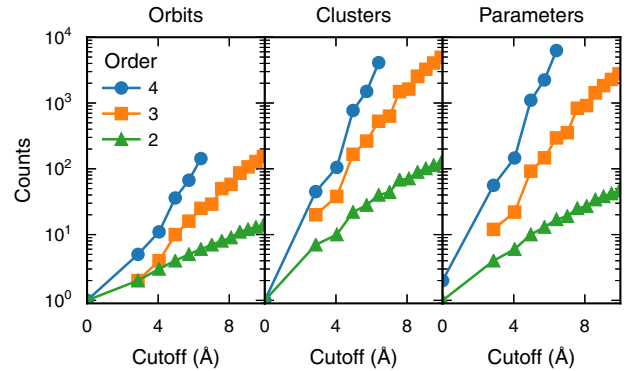
$$\sum_i \Phi_{ij\dots}^{\alpha\beta\dots} = 0 \quad (5)$$

which apply for all Cartesian indices independently.<sup>[23]</sup>

Rotational invariance is generally harder to enforce than translational invariance. This is partly due to subtle difficulties that arise when combining periodic boundary conditions with a rotationally invariant expansion around equilibrium.<sup>[25]</sup> Many authors have described and proposed rotational sum rules.<sup>[24,26–29]</sup> In general, it is recognized that the conservation of angular momentum leads to a set of rotational sum rules that relate force constants of order  $n$  and  $n + 1$  as<sup>[10,24]</sup>

$$\Phi_{i_1\dots i_n}^{\alpha'\dots\alpha_n} \omega_{\gamma}^{\alpha'\alpha_1} + \dots + \Phi_{i_1\dots i_n}^{\alpha_1\dots\alpha'} \omega_{\gamma}^{\alpha'\alpha_n} + \Phi_{i_1\dots i_{n+1}}^{\alpha_1\dots\alpha_{n+1}} \omega_{\gamma}^{\alpha_{n+1}\alpha'} r_{i_{n+1}}^{\alpha'} = 0 \quad (6)$$

where  $\omega$  is a generator of infinitesimal rotations with the same representation as the Levi-Civita symbol and  $r_i^{\alpha}$  denotes the position vector of atom  $i$ .



**Figure 2.** a) Number of unique types of clusters, b) total number of clusters per primitive cell, and c) number of unknown parameters in FCC aluminum as a function of the cutoff radius imposed during construction of the cluster space.

To simplify things, we only consider the two simplest sum rules derived by Born and Huang<sup>[23]</sup> and apply them to second-order FCs. The Born–Huang sum rule reads

$$\Phi_{ij}^{\alpha\beta} r_j^{\gamma} = \Phi_{ij}^{\alpha\gamma} r_j^{\beta} \quad (7)$$

which is a truncation of Equation (6), whereas the Huang invariance imposes

$$\sum_{ij} \Phi_{ij}^{\alpha\beta} r_{ij}^{\gamma} r_{ij}^{\delta} = \sum_{ij} \Phi_{ij}^{\gamma\delta} r_{ij}^{\alpha} r_{ij}^{\beta} \quad (8)$$

where  $r_{ij}$  is the distance vector between the equilibrium positions of atoms  $i$  and  $j$ .

These two constraints do not suffice to render the expansion fully rotationally invariant and thus MD simulations can potentially behave unphysical. As shown below, they do, however, enforce the correct dispersion relation near the  $\Gamma$ -point, which is crucial for, for example, 2D materials.<sup>[25]</sup>

Even with the application of symmetry and sum rules the number of orbits increases rapidly with cutoff and order (Figure 2). Since each orbit is associated with several parameters, the total number of parameters increases even more quickly. As a result for systems with low symmetry, large unit cells, and high orders, the number of model DOFs can easily outnumber the number available reference forces leading to underdetermined problem as discussed below.

## 3. Methodology

This section will start with a concise overview of the computational methods used when implementing the framework described above. In the subsequent subsections, each step is described in more detail.

The *cluster space* represents a fundamental element of the algorithm implemented in HIRHIVE. It comprises information concerning

- the clusters that an atom in the primitive cell can be part of given the cutoff,
- the organization of clusters into orbits,

- the FC for each orbit (relative to a representative cluster of that orbit),
- the free parameters allowed by symmetry in each FC, and
- the constraints imposed on the free parameters by translational and rotational invariance.

The cluster space thus contains all the information needed to construct the template FCs for any supercell of the same structure.

The information is generated by the following procedure, which requires user input in the form of a structure (defined by cell metric, basis, and chemical species) as well as a set of cutoff radii per order.

1. Find all applicable symmetry operations for the input structure using, for example, SPGLIB.<sup>[30]</sup>
2. Enumerate all possible sites in periodic images of the primitive cell located at the origin (zero-cell) that reside within the specified cutoff relative to any atom in the center cell.
3. Construct all possible clusters, which contain at least one site in the zero-cell and are consistent with the cutoff.
4. Apply symmetry operations and establish the symmetry relations between clusters; group the clusters into orbits.
5. Use the symmetry operations that map the representative cluster of an orbit onto itself to reduce the FCs into reduced components that obey the symmetries.
6. Finally, construct the system of equations describing the sum rules and find the solution space satisfying the constraints.

The cluster space can then be used together with a supercell with displaced atoms to create the so-called sensing matrix, which relates the free parameters to the resulting forces.

### 3.1. Clusters and Orbits

Finding the possible clusters in a structure can be done in several ways. Keep in mind that even though a lower-order cluster (e.g., 0-0-3-7) might be invariant under a symmetry, a higher-order cluster involving the same atoms (e.g., 0-3-3-7-7) need not be invariant under the same symmetry. One simple way to generate clusters is to generate all possible unique multisets of increasing cardinality with elements drawn from the atom enumeration of the periodic images. To decide if the cluster (multiset) should be kept, all atoms must be within the cutoff distance from each other and it must contain at least one atom in the zero-cell.

