

<https://doi.org/10.1038/s41524-024-01236-3>

# Automated all-functionals infrared and Raman spectra

Lorenzo Bastonero <sup>1</sup> ✉ & Nicola Marzari <sup>1,2,3</sup>

Infrared and Raman spectroscopies are ubiquitous techniques employed in many experimental laboratories, thanks to their fast and non-destructive nature able to capture materials' features as spectroscopic fingerprints. Nevertheless, these measurements frequently need theoretical and computational support in order to unambiguously decipher and assign complex spectra. Linear-response theory provides an effective way to obtain the higher-order derivatives needed, but its applicability to modern exchange-correlation functionals and pseudopotential formalism remains limited. Here, we devise an automated, open-source, user-friendly approach based on density-functional theory and the electric-enthalpy functional to allow seamless calculation from first principles of infrared absorption and reflectivity, together with zone-center phonons, static dielectric tensor, and Raman spectra. By employing a finite-displacement and finite-field approach, we allow for the use of any functional, as well as an efficient treatment of large low-symmetry structures. Additionally, we propose a simple scheme for efficiently sampling the Brillouin zone at different electric fields. To demonstrate the capabilities of the present approach, we study ferroelectric LiNbO<sub>3</sub> crystal as a paradigmatic example, and predict infrared and Raman spectra using various (semi)local, Hubbard corrected, and hybrid functionals. Our results also show how PBE0 and extended Hubbard functionals (PBEsol+U+V) yield for this case the best match in term of peak positions and intensities, respectively.

Vibrational spectroscopies are one of the most powerful, fast and reliable methods for materials' structure identification and characterization. Among the different techniques probing atomic motions, infrared (IR) absorption and Raman experiments are widespread in most experimental setups, thanks to their low cost and precision in identifying microscopic features. Despite these advantages, data interpretation for many cases calls for theoretical and computational support. Density-functional theory (DFT) is most often employed as a reliable framework for accurate predictions of such spectra, providing in principle exact estimates of the second- and third-order derivatives of the energy functional that are needed to predict these vibrational spectra. Over the past decade, many techniques have been developed to efficiently obtain non-resonant spectra of insulating crystals, such as linear-response theory<sup>1-5</sup> for infrared absorption, second-order response of the density matrix<sup>6</sup> and the  $2n + 1$  theorem<sup>7</sup> for Raman couplings. While being elegant and attractive, these methods come at the cost of rather complex code implementations, limiting in practice applications to selected semi-local functionals and often only to norm-conserving

pseudopotentials. This is especially the case for the third-order derivatives of the total energy, required for the Raman calculations. One common expedient for the latter is to resort to finite differences of the dielectric tensor upon atomic displacements, computed using density-functional perturbation theory (DFPT)<sup>8,9</sup>. However, this approach scales in general as  $6N_{at}$  with the number of atoms  $N_{at}$ , a burden in low-symmetry systems, and becomes out of reach for the non-linear optical susceptibility tensor, which has been shown to be fundamental to achieve good agreement for Raman intensities both in single crystals<sup>10</sup> and powder spectra<sup>11</sup>. An alternative framework for carrying out such derivatives is the finite numerical differentiation of forces and polarization upon the application of a homogeneous electric field<sup>7,12-15</sup>. With this technique, along with finite displacements<sup>16</sup>, one can compute straightforwardly all the quantities needed for IR and Raman spectra, as well as allowing to use any functional (Hubbard corrected, hybrids, ...), and pseudopotential formalism (norm-conserving, ultrasoft, PAW)<sup>17</sup> with little effort. Furthermore, the use of finite fields overcomes the expensive computation and poor scaling of the finite dielectric tensor approach, allowing

<sup>1</sup>U Bremen Excellence Chair, Bremen Center for Computational Materials Science, and MAPEX Center for Materials and Processes, University of Bremen, D-28359 Bremen, Germany. <sup>2</sup>Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. <sup>3</sup>Laboratory for Materials Simulations, Paul Scherrer Institut, 5232 Villigen PSI, Switzerland. ✉ e-mail: [lbastone@uni-bremen.de](mailto:lbastone@uni-bremen.de)

for the study of complex materials such as amorphous and defective structures<sup>18–20</sup>. In fact, with the present approach, one would need to perform typically only 12 ground-state calculations in the presence of a homogeneous electric field (see Eq. (19) and related discussion). While appealing, this avenue has seen scarce applicability, possibly due to the many steps and pitfalls related to the underlying algorithm, which can be highly non-trivial even for expert users.

A key step forward is to streamline all of these operations through a modern workflow manager, adhering in passing to the FAIR<sup>21</sup> principles for data sharing, while being able of automatically submitting, parsing, and processing the outputs generated by multiple calculations. In this study, we present a comprehensive formulation for a unified finite-displacement and finite-field approach, along with efficient Brillouin zone sampling, and encode this formalism in a self-contained Python package that operates within the AiiDA framework<sup>22,23</sup>, offering a scalable computational infrastructure for automated and reproducible workflows and data provenance. To showcase these capabilities, we demonstrate its effectiveness in predicting the infrared and Raman spectra of LiNbO<sub>3</sub> using seven different functionals: LDA, PBE, PBEsol, PBEsol+U, PBEsol+U+V, PBE0, and HSE06.

## Results

### The finite-displacement and finite-field method

Infrared and Raman spectroscopy are investigation methods probing the atomic vibrations of materials. In both techniques, a source of light shines on the sample, and the incident photons are either absorbed or scattered by the material. In IR experiments, peaks of the absorption spectra correspond to collective atomic motions in resonance with that frequency, corresponding to long-wavelength polar optical phonons. Raman spectroscopy records instead a scattered frequency  $\omega_S$ ; in this case, the excited lattice vibration corresponds to the shift  $\pm \omega_\nu = \omega_L - \omega_S$  from the incoming laser frequency  $\omega_L$ , where the plus (minus) sign refers to (anti-)Stokes process for which a phonon is created (annihilated). Far from resonance the two phenomena can be captured by the change in polarization  $\mathbf{P}$  and the change in susceptibility  $\chi$ , for IR and Raman respectively, leading to different selection rules<sup>24</sup> for the activated phonon modes which ultimately provide a comprehensive fingerprint of the material. The former mechanism is coupled to the Born effective charge  $\mathbf{Z}_I^{*5,25}$  of an atom  $I$  in the unit cell, and the latter instead to the Raman tensor defined as  $d\chi/d\tau_I$  in the Placzek approximation<sup>26–28</sup>, where  $\chi$  is the electronic susceptibility tensor and  $\tau_I$  the atomic displacement. In terms of these quantities, the absorption intensities in the infrared regime of a normal mode  $\nu$  with angular frequency  $\omega_\nu$  are:

$$I_{\text{IR}}^\nu = \frac{|\bar{\mathbf{Z}}^\nu|^2}{\omega_\nu}, \quad (1)$$

associated to the imaginary part of the complex dielectric function<sup>5,18,29</sup>

$$\epsilon_{ij}(\omega) = \epsilon_{ij}^\infty + \frac{4\pi}{\Omega} \sum_\nu \frac{\bar{Z}_i^\nu \bar{Z}_j^\nu}{\omega_\nu^2 - \omega^2}, \quad (2)$$

which is connected to other measurable quantities, such as the energy loss function and reflectivity spectrum<sup>5,18,29</sup> ( $\epsilon_{ij}^\infty$  is the high-frequency dielectric tensor). The amplitudes are defined via the polarization vector

$$\bar{Z}_i^\nu = \sum_{I,k} Z_{I,ik}^* \frac{e_{I,k}^\nu}{\sqrt{m_I}}, \quad (3)$$

where  $e_{I,k}^\nu$  is the phonon eigenvector of the mode  $\nu$ , describing the collective dynamics of atoms (index  $I$ ) in space (index  $k$ ). In particular, the relationship to the reflectivity  $\mathcal{R}(\omega)$  of optical waves oriented perpendicularly to the surface and having their electric field