The symmetry operations can now be used to categorize the clusters into orbits, each of which contains a representative cluster. All other clusters in the orbit are related to the representative (prototype) cluster by a symmetry operation together with a permutation. If during the categorization process a symmetry maps a cluster onto a permutation of itself, it can be used to reduce the number of free components in the FC as described below.

### 3.2. Cluster Symmetries and Invariant Bases

If a representative cluster is symmetric under a set of crystal symmetries  $\{S_i\}$ , the corresponding representative FC must

obey the same symmetries. Using multi-index notation  $\cdot^\Lambda = \{i_{ij}^{\alpha\beta}, i_{ijk}^{\alpha\beta\gamma}, \dots\}$ , the problem of finding valid solutions can be transformed into an eigenvalue problem by flattening the tensors

$$\Phi^\Lambda = \Phi^{\Lambda'} \tilde{R}^{P(\Lambda)\Lambda} \quad (9)$$

The eigenvectors, after back-transformation, are then solutions  $\varphi$  to the original Equation (4), which form an invariant basis with respect to the symmetries of the cluster. An FC associated with the respective cluster that fulfills the symmetry can then be constructed as a linear combination of these basis functions

$$\Phi = \sum_p a_p \varphi_p \quad (10)$$

where  $a_p$  represents the parameters that must be found by optimization (see below) and cannot be determined by symmetry alone and the basis elements  $\varphi$  we denote as eigentensors. Equation (9) can in principle be solved by any suitable algorithm. It is, however, preferable to work in scaled coordinates since then the rotation matrices are integer matrices allowing one to use an optimized algorithm. This has some additional advantages including exact precision and conservation of sparsity, that is, a  $3 \times 3$  matrix with nine unknowns can be decomposed into nine dense  $3 \times 3$  matrices or nine sparse  $3 \times 3$  matrices with only one element per matrix.

### 3.3. Translational Sum Rules

While the parametrization obtained after enforcing the crystal symmetries could in principle be used as the final parametrization of the model, the resulting FCs are not guaranteed to be translational or rotational invariant. The constraints due to sum rules can be enforced by projecting the parameters onto a subspace of the parameter space (i.e., the nullspace of Equation (5)) that fulfills the sum rule. As noted below, this method is used in HIPHIVE to enforce the rotational sum rules.

Sum rules can also be applied via re-parametrization by only considering linear combinations of parameter vectors, which span the aforementioned nullspace of the sum rule. HIPHIVE adopts this approach to enforce the translational sum rules using integer arithmetics via SYMPY to create the nullspace.<sup>[31]</sup> The corresponding constraint matrix is constructed as described below.

Let  $\Theta$  denote an orbit and  $T_{\Theta c}$  the combined rotation and permutation operation that maps the representative cluster  $c_\Theta$  onto a cluster  $c$  belonging to the same orbit. Then, together with Equation (10) the sum rule Equation (5) can be written as

$$\sum_i \sum_{\Theta} \sum_p a_{p\Theta} \varphi_{p\Theta}^{\alpha'\beta' \dots} T_{\Theta(ij \dots)}^{\alpha'\beta' \dots \alpha\beta \dots} = 0 \quad \forall j \dots \alpha\beta \dots \quad (11)$$

$$= \mathbf{C}^{\text{trans}} \mathbf{a} = \mathbf{0}$$

where  $\mathbf{C}^{\text{trans}}$  is the translational constraint matrix. Interpreting  $p_\Theta$  as a multi-index, the equation above represents a system of linear equations in the parameters  $a_{p\Theta}$ . The solutions to this system are referred to as constraint vectors  $A_{p\Theta, i}$  and specify how the parameters  $a_{p\Theta}$  must be related in order to fulfill the sum

rules

$$a_{p\ominus} = A_{p\ominus,i} \tilde{a}_i$$

where  $\tilde{a}$  is the new parameters guaranteed to fulfill the translational sum rules.

### 3.4. Sensing Matrix

To extract the independent parameters, the so-called sensing (or fit) matrix  $\mathbf{M}$  must be constructed for each input structure. This matrix, which depends on the  $3N_{\text{at}}$ -dimensional displacement vector  $\mathbf{u}$ , relates the  $3N_{\text{at}}$ -dimensional vector of predicted forces  $\mathbf{f}$  to the  $N_{\text{par}}$ -dimensional parameter vector  $\tilde{\mathbf{a}}$

$$\mathbf{f}(\mathbf{u}, \tilde{\mathbf{a}}) = \mathbf{M}(\mathbf{u})\tilde{\mathbf{a}} \quad (12)$$

where  $N_{\text{at}}$  and  $N_{\text{par}}$  denote the number of atoms in the supercell and the number of parameters, respectively. This form is possible due to Equation (1), which has the form of an interatomic potential with tunable parameters. The crucial step here is to recognize that the forces  $\mathbf{f}$  are linearly related to the FC parameters  $\tilde{\mathbf{a}}$  for given displacements  $\mathbf{u}$ . The FCs in turn are linearly dependent on the FC parameters via the crystal symmetries and the FC parameters associated with the orbits, where the parameters are either the true expansion parameters associated with the orbits or the constrained parameters that include the translational sum rules.

### 3.5. Rotational Sum Rules

In principle, the procedure used to enforce the translational sum rules can also be employed to apply the rotational sum rules. Since the rotational sum rules involve the positions of the lattice sites (which can assume any real value) the algorithm to extract the nullspace must, however, be numerically very robust. Alternatively, one can project a previously determined parametrization onto the correct subspace while maintaining lattice symmetries and translational invariance, which is the approach adopted here.