along a crystal optical axis  $\mathbf{q}$  is

$$\mathcal{R}(\omega) = \left| \frac{\sqrt{\epsilon_{\hat{\mathbf{q}}}(\omega) - 1}}{\sqrt{\epsilon_{\hat{\mathbf{q}}}(\omega) + 1}} \right|, \quad (4)$$

where  $\epsilon_{\hat{\mathbf{q}}}(\omega) = \hat{\mathbf{q}} \cdot \boldsymbol{\epsilon}(\omega) \cdot \hat{\mathbf{q}}$  and  $\hat{\mathbf{q}}$  is the versor of  $\mathbf{q}$ . The Raman scattering intensities are instead, in the Placzek approximation<sup>26,27</sup>

$$I_{\text{Raman}}^\nu \propto \frac{\omega_S^4}{\omega_\nu} (n_\nu(T) + 1) |\mathbf{e}_S \cdot \boldsymbol{\alpha}^\nu \cdot \mathbf{e}_L|^2 \quad (5)$$

for Stokes processes for which  $\omega_S = \omega_L - \omega_\nu$ , and where  $\mathbf{e}_{L,S}$  represent the photon polarization vectors, orthogonal to their propagation vectors  $\mathbf{k}_{L,S}$ , and  $n_\nu(T)$  is the Bose-Einstein occupation function;  $\boldsymbol{\alpha}^\nu$  is the Raman susceptibility tensor

$$\alpha_{ij}^\nu = \sqrt{\Omega} \sum_{I,k} \frac{\partial \chi_{ij}}{\partial \tau_{I,k}} \frac{e_{I,k}^\nu}{\sqrt{m_I}}, \quad (6)$$

where  $m_I$  is the mass of the atom  $I$  and  $\Omega$  the volume of the unit cell. In anti-Stokes processes, the prefactor  $(n_\nu(T) + 1)$  must be replaced with  $n_\nu(T)$ , and  $\omega_S = \omega_L + \omega_\nu$ . The ratio of Stokes and anti-Stokes peak intensities is therefore temperature dependent, and, interestingly, can be used as a microscopic thermometer<sup>30</sup>. In the harmonic approximation, the phonon normal modes are obtained from the interatomic force constants (IFCs)

$$\Phi_{IJ,ij}(a, b) = \frac{\partial^2 E}{\partial \tau_{I,i}^a \partial \tau_{J,j}^b}, \quad (7)$$

where  $E$  is the potential-energy surface, and  $a$  and  $b$  indicate the periodic images with respect to the origin (i.e., the unit cell). The IFCs can be written in terms of first-order derivative of atomic forces  $\mathbf{F}_I$  with respect to atomic displacements, and thus can also be obtained using a set of small displacement configurations in supercells, as discussed in the literature<sup>5,16</sup>, rather than the more efficient density-functional perturbation theory approach. One can exploit periodic-boundary conditions to conveniently Fourier transform the IFCs, and rewrite the latter as a function of the wavevector  $\mathbf{q}$ ; the mass-scaled of this Fourier transform is the dynamical matrix describing the harmonic ion dynamics

$$D_{IJ,ij}(\mathbf{q}) = \frac{1}{\sqrt{m_I m_J}} \sum_b \Phi_{IJ,ij}(0, b) e^{i\mathbf{q} \cdot (\mathbf{R}_J^b - \mathbf{R}_I^0)}, \quad (8)$$

where  $\mathbf{R}_J^b$  is the equilibrium position of atom  $J$  in the  $l$ -th unit cell image. The diagonalization of the dynamical matrix leads to the phonon frequencies and eigenvectors

$$\sum_{Jj} D_{IJ,ij}(\mathbf{q}) e_{J,j}^\nu(\mathbf{q}) = [\omega_\nu(\mathbf{q})]^2 e_{I,i}^\nu(\mathbf{q}). \quad (9)$$

IR and Raman spectroscopies utilize photons which carry a negligible momentum  $\mathbf{q} \approx \mathbf{0}$ , thus only probing phonons in the vicinity of the zone center. For exactly  $\mathbf{q} = \mathbf{0}$ , the phase factor in Eq. (8) becomes the unit, allowing for the calculation of the IFCs via finite-displacements in the primitive cell only, as the displacement patterns are defined uniquely in such cell. Nevertheless, the evaluation of the dynamical matrix in the long-wavelength limit, i.e., when  $\mathbf{q} \rightarrow \mathbf{0}$ , in the case of polar materials requires a proper treatment of its non-analytical behavior<sup>4,5</sup>, accounting for the coupling between phonon polarization densities and electric fields, that adds an extra cost for displacements that create a macroscopic electric field, absent in periodic-boundary calculations of the primitive cell only. The analytical shape of this coupling depends on the dimensionality of the

material<sup>4,5,31,32</sup> and on the multipolar order<sup>33</sup>. At the dipolar level, the dynamical matrix in such limit reads

$$D_{ll,ij}(\mathbf{q} \rightarrow \mathbf{0}) = D_{ll,ij}(\mathbf{q} = \mathbf{0}) + D_{ll,ij}^{NA}(\mathbf{q}) \quad (10)$$

where the non-analytical contribution is

$$D_{ll,ij}^{NA}(\mathbf{q}) = \frac{1}{\sqrt{m_l m_j}} \frac{W^d(\mathbf{q})}{\Omega} \left[ \left( \sum_l q_l Z_{l,li}^* \right) \left( \sum_{l'} q_{l'} Z_{l',lj}^* \right) \right], \quad (11)$$

valid for  $\mathbf{q} \rightarrow \mathbf{0}$  and where  $W^d$  is the Coulomb screening whose shape depends on the dimensionality  $d$  (see e.g., refs. <sup>31,32</sup>). The directional dependence of the polar optical phonon wavevector  $\mathbf{q}$  of the dynamical matrix gives rise to a shift in frequency and a change in eigenvector compared to the transverse optical modes (i.e.,  $\bar{\mathbf{Z}}(\mathbf{q}) \cdot \mathbf{q} = 0$ ), giving rise to the well known splitting between the longitudinal optical (LO) and transverse optical (TO) modes<sup>25</sup>; this shift is largest when  $\bar{\mathbf{Z}}(\mathbf{q}) \parallel \mathbf{q}$ . Analogously to the IFCs, the Born effective charges and Raman tensors can also be defined as derivatives of forces with respect to a macroscopic electric field  $\mathcal{E}$ <sup>4</sup>. For the former we have:

$$Z_{l,ij}^* = \frac{\partial F_{l,j}}{\partial \mathcal{E}_i} \quad (12)$$

while for the latter, the Raman tensor reads:

$$\frac{\partial \chi_{ij}}{\partial \tau_{l,k}} = \frac{1}{\Omega} \frac{\partial^2 F_{l,k}}{\partial \mathcal{E}_i \partial \mathcal{E}_j} \quad (13)$$