The sum rules, for example,

$$\Phi_{ij}^{\alpha\beta} r_j^\gamma - \Phi_{ij}^{\alpha\gamma} r_j^\beta = 0$$

are constructed and flattened to a column in a new constraint matrix. This is repeated for all parameters  $\tilde{\mathbf{a}}$  leading to a  $N_{\text{prim}} \times \tilde{N}_{\text{par}}$  matrix where  $N_{\text{prim}}$  is the number of atoms in the primitive cell times  $3^4$  and  $\tilde{N}_{\text{par}}$  is equal to the number of independent parameters after the enforcement (re-parametrization) of the translational sum rules. The sum rules using this new constraint matrix  $\mathbf{C}^{\text{rot}}$  and the parameters  $\tilde{\mathbf{a}}$  is written analogous to the translational sum rules as

$$\mathbf{C}^{\text{rot}}\tilde{\mathbf{a}} = \mathbf{0}$$

Given a solution  $\tilde{\mathbf{a}}$ , the above is in general not fulfilled. Let us assume  $\mathbf{C}^{\text{rot}}\tilde{\mathbf{a}} = \mathbf{d}$ , where  $\mathbf{d}$  is a vector describing how well the sum rule is fulfilled. Now, we want to find a correction  $\Delta\tilde{\mathbf{a}}$  to

$\tilde{\mathbf{a}}$ , which is as small as possible and ensures the sum rules are fulfilled

$$\mathbf{C}^{\text{rot}}(\tilde{\mathbf{a}} + \Delta\tilde{\mathbf{a}}) = \mathbf{0} \quad (13)$$

Thus, we have a new problem, that is, to find  $\Delta\tilde{\mathbf{a}}$  such that  $\|\mathbf{C}^{\text{rot}}\Delta\tilde{\mathbf{a}} + \mathbf{d}\| < \varepsilon_1$  while  $\|\Delta\tilde{\mathbf{a}}\| < \varepsilon_2$ , where  $\varepsilon_i$  are numerical tolerance parameters. This problem can be solved efficiently by, for example, ridge regression using the  $\ell_2$ -norm.

### 3.6. Reference Structures

Extracting FCs requires a set of structures with reference forces that are most commonly obtained from DFT calculations. The compilation of these structures depends on the intended usage. In the present context, we are usually faced with FC extraction in one of the following situations.

1. FCs to be used in lattice dynamics theory (phonon frequencies, lifetimes, etc. and derived properties such as harmonic free energy and thermal conductivity)
2. FCs to be used as an interatomic potential in, for example, MD simulations for sampling strongly anharmonic PESs
3. effective lower-order FCs directly fitted to a MD trajectory

In the last case, structure selection is trivial as the MD trajectory naturally delivers the structures of interest. The other two scenarios require, however, more attention.

In case (i), often only second- and third-order FCs are of interest. For this purpose, we have found “rattled” structures to work well, which can be obtained by imposing displacements drawn from a Gaussian distribution. Displacement amplitudes of  $\approx 0.01$ – $0.05$  Å commonly yield accurate force constants, which is comparable to the default displacement amplitude of 0.01 Å used, for example, by PHONOPY.

When using the rattle approach, we have found that it can become difficult to untangle the force contribution if higher orders are contributing to the forces. This leads to higher-order contributions effectively being included in the fitted force constants thus reducing their accuracy. Therefore, even if the goal is to extract only second or third-order FCs, it is beneficial to include terms corresponding to the respective higher order (third or fourth-order) during FCs extraction.

In case (ii), it is more difficult to produce good training structures without any prior knowledge of the PES. When constructing reference structures for a fourth, sixth, or even higher-order model, rather large displacements must be included in training set. Here, the rattle method with standard deviations of  $\gtrsim 0.1$  Å will fail as it commonly leads to some very short interatomic distances with huge repulsive forces. This can be overcome using a Monte Carlo (MC) approach, which aims to produce large random displacement while preserving interatomic distances. A randomly selected atom is displaced by a small amount and the new position is accepted with probability

$$P = \frac{1}{2} \left[ \operatorname{erf} \left( \frac{d_{\text{min}} - d_{\text{threshold}}}{d_{\text{width}}} \right) + 1 \right]$$



where  $d_{\min}$  is the new minimum distance to any another atom,  $d_{\text{threshold}}$  is a user defined threshold for interatomic distances, and  $d_{\text{width}}$  acts as an effective temperature typically of order  $\approx 0.1\text{\AA}$ .

There is, however, a more elegant and physically sensible approach. Using some rough estimate for the second-order FCs, one can obtain a set of normal modes, which can be subsequently randomly populated with an average energy of  $k_B T/2$  to generate a physically sensible displacement pattern<sup>[32]</sup>

$$u_i^\alpha = \sqrt{\frac{2k_B T}{m_i}} \sum_s \frac{1}{\omega_s} W_{is}^\alpha \sqrt{-\ln Q_s} \cos(2\pi U_s)$$

where  $W$  is the polarization vector of mode  $s$ , and  $Q$  and  $U$  are uniform random numbers between zero and one. This approach can be used to generate displacements corresponding to specific temperatures as shown, for example, in ref. [33]. The initial second-order FCs can, for example, be obtained from a fit to a minimal set of rattled structures with a small displacement amplitude.

These two methods allow one to generate sensible training structures with large displacements with minimal knowledge of the reference PES. While we have found these methods to work very well in most cases, it is of course also possible to improve models iteratively.