In a finite-difference scheme, this means that a small homogeneous electric field must be applied in order to carry out the numerical differentiation. In periodic-boundary conditions, this can be accomplished by extending the energy functional  $E[\psi]$  to include a term<sup>12,13,34</sup> accounting for the coupling of the electric field and the total polarization of the system (often referred as *electric-enthalpy* functional):

$$\mathcal{F}[\psi, \mathcal{E}] = E[\psi] - \Omega \mathcal{E} \cdot [\mathbf{P}^{\text{ion}} + \mathbf{P}^{\text{el}}[\psi]], \quad (14)$$

where  $\mathbf{P}^{\text{ion}}$  and  $\mathbf{P}^{\text{el}}$  are the ionic and electronic polarizations, respectively; the latter being described by the modern theory of polarization<sup>35,36</sup>. Such functional, while not bounded from below, admits local minima representing self-consistent stationary solutions<sup>12,13</sup>, if  $\mathcal{E}$  is properly chosen. We note that in many implementations (such as the present one) the electric enthalpy functional is not directly minimized<sup>37,38</sup>, but it is recast in a non-local operator which couples neighboring  $k$ -points and which is solved self-consistently among all  $k$ -points (see refs. <sup>39-41</sup> for details). The self-consistent solution of the functional guarantees the simultaneous evaluation of Hellmann-Feynman forces, and electronic polarization, at convergence. As a consequence, the electronic polarization (that comes for free) can be exploited to extract its first and second-order derivatives with respect to an electric field<sup>7,13,18</sup>. These correspond respectively to the high-frequency dielectric tensor  $\epsilon^\infty$ , used to calculate the screened Coulomb interaction in Eq. (11)<sup>4,31,32</sup>, the full complex dielectric tensor (Eq. (4)), and the non-linear optical susceptibility  $\chi^{(2)}$ , key in non-centrosymmetric crystals where the Fröhlich contribution (see Eq. (17)) to the Raman tensor could be sizable<sup>10,11</sup>. Their relationship with the electronic polarization derivatives is as follows:

$$\epsilon_{ij}^\infty = \delta_{ij} + \frac{4\pi}{\Omega} \frac{\partial P_i^{\text{el}}}{\partial \mathcal{E}_j}, \quad (15)$$

$$\chi_{ijk}^{(2)} = \frac{2\pi}{\Omega} \frac{\partial^2 P_i^{\text{el}}}{\partial \mathcal{E}_j \partial \mathcal{E}_k}. \quad (16)$$

For non-vanishing  $\chi^{(2)}$ , the Raman tensor in 3D solids is modified as follows<sup>10,11,42</sup>:

$$\frac{\partial \chi_{ij}}{\partial \tau_{l,k}}(\mathbf{q}) = \frac{\partial \chi_{ij}}{\partial \tau_{l,k}} - \frac{8\pi}{\Omega} \frac{\sum_l q_l Z_{l,lk}^*}{\sum_{s,s'} q_s \epsilon_{ss'}^\infty q_{s'}} \sum_l \chi_{ijl}^{(2)} q_l, \quad (17)$$

which gives a second  $\mathbf{q}$  dependence contribution to the Raman spectra intensities (the first being given by the change in phonon eigenvectors and frequencies).

### Numerical differentiation

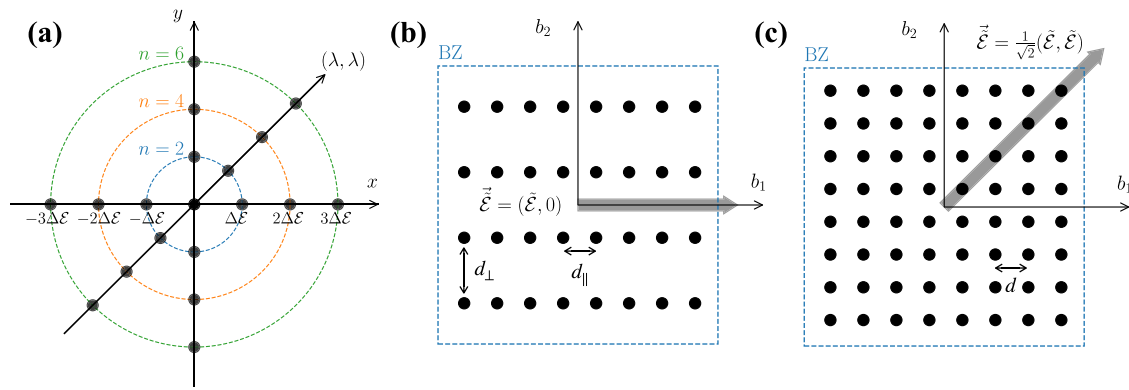
The possibility of consistently describing homogenous electric fields allows to access any derivative in atomic displacements and electric field, while being easily applicable to advanced functionals and pseudopotential formalisms<sup>15,17</sup>. We calculate here the tensors for infrared and Raman cross-sections using a central derivative formulation. These derivatives are performed around  $\mathcal{E} = \mathbf{0}$ , which resembles the common experimental setup where an external electromagnetic source is negligible, either natural or artificial. Indicating (by dropping the indices) as  $A$  either a force or the electronic polarization, we calculate its  $m$ -th derivatives with a numerical accuracy of  $n$ -th order via the discretization formula:

$$\frac{\partial^m A}{\partial \mathcal{E}_i^m} = \frac{1}{\Delta \mathcal{E}_i^m} \sum_{l=-n/2}^{n/2} c_l^m A(l \cdot \Delta \mathcal{E}_i) + \mathcal{O}(\Delta \mathcal{E}_i^n), \quad (18)$$

where  $\Delta \mathcal{E}_i > 0$  is the small electric field, which we refer to as the finite difference step, along the  $i$ -th Cartesian direction, and  $c_l^m$  are the central derivative coefficients that can be automatically found using the Fornberg's algorithm<sup>43</sup>. The discretization is performed using uniform steps such that  $l = -\frac{n}{2}, \dots, 0, \dots, \frac{n}{2}$  is always an integer number and the applied fields are evenly spaced (see Fig. 1a). This is at variance with the approach of calculating gradients for Bravais lattices with finite differences of ref. <sup>44</sup>, to use standard expressions for the higher-order and mixed derivatives. For mixed second-order derivatives of  $A$ , i.e., for the Raman and non-linear susceptibility tensors, we exploit a formula devised in ref. <sup>18</sup> to reduce the number of calculations: we assign  $\mathcal{E}_i = \mathcal{E}_j = \lambda$  and express the mixed derivative as:

$$\frac{\partial^2 A}{\partial \mathcal{E}_i \partial \mathcal{E}_j} = \frac{1}{2} \left[ \frac{\partial^2 A}{\partial \lambda^2} - \frac{\partial^2 A}{\partial \mathcal{E}_i^2} - \frac{\partial^2 A}{\partial \mathcal{E}_j^2} \right], \quad (19)$$

where each second-order derivative is evaluated through Eq. (18); this can be visualized schematically in Fig. 1(a). For example, using a 2nd-order numerical accuracy, we would need only 12 self-consistent field (SCF) calculations with a non-zero electric ( $\Delta \mathcal{E}$ -SCF) field to determine all the tensors' components needed both for IR and Raman spectra, even in low-symmetry systems such as amorphous materials<sup>18,20</sup>; this number can be reduced if one exploits the symmetries of the material. In fact, there could exist symmetries belonging to the point group of the crystal that transform an electric field direction into another. For example, if the crystal satisfies inversion symmetry, this is sufficient to reduce by a half the number of  $\Delta \mathcal{E}$ -SCF calculations needed, since forces and electronic polarization produced by  $\Delta \mathcal{E}_i$  are connected by a similarity transformation to the one produced by  $-\Delta \mathcal{E}_i$ ; keeping the previous numerical example, we would need then just 6  $\Delta \mathcal{E}$ -SCF calculations. For highly-symmetric crystals, such as cubic Si, only two independent electric field directions must be evaluated (e.g., along Z and the mixed direction XZ); in this case, the number of calculations reduces further to only 2. In general, the number of  $\Delta \mathcal{E}$ -SCF calculations needed scales with the numerical accuracy  $n$  of the central formula as  $\mathcal{N}_n = 6 \cdot n$ , not accounting for symmetries. Interestingly, it can be noticed from Fig. 1 that the evaluation of the numerical derivatives at  $n > 2$  provides automatically derivatives at lower accuracy  $n$ ; this is helpful to



**Fig. 1 | Illustration of the numerical derivative discretization and of the application of the directional sampling scheme applied to a simple lattice.** **a** Schematic illustration of the discretization needed in Eqs. (18) and (19) for derivatives with respect to the electric field in the directions  $i = \{x, y\}$  and for mixed derivatives along  $(\lambda, \lambda)$ . The black dots represent the applied electric fields used in the SCF calculations to evaluate the electric-field-dependent  $A$  (forces or polarization) used in Eqs. (18) and (19), while the dashed colored circles are associated to the accuracy  $n$  achieved in the numerical differentiation (here shown up to order  $n = 6$ ). **b, c** Square reciprocal lattice showing possible  $k$ -point samplings (black dots) of the BZ generated via Eqs. (23)–(24) for the electric field  $\mathcal{E}$  shown in crystal coordinates. **b** shows the case when the electric field is parallel to a lattice vector; in this scenario, more  $k$ -points are

generated along such direction, by using  $d_{\parallel}$ , where the induced polarization is expected to be larger. The increased number of  $k$ -points in such direction improves the convergence of the calculated polarization via Eq. (21). A more accurate polarization can then be obtained by decreasing the value of  $d_{\parallel}$ , which in this case turns to linearly increasing the number of  $k$ -points. **c** shows the special case  $\tilde{\mathcal{E}}_1 = \tilde{\mathcal{E}}_2 \equiv \tilde{\mathcal{E}}/\sqrt{2}$ , where the induced polarization is expected along the mixed direction. Since the polarization is a sum of the polarization projected along  $b_1$  and  $b_2$ , Eq. (22), then the  $k$ -points are increased along both directions by using  $d_{\parallel}$ . The generated mesh is equivalent to a uniform mesh defined by a single  $k$ -point distance  $d \equiv d_{\parallel}$ . In this case, when more accurate values are desired, the number of  $k$ -points will scale quadratically when decreasing  $d_{\parallel}$ .

check the convergence *a posteriori* with respect to the finite step size  $\Delta\mathcal{E}$ . Some benchmarks and illustrations on a prototypical material are provided in the Supplementary Discussion.

While this approach is advantageous, we also highlight that the choice of the field magnitude is restricted. As already stated before, the electric-enthalpy functional (14) has no global minimum, and long-lived metastable states exist only below a certain critical field  $\mathcal{E}^c$ , where Zener tunneling is suppressed<sup>12,34</sup>. An estimate of such critical value can be performed using<sup>12</sup>:

$$e|\mathcal{E}^c \cdot \mathbf{a}_i| \simeq E_{\text{gap}}/N_i \quad (20)$$

where  $e$  is the electric charge,  $\mathbf{a}_i$  the lattice vectors defining the unit cell,  $E_{\text{gap}}$  the electronic band gap, and  $N_i$  the number of  $k$ -points sampling the Brillouin zone in the  $i$ -th crystal direction. For semiconductors with small band gaps this critical field can be very small, limiting in practice the convergence due to numerical noise. In these circumstances, a balance must be found; nevertheless, for the majority of cases this does not represent an actual limitation. The last subtlety lies in the choice of the logarithm branch when calculating the Berry-phase polarization; a detailed explanation of the issue along with practical solutions can be found in the Supplementary Methods.

**Directional sampling**

An additional subtle issue related to the Berry phase formalism used in the electric-enthalpy formulation is the dense sampling of the Brillouin zone required to achieve well converged values of the polarization. In fact, the polarization is a  $k$ -space integral of a Berry connection<sup>35,36</sup>, and the discretized formula for the gradient converges slowly with respect to the  $k$ -point distance<sup>14</sup> (see also Supplementary Discussion). One can express the electronic polarization using the string-averaged discretized Berry phase formulation<sup>35,36</sup> along the direction of a reciprocal lattice vector  $\mathbf{b}_i$  as:

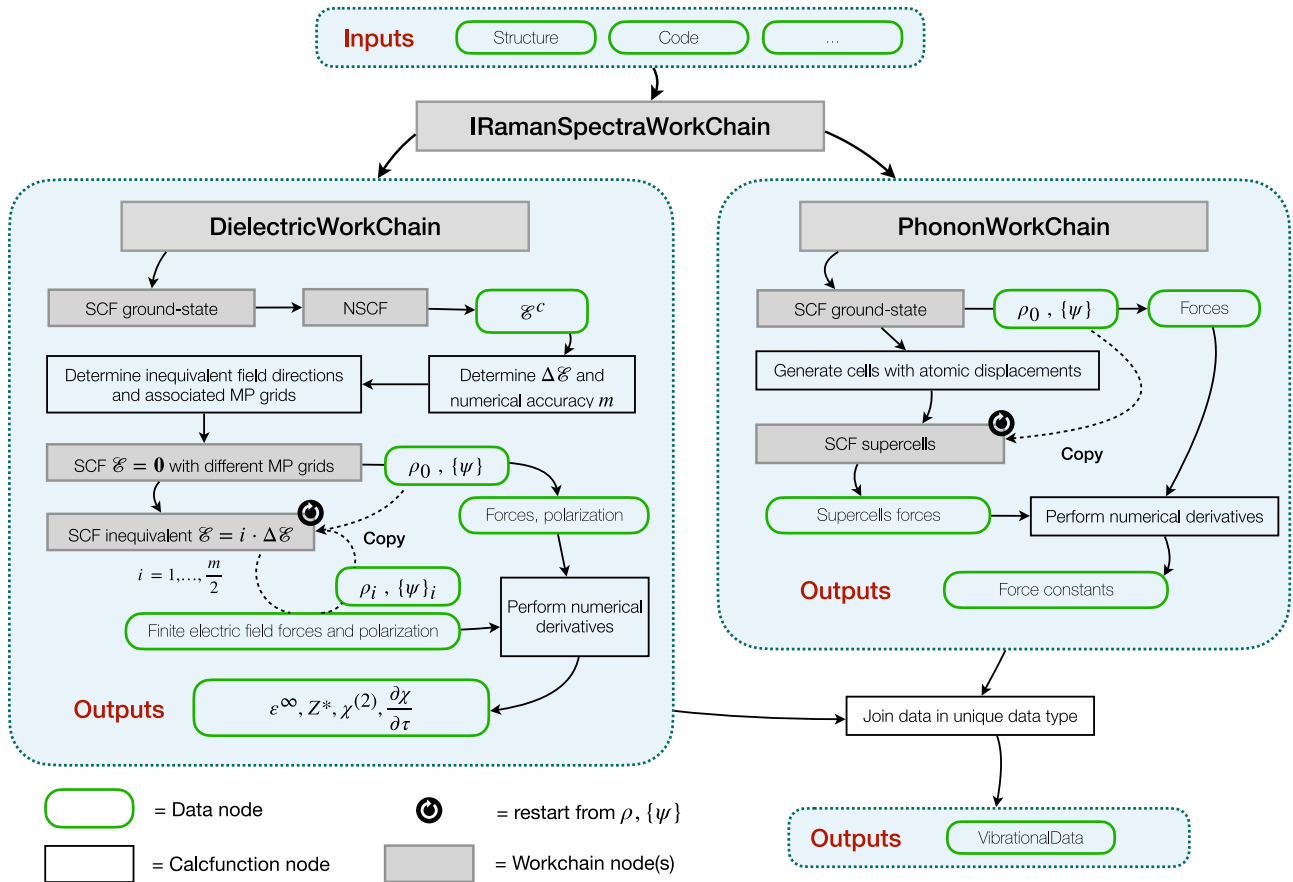
$$\mathbf{P}^{\text{el}} \cdot \mathbf{b}_i = \frac{2e}{N_{\perp}^{(i)}\Omega} \sum_{l=1}^{N_{\perp}^{(i)}} \text{Im} \ln \prod_{j=0}^{N_{\parallel}^{(i)}-1} \det[S(\mathbf{k}_{1j}^{(i)}, \mathbf{k}_{1j+1}^{(i)})]. \quad (21)$$

Here,  $S_{nm}(\mathbf{k}, \mathbf{k}') = \langle u_{n\mathbf{k}} | u_{m\mathbf{k}'} \rangle$  represents the scalar product between Bloch’s state of band  $n$  and  $k$ -point  $\mathbf{k}$ , and state of band  $m$  and  $k$ -point  $\mathbf{k}'$ .