## 4. Workflow

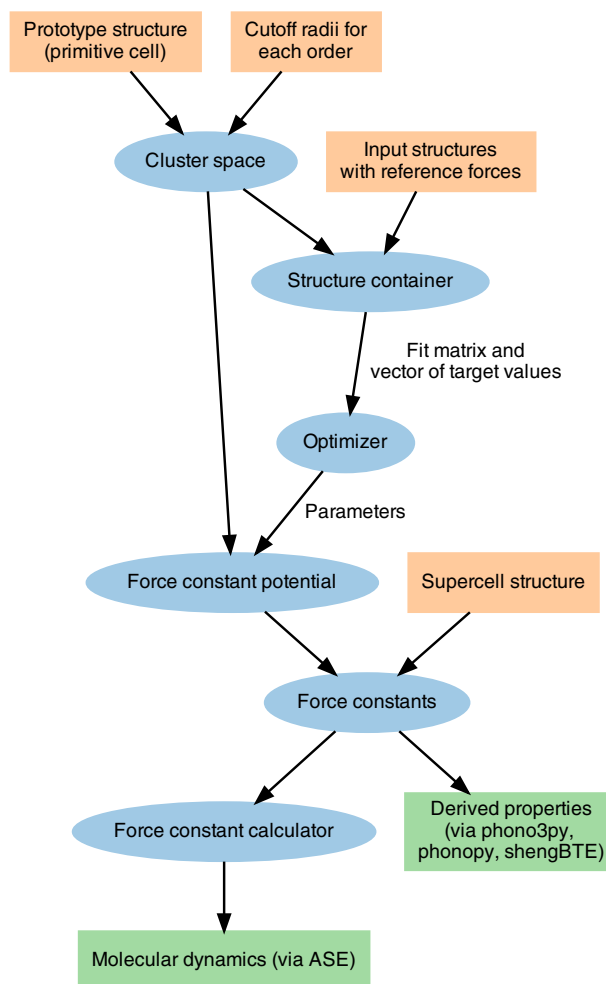
The approach outlined above has been implemented in the Python package HIPHIVE. The latter also provides functionality for performing related tasks such as fitting and validation, can be interfaced with other Python packages such as, for example, PHONOPY<sup>[34]</sup> and SCIKIT-LEARN.<sup>[19]</sup> The following subsections described the key steps involved in creating a force constant potential (FCP) (Figure 3).

### 4.1. Cluster Space

To build an FCP, one must create a cluster space object based on a prototype structure and a set of cutoffs, which specify the maximum interaction range considered for each order. As described in Section 3, the cluster space compiles the information needed to completely specify the clusters and eigentensors of any supercell based on a certain primitive cell. The cluster space is associated with a number of free parameters, which can be extracted by fitting to forces obtained by pseudo-random displacements for a number of supercell structures.

### 4.2. Structure Container

In order to fit the parameters, one requires a set of reference forces and displacements for some structures. The reference structures should span the configuration space of interest and for computational reasons, it is desirable to use as few structures as possible (see Section 3.6).



**Figure 3.** Overview of HIPHIVE workflow. Square orange nodes represent data supplied by a user. Blue ellipses represent HIPHIVE objects. Green squares represent output data that can be used directly or processed further using other packages or programs.

HIPHIVE can handle any combination of structures as long as they share the same equivalent primitive cell. For convenience, all configurations available for training and validation are compiled into a structure container. Upon addition of a structure, its corresponding sensing matrix is constructed as described in Section 3. The structure container then provides functionality for selecting subsets of structures in the form of sensing matrices and target forces suitable for training and validation.

### 4.3. Training and Validation

Having constructed cluster space and structure container, one can train the associated parameters. This involves solving a set of linear equations, which can be readily achieved by a number of linear regression techniques. The equation system can be over or underdetermined and in both cases some form of regularization is useful since the solution is often sparse and/or the data contains noise originating from the input data or the truncation

of the cluster space. The validity of the solution can be assessed by cross validation techniques including, e.g., shuffle-split or k-fold.

HIPHIVE supports a number of popular regularized training and validation techniques via the SCIKIT-LEARN machine learning library. This includes, for example, methods such as least absolute shrinkage and selection operator (LASSO), automatic relevance detection regression (ARDR), singular-value decomposition, elastic net, Bayesian-ridge regression, and recursive feature elimination. There is also functionality for generating ensembles of models. Further training/validation protocols can be readily implemented thanks to the modular structure of the approach.

#### 4.4. Force Constant Potential

Once the free parameters have been obtained, the cluster space can be transformed into an FCP. During this step, it is also possible to enforce rotational sum rules. The FCP enables calculating the FCs in any supercell compatible with the original primitive cell.

The final FCs can be analyzed using PHONOPY to obtain, for example, phonon dispersions or thermodynamic quantities in the (quasi-)harmonic approximation. It is also possible to compute transport properties using, for example, PHONO3PY or SHENGTE or to carry out MD simulations via an ASE calculator.

## 5. Applications

### 5.1. Phonon Dispersion of Monolayer-MoS<sub>2</sub>

2D materials such as graphene are attracting a lot of interest due their exceptional properties. As a result of their dimensionality, they exhibit a quadratic dispersion of one of the transverse acoustic modes near the  $\Gamma$ -point.<sup>[25]</sup> This is in contrast to (3D) bulk materials, for which all acoustic branches exhibit a linear dispersion in the center of the Brillouin zone. It has been shown that in order for this behavior to be captured correctly, the FCs must fulfill crystal symmetry, translational invariance, as well as rotational invariance conditions.<sup>[25]</sup> This provides an opportunity for demonstrating the impact of the rotational sum rules on the phonon dispersion.

Here, we consider a monolayer of the transition metal dichalcogenide (TMD) MoS<sub>2</sub>, which belongs to space group  $P\bar{6}m2$  (International Tables of Crystallography number 187). Input configurations were generated by imposing random displacements on ideal supercells comprising up to 300 atoms (equivalent to  $10 \times 10 \times 1$  unit cells). The average displacement amplitudes for these configurations were approximately  $0.008 \text{ \AA}$  leading to average forces of  $170 \text{ meV \AA}^{-1}$  and maximum forces of about  $1.1 \text{ eV \AA}^{-1}$ . Reference forces were obtained from DFT calculations using the projector augmented wave method<sup>[35,36]</sup> as implemented in VASP<sup>[37]</sup> and the vdW-DF-cx method, which combines semi-local exchange with nonlocal correlation.<sup>[38–41]</sup> The latter has been previously shown to perform very well with regard to the description of both structural and thermal properties of TMDs.<sup>[42]</sup> The Brillouin zone was sampled using Monkhorst-

Pack  $k$ -point grids equivalent to a  $16 \times 16 \times 1$  mesh relative to the primitive cell, except for the 300-atom cells, which were sampled using only the  $\Gamma$ -point. The plane-wave energy cutoff was set to 260 eV, a finer grid was employed for evaluating the forces, and the reciprocal projection scheme was used throughout.