Assigning momentarily  $i = 1$  to  $\mathbf{b}_i$  as a way of example, then  $N_{\perp}^{(1)} = N_2 \times N_3$  is the number of  $k$ -point strings parallel to  $\mathbf{b}_1$ , each such string having  $N_{\parallel}^{(1)} = N_1$   $k$ -points defined as  $\mathbf{k}_{lj}^{(1)} = \mathbf{k}_l^{(1)} + (j/N_1)\mathbf{b}_1$ , where  $\mathbf{k}_l^{(1)}$  belongs to the plane  $(\mathbf{b}_2, \mathbf{b}_3)$ , corresponding to a slice of the full mesh  $N_1 \times N_2 \times N_3$ , and is indexed by  $l$ . The total electronic polarization is then given by

$$\mathbf{P}^{\text{el}} = \frac{1}{2\pi} \sum_{i=1}^3 \mathbf{a}_i (\mathbf{P}^{\text{el}} \cdot \mathbf{b}_i), \quad (22)$$

where  $\mathbf{a}_i$  are the direct lattice vectors. In this way, the calculation of the polarization in real space requires to evaluate Eq. (21) for each (reciprocal) lattice vector and on top of very dense meshes. Once the self-consistent electronic charge-density is found, hence the Kohn-Sham Hamiltonian is fixed, one can increase straightforwardly the amount of  $k$ -points needed for accurate polarization calculations, by diagonalizing non self-consistently the Kohn-Sham Hamiltonian at the desired grid and thus avoiding the expensive self-consistent cycle. Yet, such diagonalization can be computationally demanding. A smart solution is to compute each  $\mathbf{P}^{\text{el}} \cdot \mathbf{b}_i$  contribution at a different sampling, noticing that the right-hand side of Eq. (21) converges faster with respect to  $N_{\perp}^{(i)}$ , and consequently one can simply increase the number of  $k$ -points along  $\mathbf{b}_i$ , i.e., by increasing  $N_i$  only. This allows to scale linearly with the number of points in the string, instead of increasing all  $N_i$  (cubic scaling), at the price of performing three calculations instead of a single, overall more expensive, calculation on a very dense mesh. Importantly, one needs to carefully check convergence with respect to both the number of “orthogonal”  $k$ -points  $N_{\perp}^{(i)}$  and the number of  $k$ -points per string  $N_{\parallel}^{(i)} = N_i$ . Unfortunately, in the case of a finite electric field, a non-local operator which couples neighboring  $k$ -points is introduced in the Hamiltonian<sup>39–41</sup>. This means that, on one hand, one cannot simply diagonalize in a post-processing step the Hamiltonian, and on the other hand, one is in principle not justified to split the calculation in three parts as formerly explained. We then seek a sampling strategy that can simultaneously alleviate the  $k$ -points scaling, and that can still provide accurate polarizations. The main idea is to perform a single calculation with a *special*  $k$ -point grid. We can exploit the fact that we are interested in calculating the induced polarization by an homogeneous electric field, contrary to other type of calculations where the direction of the induced polarization, e.g., due to atomic displacements, is less obvious (one can still



**Fig. 2 | Schematic representation of the automated workflow for the calculation of infrared and Raman spectra.** Computational workflow scheme of the IRamanSpectraWorkChain and of its sub-workflows, the DielectricWorkChain and the PhononWorkChain. The inputs can be easily, fully

defined via the `get_builder_from_protocol` while specifying only a `StructureData` and a `Code`. The black circles on top of boxes mean that the calculation starts using charge density and wavefunctions from a previous calculation.

exploit the strategy proposed above, if an electric field is absent). Let's assume that the induced polarization by an electric field is larger in the direction of the applied electric field, meaning we can focus on computing it more accurately along such direction, and mimic the strategy reported previously, but using a *sole*, electric field-dependent, *directional* k-point mesh. To explain this further, let  $\tilde{\mathcal{E}} = (\tilde{\mathcal{E}}_1, \tilde{\mathcal{E}}_2, \tilde{\mathcal{E}}_3)$  be the electric field in the lattice reference system. The product  $\mathbf{P}^{\text{el}} \cdot \mathbf{b}_i$  is then expected to be larger where  $\tilde{\mathcal{E}}_i$  is larger, and, as a consequence, one can increase  $N_{\parallel}^{(i)} = N_i$  weighted upon the relative value of  $\tilde{\mathcal{E}}_i$  with respect to the other components of  $\tilde{\mathcal{E}}$ . To automate the choice of  $N_i$  in a weighted manner, we start by defining two quantities: an *orthogonal* and a *parallel* k-point distance,  $d_{\perp}$  and  $d_{\parallel}$ , respectively. The *orthogonal* distance represents a minimum distance between k-points in the BZ, thus usually producing small  $N_i$  - this quantity is used as a reference minimum for  $N_i$  in the unique mesh, resembling essentially  $N_i$  employed for orthogonal meshes in the three calculations strategy. The *parallel* distance instead is used to define a weighted  $N_i$  which should be greater than its reference minimum, defined previously by the orthogonal distance, when  $\tilde{\mathcal{E}}_i$  is considered large enough (see Fig. 1b, c). To quantify the relative magnitude of  $\tilde{\mathcal{E}}_i$ , and consequently choosing the weighting, we assign the following weights:

$$w_i = \frac{|\tilde{\mathcal{E}}_i|}{\max\{|\tilde{\mathcal{E}}_1|, |\tilde{\mathcal{E}}_2|, |\tilde{\mathcal{E}}_3|\}}, \quad (23)$$

aiming at increasing  $N_i$  for lattice directions where the electric field is larger, and where we expect the induced electric polarization to be greater. Finally,

we can assign the number  $N_i$  of the mesh as:

$$N_i = \max \left\{ w_i \frac{|\mathbf{b}_i|}{d_{\parallel}}, \frac{|\mathbf{b}_i|}{d_{\perp}} \right\}. \quad (24)$$

This methodology is easy to implement and generates meshes that significantly reduce the number of k-points needed instead when using a uniform sampling, i.e.,  $N_i = |\mathbf{b}_i|/d$ , where  $d$  is a constant k-points distance (see Supplementary Discussion for detailed numerical experiments). This scheme allows for linear scaling of k-points with respect to  $1/d_{\parallel}$  when the applied electric field is perfectly aligned along one of the lattice vectors, as it can be seen in the simple case of Fig. 1b. In orthorhombic systems this is particularly convenient, as the three lattice vectors are parallel to the three Cartesian axis. In general, when the electric field direction is not parallel to one of the lattice vectors, the number of k-points scales as a power of  $1/d_{\parallel}$ . In the extreme case when  $\tilde{\mathcal{E}}_1 = \tilde{\mathcal{E}}_2 = \tilde{\mathcal{E}}_3$ , then the number of k-points will scale cubically with respect to the parallel distance, equivalent to a uniform mesh defined by a single k-point distance  $d \equiv d_{\parallel}$ . This is schematically depicted in Fig. 1c for a simple case in 2D.

### Computational workflow

In this section, we finally present the computational workflows that allows to simulate the desired spectra—infrared, Raman, or both—in an automated fashion. We designed two main independent workflows as AiiDA WorkChains<sup>22</sup>: one carrying out solely the finite displacement calculations



for phonon modes to evaluate the IFCs, which we called `PhononWorkChain`, and the other for the mixed total-energy derivatives via the homogeneous electric fields, the `DielectricWorkChain`. When joined together, the above two provide the `IRamanSpectraWorkChain`, as schematically outlined in Fig. 2.