FCPs were constructed using second-order cutoffs of up to  $15 \text{ \AA}$  and including third-order terms up to a range of  $3.0 \text{ \AA}$ . The latter procedure was found to stabilize the convergence of the second-order terms. For the largest supercell (300 atoms) and cutoff radius  $15 \text{ \AA}$ , one obtains 319 parameters. Parameters were optimized by conventional least-squares regression. For validation, the phonon dispersion was also computed using PHONOPY.<sup>[2]</sup>

The phonon dispersion obtained using HIPHIVE with all invariance conditions imposed is virtually indistinguishable from the one generated by PHONOPY (Figure 4a). Here, the lowermost transverse acoustic branch clearly exhibits a quadratic dispersion. If the rotational sum rules are deactivated, the dispersion is almost unchanged save for the emergence of a small imaginary pocket in the immediate vicinity of the Brillouin center (Figure 4b), and thus does *not* yield a quadratic dispersion. As noted in ref. [25], this seemingly small error can have a pronounced effect whenever the acoustic modes contribute substantially to a property, as in the case of, for example, the thermal conductivity.

This example also enables us to illustrate the effect of supercell size on the phonon dispersion and thus effectively the range of the FCs (Figure 4c). Convergence is achieved at a cutoff of just over  $9 \text{ \AA}$  equivalent to a supercell size of  $6 \times 6 \times 1$  (108 atoms).

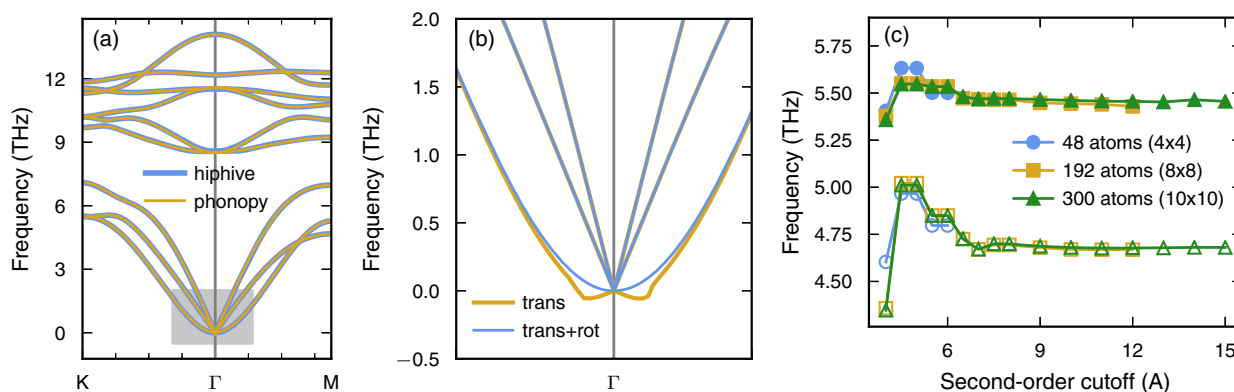
### 5.2. Thermal Conductivity of Monolayer-MoS<sub>2</sub>

Since modeling the thermal conductivity requires both accurate second- and third-order FCs, it thereby provides a viable test for the extraction of the higher-order FCs. The in-plane thermal conductivity of MoS<sub>2</sub> has been studied extensively both experimentally<sup>[43–45]</sup> and computationally,<sup>[42,46,47]</sup> and thus constitutes a good test case.

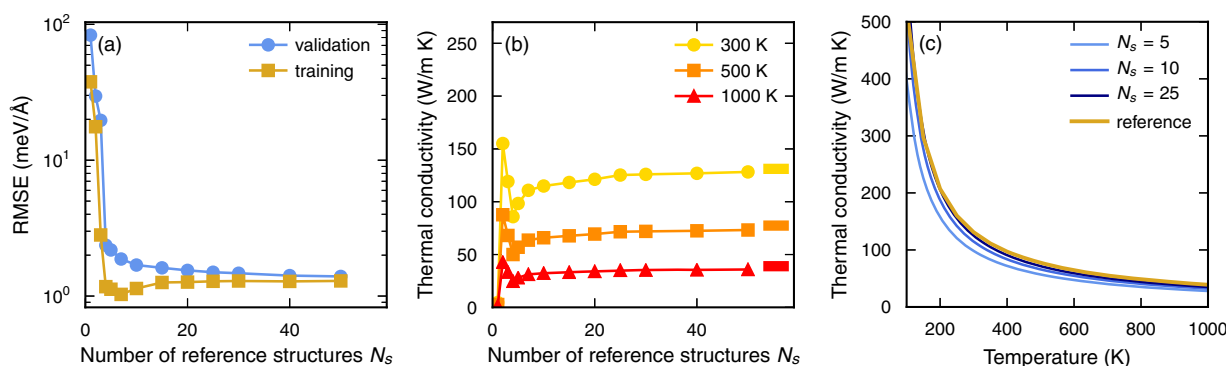
We used 192-atom ( $8 \times 8 \times 1$ ) supercells and generated structures for reference force calculations by rattling using an average displacement of  $0.048 \text{ \AA}$ . Reference DFT calculations were carried out using the same computational parameters as for the ideal monolayer, yielding an average (maximum) force of  $993 \text{ meV \AA}^{-1}$  ( $4.84 \text{ eV \AA}^{-1}$ ).

The cluster space was constructed using cutoffs of  $10.0$ ,  $6.5$ , and  $3.5 \text{ \AA}$  for second, third, and fourth-order terms, respectively, yielding 133 orbits, 2209 clusters, and 2004 parameters. FCP parameters were determined using recursive feature elimination with  $\ell_2$ -minimization. During training and validation, the set of reference forces was split at ratio of 4 to 1 into training and test sets, whereas the final FCPs were obtained by fitting against all available data.

The thermal conductivity was evaluated in the framework of phonon Boltzmann transport theory in the relaxation time approximation using the PHONO3PY<sup>[3]</sup> code. The integration of the Brillouin was carried out using the tetrahedron method and a  $40 \times 40 \times 1$   $q$ -point mesh. For clarity of the analysis, only phonon–phonon scattering was considered as a rate-limiting process. Reference calculations were carried out including pairs up to a distance of  $5.9 \text{ \AA}$ .



**Figure 4.** Vibrational properties of an ideal monolayer of MoS<sub>2</sub>. a) Phonon dispersions obtained using `HIHIVE` and `PHONOPY`, respectively. b) Effect of rotational invariance conditions on the dispersion in the gray shaded region in (a) near the  $\Gamma$ -point. c) Convergence of the lowermost modes at K (filled symbols) and M (open symbols) with supercell size and the cutoff for second-order terms.



**Figure 5.** Thermal conductivity of an ideal monolayer of MoS<sub>2</sub>. Convergence of a) RMSE and b,c) thermal conductivity with the number of structures in the reference data set. The thermal conductivity obtained using the conventional `PHONO3PY` approach is shown by the horizontal bars on the right hand side of panel (b) and the orange line in panel (c).

The root-mean-square error (RMSE) over the validation set quickly drops with the number of structures in the reference set (Figure 5), levelling out at a value of about  $1.5 \text{ meV } \text{\AA}^{-1}$  from about 15 structures onward. The thermal conductivity converges slightly more slowly as it requires about 20 to 25 configurations to achieve a converged result (Figure 5b). Most importantly, the converged values are in very good agreement with the reference data from `PHONO3PY` over the entire temperature range (Figure 5c), demonstrating the accuracy of the extracted FCs.

Crucially, this result was achieved at a fraction of the computational cost. The equivalent `PHONO3PY` calculations require 441 configurations for a cutoff of  $5.9 \text{ \AA}$  (571 configurations using a cutoff of  $6.5 \text{ \AA}$ ). In the case of `SHENGBTE`,<sup>[4]</sup> the numbers are somewhat smaller but still comparable in magnitude (324 structures for a cutoff of  $5.9 \text{ \AA}$ ). By contrast, `HIHIVE` required only 20 to 25 configurations to reach a converged result.

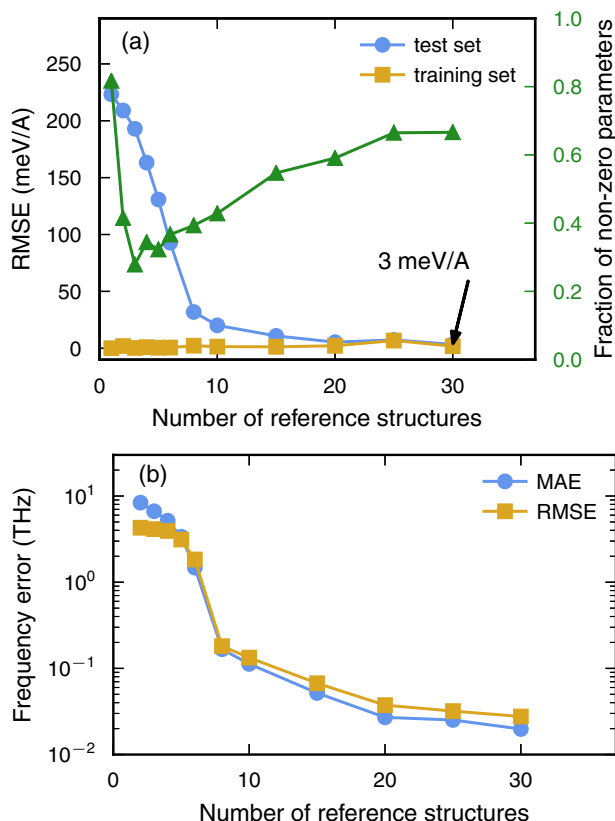
### 5.3. Sulfur Vacancy in Monolayer-MoS<sub>2</sub>

Defects in general and sulfur vacancies, in particular, are present in comparably large numbers in monolayers of MoS<sub>2</sub><sup>[48]</sup> and they can have a very pronounced effect on the thermal conductivity.<sup>[49]</sup>

To analyze this effect, one requires the second-order FCs matrix of a defect supercell, which enables one to evaluate the perturbation matrix connecting the ideal and defective systems. Due to the size of a defect configuration and its low symmetry, computing the FCs using the enumeration approach alluded to above commonly requires a large number of individual displacement/force calculations. Here, we demonstrate that this number can be considerably reduced using `HIHIVE` and advanced linear regression techniques.

The sulfur vacancy was described using a 192-atom supercell ( $8 \times 8 \times 1$  unit cells), which has been found above to yield well-converged FCs in the case of the ideal monolayer. Reference DFT calculations were carried out using the same computational parameters as for the ideal monolayer. Structures for reference force calculations were generated by rattling using an average displacement of  $0.016 \text{ \AA}$ , with an average (maximum) force of  $330 \text{ meV/\AA}$  ( $1.02 \text{ eV } \text{\AA}^{-1}$ ). The cluster space was constructed using cutoffs of  $9.0$  and  $3.0 \text{ \AA}$  for second and third-order terms, respectively, yielding 1670 orbits, 9376 clusters, and 13 030 parameters. FCP parameters were determined using ARDR with a  $\lambda$ -threshold of  $10^4$  and without standardization. During training and validation, the set of reference forces was split at a ratio of 4 to 1 into training and test sets, whereas the final FCPs were obtained by fitting against all available data.





**Figure 6.** Vibrational properties of a sulfur vacancy in a monolayer of MoS<sub>2</sub>. Convergence of the a) RMSE and b) the MAE and RMSE with the number of available structures.

The RMSE converges rather quickly with the number of structures (Figure 6a). Already for about 25–30 structures, the RMSE for the force components in the test set relative to the DFT reference calculations is only about 3 meV Å<sup>-1</sup>, where the PHONOPY analysis required 215 individual DFT calculations. This convergence behavior also translates to the mean absolute error (MAE) and RMSE over the frequencies (Figure 6b) as well as the overall phonon dispersion.