The `IRamanSpectraWorkChain` takes as inputs a structure, a code and other details on how to run the workflows (e.g., SCF parameters, wall-time, parallelization options, and so on). These information are then split and/or shared between the two sub-workflows. Both start running an initial SCF ground-state calculation, meant to produce the charge density  $\rho_0$  and wavefunctions  $\{\psi_i\}$  that will be used as a starting point for the next runs. We represent this restart throughout the flowchart of Fig. 2 using a rotating white arrow in a black circle, and we use the dashed arrows to indicate that the files have been copied over to a fresh new folder. When these calculations are finished, the `PhononWorkChain` generates the structures with irreducible atomic displacements, using `Phonopy`<sup>16</sup>, and computes their forces. The latter are then used to produce the matrix of IFCs in the unit cell (although the `PhononWorkChain` is designed to compute the IFCs in a generic supercell), which are employed in a post-processing step to obtain the phonon frequencies and eigenvectors by diagonalizing the corresponding dynamical matrix, Eq. (10), for  $\mathbf{q} \rightarrow \mathbf{0}$ . In parallel, the `DielectricWorkChain` evaluates  $\mathcal{E}^c$  through a non-SCF calculation and extracts the numerical accuracy  $n$  and the electric field step  $\Delta\mathcal{E}$ , as follows: when the critical value is larger than  $10^{-4}$  Ry a.u. (we use Rydberg atomic units; 1 Ry a.u.  $\approx 36.3609$  V/Å), a 4th-order numerical expansion is chosen, otherwise a 2nd order is used. The 4th-order allows to achieve a greater numerical accuracy for finite differences; thus, it is utilized whenever it is possible to apply an electric field above threshold. If the electric field that can be applied is smaller than  $10^{-4}$  Ry a.u., a 2nd-order expansion is sufficient, thus avoiding calculations that would be affected by too much numerical noise<sup>38</sup>; the amplitude of the electric field step in fact may result too small to produce significant digit changes in forces and polarization, hence possibly worsening the numerical derivatives even if adding extra points to the formula. The step  $\Delta\mathcal{E}$  is instead selected according to:

$$\Delta\mathcal{E} = \begin{cases} \frac{10^{-3}}{n} & \text{if } \mathcal{E}^c > 10^{-3}, \\ \frac{\text{round}(\mathcal{E}^c \cdot 5)}{n} & \text{if } 10^{-4} < \mathcal{E}^c \leq 10^{-3}, \\ \frac{\text{round}(\mathcal{E}^c \cdot 6)}{n} & \text{if } \mathcal{E}^c \leq 10^{-4}, \end{cases} \quad (25)$$

where the electric field units are again expressed in Ry a.u. and  $\text{round}(x,y)$  rounds the number  $x$  to the  $y$  digit after the decimal point. These choices should guarantee larger values of the finite step to reduce the numerical noise on forces and polarization, and at the same time higher accuracy to remove the step size dependence of the finite-difference formula, which can be verified a posteriori when  $n > 2$  (see also Fig. 1a). We point out here that eventually the user can still enforce the accuracy  $n$  and/or the  $\Delta\mathcal{E}$  value that will be used in Eqs. (18) and (25). Once an electric field is chosen and the ground-state SCF calculation is finished, a symmetry analysis is performed to find the independent directions  $\hat{\mathcal{E}}$  of the electric fields to apply, where  $\hat{\mathcal{E}}$  is a versor. For each such independent direction a MP mesh is determined through Eq. (24) and a series of SCF calculations are then launched at  $\mathcal{E} = \mathbf{0}$  with the corresponding mesh restarting from the ground-state density of the previous SCF. The reason for computing SCF calculations at  $\mathcal{E} = \mathbf{0}$  with these meshes is twofold: first, they provide the ground-state charge density and wavefunctions for the next finite electric field calculations, resulting in faster convergence of the self-consistent cycles, and, second, they guarantee the same level of convergence on forces and polarization for the finite differentiation. Once finished, electric-enthalpy calculations start for each independent direction  $\hat{\mathcal{E}}$  at the smallest step value  $\Delta\mathcal{E} = \Delta\mathcal{E} \cdot \hat{\mathcal{E}}$ , to provide an adiabatic switching of the perturbation. If the numerical accuracy  $n$  is greater than 2, further SCF calculations with homogeneous electric fields are run, restarting from the previous steps till the maximum value  $(n/2)\Delta\mathcal{E}$ ; this loop corresponds to increasing/decreasing the value of  $l$  in Eq. (18). At this point, the computed forces and polarizations are gathered together to

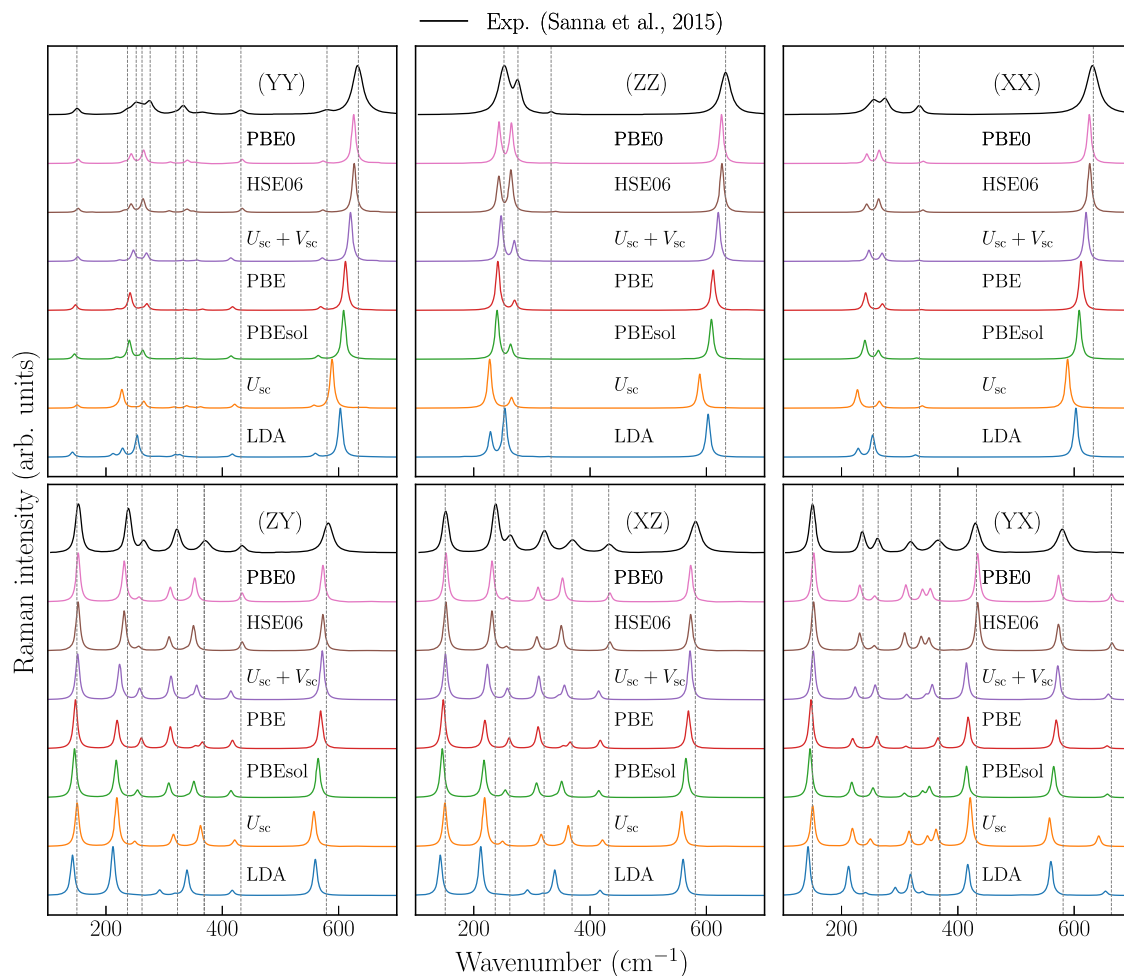
perform the numerical differentiation following Eqs. (18) and (19) to obtain the second- and third-rank tensors (12)-(13)-(15). As a final step, the `IRamanSpectraWorkChain` joins all the information in a unique AiiDA data type, named `VibrationalData`, able to efficiently store large arrays and information in the AiiDA repository and database, and with provenance graph for full reproducibility<sup>22</sup>. This data type is a self-contained Python class, and it is provided with a number of post-processing methods and features that allow the user to compute powder and single-crystal polarized vibrational intensities and frequencies, along with the symmetry labels of the respective phonon modes. We also implemented an `IntensitiesAverageWorkChain` which performs the spherical-average spectra, important in Raman for non-centrosymmetric crystals<sup>11</sup> when in a polycrystalline powder form.