The regression produces sparse models as evident from the fraction of nonzero parameters, which remains below 70% (green triangles in Figure 6a). This is in line with physical intuition, according to which FCs ought to decay rather quickly with interaction distance and order.

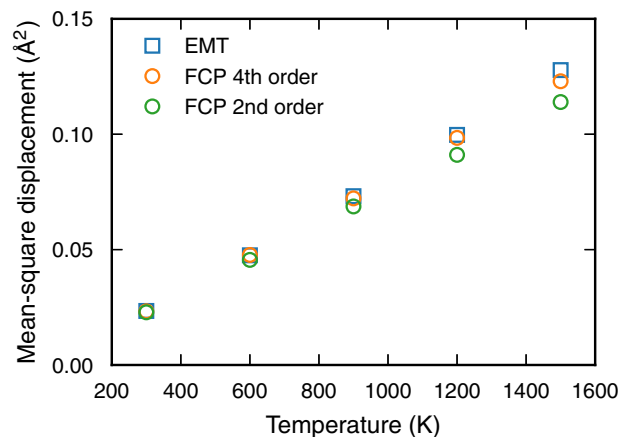
#### 5.4. Molecular Dynamics Simulations of Nickel

Anharmonic FCPs can in principle be sampled using MD simulations, which provide access to dynamic properties including, for example, dynamical structure factors, velocity auto-correlation functions, and free energies.<sup>[50]</sup> As illustrated in the following, HIPHIVE enables such simulations by providing an ASE calculator class that merely requires a set of FCs as input.

A fourth-order FCP was created for bulk Ni using cutoff radii of 5.0, 4.0, and 4.0 Å for second, third, and fourth-order terms, respectively, corresponding to 171 clusters in the unit cell. The

**Table 1.** Number of orbits and parameters (in brackets) by order and body.

	One-body		Two-body		Three-body		Four-body	
Second order	1	(1)	4	(12)	–	–	–	–
Third order	0	(0)	2	(8)	2	(14)	–	–
Fourth order	1	(2)	4	(29)	3	(75)	3	(40)



**Figure 7.** Atomic mean-square displacement in FCC Ni as a function of temperature as obtained within the harmonic approximation as well as from MD simulations based on the original potential (EMT) and a fourth-order Hamiltonian.

associated cluster space contained 20 unique orbits (5, 4, and 11 for second-, third-, and fourth-order, respectively) with interactions up to the fourth nearest neighbor for the pairs, leading to 119 free parameters.

Training data comprised five structures with 256 atoms (4 × 4 × 4 conventional unit cells) for a total of 3840 force components. The structures were obtained by applying displacements randomly drawn from a normal distribution, modified to avoid interatomic distances shorter than 2.3 Å. The resulting average atomic displacement were about 0.13 Å. Reference forces were obtained using an effective medium theory model as implemented in ASE.<sup>[20]</sup> The parameters of the model were trained by standard least-squares fitting as the system is heavily over-determined (Table 1).

The integration of the equations of motion was carried out using functionality provided by ASE while HIPHIVE was used to provide an interaction model in the form of an ASE calculator object. Simulations were carried out using an 864-atom supercell (6 × 6 × 6 conventional unit cells). The equations of motion were integrated for 50 ps using a time step of 5 fs at temperatures 300, 600, 900, 1200, and 1500 K using a Langevin thermostat.

The atomic mean-square displacements computed using the full fourth-order FCP as well as using only the second-order FCs are practically identical to those obtained using the effective medium theory model at low temperatures (Figure 7). At higher temperatures where anharmonic effects are more important, the second-order model yields a systematic error whereas the fourth-order model remains very accurate. Very close to the melting point even the fourth-order model starts to deviate from

the exact solution, which is due to even higher-order terms becoming relevant at these temperatures.

## 6. Conclusions and Outlook

Second and higher-order FC are fundamental to the description of the thermodynamics of materials. Here, we have introduced the HIPHIVE package that enables their efficient extraction from first-principles calculations using regression techniques with regularization. The implementation take advantage of symmetry and sum rules in order to constrain the number of DOFs and is computationally efficient. Potential applications have been illustrated by several examples including ideal and defective systems, phonon analysis as well as dynamic simulations.

The package is designed for integration in various workflows including, for example, applications in high-throughput calculations, as it can be readily interfaced with a large number of electronic structure codes via ASE and machine learning techniques via SCIKIT-LEARN. The flexibility of the interface also enables one to systematically explore the efficacy of different optimization algorithms for the construction of FCPs and the compilation of databases of such models for future use.

## Acknowledgements

This work was funded by the Knut and Alice Wallenberg Foundation, the Swedish Research Council, the Swedish Foundation for Strategic Research (SSF) as well as the Swedish National Infrastructure for Computing (SNIC). Computer time allocations by the SNIC at C3SE (Gothenburg), NSC (Linköping), and PDC (Stockholm) are gratefully acknowledged.

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

anharmonicity, force constants, machine learning, method, phonons

Received: November 22, 2018

Revised: January 10, 2019

Published online:

- [1] J. C. Thomas, A. V. der Ven, *Phys. Rev. B* **2013**, *88*, 214111.  
 [2] A. Togo, I. Tanaka, *Scr. Mater.* **2015**, *108*, 1.  
 [3] A. Togo, L. Chaput, I. Tanaka, *Phys. Rev. B* **2015**, *91*, 094306.  
 [4] W. Li, J. Carrete, N. A. Katcho, N. Mingo, *Comput. Phys. Commun.* **2014**, *185*, 1747.  
 [5] J. Carrete, B. Vermeersch, A. Katre, A. van Roekeghem, T. Wang, G. K. H. Madsen, N. Mingo, *Comput. Phys. Commun.* **2017**, *220*, 351.  
 [6] J. J. Plata, P. Nath, D. Usanmaz, J. Carrete, C. Toher, M. Jong, M. Asta, M. Fornari, M. B. Nardelli, S. Curtarolo, *npj Comput. Mater.* **2017**, *3*, 45.  
 [7] X. Gonze, F. Jollet, F. A. Araujo, D. Adams, B. Amadon, T. Applencourt, C. Audouze, J.-M. Beuken, J. Bieder, A. Bokhanchuk, E. Bousquet, F. Bruneval, D. Caliste, M. Côté, F. Dahm, F. Da Pieve, M. Delaveau, M. Di Gennaro, B. Dorado, C. Espejo, G. Geneste, L. Genovese, A. Gerossier, M. Giantomassi, Y. Gillet, D. Hamann, L. He, G. Jomard, J. L. Janssen, S. Le Roux et al., *Comput. Phys. Commun.* **2016**, *205*, 106.  
 [8] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. D. Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio Jr, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri et al., *J. Phys. Condens. Matter* **2017**, *29*, 465901.  
 [9] X. Gonze, J. P. Vigneron, *Phys. Rev. B* **1989**, *39*, 13120.  
 [10] K. Esfarjani, H. T. Stokes, *Phys. Rev. B* **2008**, *77*, 144112.  
 [11] T. Andersson, *Master's Thesis*, Chalmers University of Technology, **2012**.  
 [12] T. Tadano, Y. Gohda, S. Tsuneyuki, *J. Phys. Condens. Matter* **2014**, *26*, 225402.  
 [13] O. Hellman, I. A. Abrikosov, *Phys. Rev. B* **2013**, *88*, 144301.  
 [14] E. Candes, M. Wakin, *IEEE Signal Process. Mag.* **2008**, *25*, 21.  
 [15] F. Zhou, W. Nielson, Y. Xia, V. Ozoliņš, *Phys. Rev. Lett.* **2014**, *113*, 185501.  
 [16] T. Tadano, S. Tsuneyuki, *Phys. Rev. B* **2015**, *92*, 054301.  
 [17] F. Zhou, W. Nielson, Y. Xia, V. Ozoliņš, arXiv:1805.08903, **2018**.  
 [18] F. Zhou, B. Sadigh, D. Åberg, Y. Xia, V. Ozoliņš, arXiv:1805.08904, **2018**.  
 [19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.  
 [20] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Duřak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge et al., *J. Phys. Condens. Matter* **2017**, *29*, 273002.  
 [21] L. J. Nelson, G. L. W. Hart, F. Zhou, V. Ozoliņš, *Phys. Rev. B* **2013**, *87*, 035125.  
 [22] hiPhive – High-order force constants for the masses, <https://hiphive.materialsmodeling.org/> (accessed: January 2019).  
 [23] M. Born, K. Huang, *Dynamical Theory of Crystal Lattices*, Oxford University Press, Oxford **1954**.  
 [24] D. C. Wallace, *Thermodynamics of Crystals*, Dover, Mineola, New York **1998**.  
 [25] J. Carrete, W. Li, L. Lindsay, D. A. Broido, L. J. Gallego, N. Mingo, *Mater. Res. Lett.* **2016**, *4*, 204.  
 [26] D. C. Gazis, R. F. Wallis, *Phys. Rev.* **1966**, *151*, 578.  
 [27] S. K. Sarkar, S. Sengupta, *Phys. Status Solidi B* **1977**, *83*, 263.  
 [28] S. Wang, H. Zhang, X. Wu, *J. Phys. Condens. Matter* **2007**, *19*, 386233.  
 [29] L. Elcoro, J. Etxebarria, *Eur. J. Phys.* **2011**, *32*, 25.  
 [30] A. Togo, I. Tanaka, arXiv:1808.01590, **2018**.  
 [31] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, S. Roučka, A. Saboo, I. Fernando, S. Kula, R. Cimrman, A. Scopatz, *Peer J. Comput. Sci.* **2017**, *3*, e103.  
 [32] G. Eckold, *International Tables for Crystallography*, Vol. D, Wiley, New York **2004**, Sec. 2.1.  
 [33] D. Åberg, P. Erhart, V. Lordi, *Phys. Rev. B* **2013**, *88*, 045201.  
 [34] A. Togo, I. Tanaka, *Scr. Mater.* **2015**, *108*, 1.  
 [35] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.  
 [36] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758.  
 [37] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **1996**, *6*, 15.

- [38] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **2004**, *92*, 246401.
- [39] J. Klimeš, D. R. Bowler, A. Michaelides, *Phys. Rev. B* **2011**, *83*, 195131.
- [40] K. Berland, P. Hylgaard, *Phys. Rev. B* **2014**, *89*, 035412.
- [41] T. Björkman, *J. Chem. Phys.* **2014**, *141*, 074708.
- [42] D. O. Lindroth, P. Erhart, *Phys. Rev. B* **2016**, *94*, 115205.
- [43] J. Liu, G.-M. Choi, D. G. Cahill, *J. Appl. Phys.* **2014**, *116*, 233107.
- [44] X. Zhang, D. Sun, Y. Li, G.-H. Lee, X. Cui, D. Chenet, Y. You, T. F. Heinz, J. C. Hone, *ACS Appl. Mater. Interfaces* **2015**, *7*, 25923.
- [45] P. Jiang, X. Qian, X. Gu, R. Yang, *Adv. Mater.* **2017**, *29*, 1701068.
- [46] W. Li, J. Carrete, N. Mingo, *Appl. Phys. Lett.* **2013**, *103*, 253103.
- [47] B. Peng, H. Zhang, H. Shao, Y. Xu, X. Zhang, H. Zhu, *Ann. Phys.* **2016**, *528*, 504.
- [48] W. Zhou, X. Zou, S. Najmaei, Z. Liu, Y. Shi, J. Kong, J. Lou, P. M. Ajayan, B. I. Yakobson, J.-C. Idrobo, *Nano Lett.* **2013**, *13*, 2615.
- [49] A. Katre, J. Carrete, B. Dongre, G. K. Madsen, N. Mingo, *Phys. Rev. Lett.* **2017**, *119*, 075902.
- [50] dynasor – Correlations for everyone, <https://dynasor.materialsmodeling.org/> (accessed: January 2019).