As mentioned earlier, the inputs of these workflows can be fully customized further by the user. We do not report here the full list of these inputs, which can be explored through the usual command line interface of AiiDA<sup>22</sup> and through the on-line documentation made available (<https://aiida-vibrospectro.readthedocs.io/en/latest/>). However, we emphasize the possibility of fully defining the entire set of inputs required via a limited subset, as already done in other AiiDA packages<sup>45</sup>; in similar fashion to ref. <sup>45</sup>, we identify the fundamental inputs to be the `code`, i.e., the Quantum ESPRESSO<sup>46-48</sup> `pw.x` binary, associated with the computer, that runs the calculations, the relaxed `structure` and the `protocol`, a string defining in a general and user-friendly fashion the accuracy of the simulation. This minimal set, used through the `get_builder_from_protocol` method<sup>45</sup> designed for all the workchains, provides the user with a pre-filled builder (the main unit in AiiDA used to run workflows) ready to be submitted. This is adequate for non-experts in the field that might be interested in simulating spectra of a given structure with minimal computational knowledge, but is also very precious in the context of high-throughput searches or for teaching or testing.

### All-functionals LiNbO<sub>3</sub> spectra

We showcase the strength of this approach by studying the ferroelectric phase of LiNbO<sub>3</sub> (space group R3c, no. 161) for which the vibrational spectra have long been debated in the literature<sup>49-52</sup> and exhibit large LO-TO phonon splittings at  $\Gamma$ <sup>49</sup> that are ideal for exploring different exchange-correlation functionals. We selected three commonly used local and semi-local functionals, LDA, PBE and PBEsol; Hubbard-corrected PBEsol in the standard and extended version, namely DFT+U<sup>53,54</sup> and DFT+U+V<sup>55-57</sup>; and two of the most used hybrids functionals for inorganic crystalline materials, HSE06<sup>58,59</sup> and PBE0<sup>60</sup>.

Through the use of the `IRamanSpectraWorkChain`, we carried out the simulation of the vibrational spectra for all the functionals above and compare our results to experimental measurements from the literature<sup>11,51,52,61,62</sup>. In particular, we compare the simulations presented in Fig. 3 with different experimental Raman single-crystal polarization setups where LO modes are filtered out<sup>52</sup>; in Fig. 4 with single-crystal and polycrystalline powder Raman spectra including the LO modes<sup>11,51,61</sup>; and in Fig. 5 with single-crystal infrared absorption<sup>62</sup> and reflectivity spectra<sup>51</sup>. We make use of the Porto notation for Raman setups: the labels within the parenthesis refer to the *polarization* of light  $\mathbf{e}_L$  and  $\mathbf{e}_S$ , the external labels to the *propagation* of light  $\mathbf{k}_L$  and  $\mathbf{k}_S$ . When only TO modes are shown, the external labels are not specified. For spectra showing LO modes, we set  $\mathbf{q} \parallel \mathbf{k}_L - \mathbf{k}_S$  in single crystals, while polycrystalline powder spectra are modeled via the spherical-average procedure, as outlined in ref. <sup>11</sup>. The conventional cell is used to define the Cartesian directions, aligning the Z axis along the  $c$  axis and the X axis along the  $a$  axis (Y is chosen to obtain a right-handed reference system). The single-crystal data of ref. <sup>52</sup>, reported in Fig. 3, allow us to compare directly the intensities (TO modes only). The experimental LO and TO frequencies were reported in ref. <sup>52</sup>, and are summarized, along with the theoretical predictions, in Supplementary Table 1. We ranked the spectra in the plots in function of the statistical estimator  $|\overline{\Delta\omega}| = \sum_{\nu=1}^N |\omega_{\nu}^{\text{exp}} - \omega_{\nu}^{\text{theo}}|/N$ , where  $N$  is the number of active modes detected. All the functionals show good agreement with experiments and



**Fig. 3 | Polarized single-crystal Raman spectra of ferroelectric LiNbO<sub>3</sub> as obtained with seven different exchange-correlation functionals, and compared to experimental measurements<sup>52</sup>.** The scattering geometries are reported in parenthesis. The theoretical spectra are ordered according to the error  $|\Delta\omega|$  (shown in

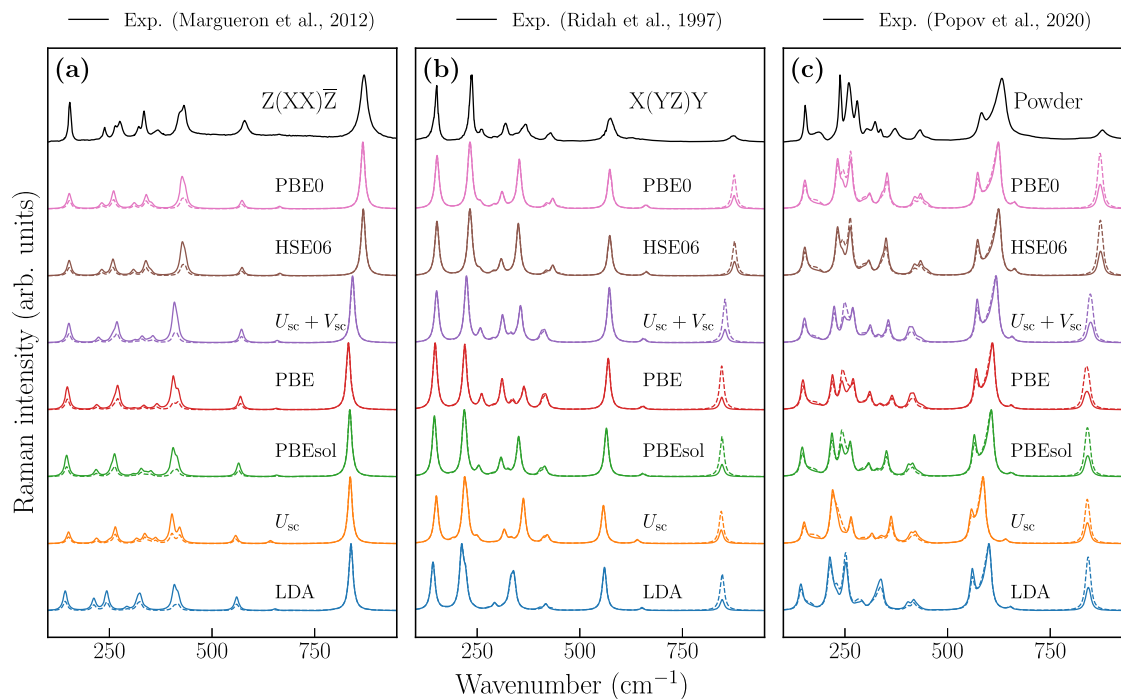
Table 1), from smallest (pink, PBE0) to largest (blue, LDA). Hubbard-corrected PBEsol +  $U_{sc}$  and PBEsol +  $U_{sc} + V_{sc}$  are labeled just  $U_{sc}$  and  $U_{sc} + V_{sc}$ , the subscript  $sc$  refers to the self-consistent calculation of  $U$  and  $V^{77,78}$ . Theoretical intensities are smeared with a  $8\text{ cm}^{-1}$  wide Lorentzian.

with the results of the literature<sup>49–52</sup>. PBEsol +  $U_{sc} + V_{sc}$ , PBE, and the two hybrids show competitive results among each other, while PBEsol, PBEsol +  $U_{sc}$  and LDA have quantitative larger discrepancies. In particular in Table 1 we report statistical estimators for the comparison of the theoretical and experimental phonon frequencies (LO and TO) and intensities (only TO, Fig. 3). These spectra setups allowed to accurately estimate the error of the intensities evaluated through the minimization of the relative error:

$$\Delta I(f) = \sqrt{\frac{\sum_{\nu=1}^N (I_{\nu}^{\text{theo}} - f I_{\nu}^{\text{exp}})^2}{\sum_{\nu=1}^N (f I_{\nu}^{\text{exp}})^2}} \quad (26)$$

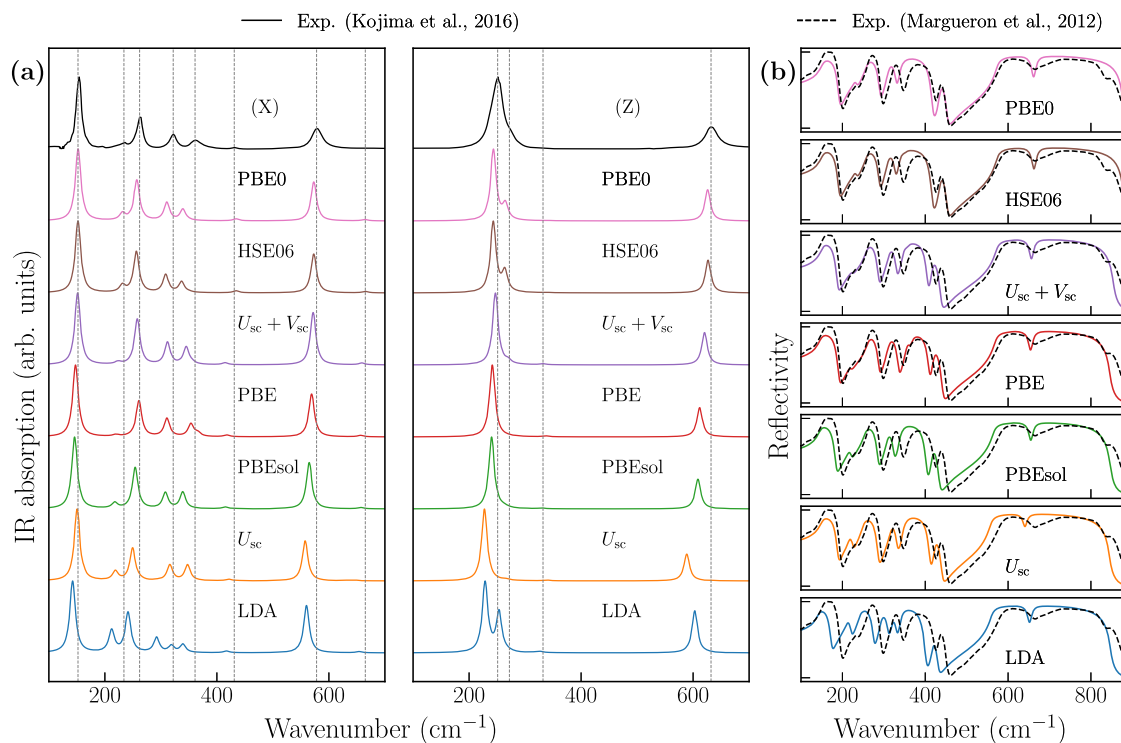
where  $f$  represents the unknown scaling factor between theoretical  $I^{\text{theo}}$  and experimental  $I^{\text{exp}}$  intensities, due to the use of arbitrary units in the experimental spectra. The smallest absolute error on frequencies  $|\Delta\omega|$  is provided by the PBE0 functional with  $7.3\text{ cm}^{-1}$ , and on intensities  $|\Delta I|$  by the PBEsol +  $U_{sc} + V_{sc}$  functional with 16%. It is to note that the latter also shows a very good  $|\Delta\omega|$  of  $10.7\text{ cm}^{-1}$ , about  $3\text{ cm}^{-1}$  apart from the best performing functional for this material, whereas the former has 24% error for  $|\Delta I|$ , 8% more with respect to PBEsol +  $U_{sc} + V_{sc}$ . We also report the renormalized absolute error  $|\Delta\omega|^{*}$ , for which the frequencies are shifted by the bare mean error  $\bar{\Delta\omega}$  before the modulus average; this gives some insights whether the discrepancy arises from a systematic error, e.g., due to the geometry used. Interestingly, after renormalization, most of the functionals get rather close error values. While this might be due to the renormalization

of the phonon frequencies due to the thermal expansion, which tends to soften the vibrational modes, all the theoretical spectra were performed at the experimental lattice geometry (see the Method section), meaning the systematic error is unrelated to this effect. While other thermal contributions should in principle be accounted for, their inclusion would require more sophisticated levels of theory<sup>63–65</sup> which go beyond the scope of the present work. Moreover, at the conditions of the experiments, such contributions are expected to produce only minor phonon shifts<sup>51</sup>. Given these points we might conclude in this case that such errors are intrinsic to the functionals. We proceed in the following with a more qualitative analysis of the other calculated spectra, as the experimental data we refer to come from different sources. In Fig. 4 we show the result of the complex interplay of long-range interactions, Eq. (11), and the non-linear dielectric response which renormalizes the Raman intensities, Eq. (17). It can be noted again that all functionals are able to describe well both the peak positions and the intensities of the different experimental measurements. In particular, the intensity of the LO mode at around  $870\text{ cm}^{-1}$  is reproduced correctly. The dashed lines in the figure highlight the error committed when neglecting the  $\chi^{(2)}$  contribution in Eq. (17), which is relevant for the accurate prediction of the intensity. As already found in ref.<sup>11</sup>, the spherical average employed to compute the spectra of the polycrystalline sample of Fig. 4c correctly reproduces the asymmetric lineshapes of the experimental peaks, which can be understood as an average of all the possible  $\mathbf{q} \rightarrow \mathbf{0}$  splitting modes which the incident light happens to activate in a polycrystalline powder sample. The absorption and reflectivity infrared spectra reported in Fig. 5 show very



**Fig. 4 | Polarized single-crystal and powder Raman spectra of ferroelectric LiNbO<sub>3</sub> as obtained with seven different exchange-correlation functionals, and compared to experimental measurements<sup>51,62</sup>.** **a, b** Raman spectra of single-crystal polarized setups; the scattering geometry is reported in the Porto notation. **c** Raman spectra of polycrystalline powder sample; the theoretical spectra are computed using

the spherical average formula<sup>11</sup>. In each panel, the dashed lines show the computed spectra when the second term in Eq. (17) is neglected. Theoretical intensities are smeared with a constant 12 cm<sup>-1</sup> broadened Lorentzian. The theoretical spectra are reported in order of decreasing error  $|\overline{\Delta\omega}|$ , and Hubbard-corrected PBEsol are abbreviated only with their corrective parameters, as in Fig. 3.



**Fig. 5 | Absorption and reflectivity spectra in the IR regime of ferroelectric LiNbO<sub>3</sub> as obtained with seven different exchange-correlation functionals, and compared to experimental measurements<sup>51,62</sup>.** **a** Infrared absorption using polarized light for E (left panel) and A<sub>1</sub> (right panel) symmetry modes. **b** Normal

reflectivity spectra of Z-cut single crystal; upper and lower ticks for each panel correspond to 1 and 0, respectively. The order of theoretical spectra and the labels are as in Fig. 3. Theoretical intensities are smeared with a constant 12 cm<sup>-1</sup> broadening; a Lorentzian function is used for **a**.



**Table 1 | Statistical error estimators between predictions and experiments<sup>52</sup> for  $E$  and  $A_1$  LO-TO modes**

	LDA	PBEsol+ $U_{sc}$	PBEsol	PBE	PBEsol + $U_{sc}$ + $V_{sc}$	HSE06	PBE0
$ \Delta\omega $	22.2	15.1	15.1	11.0	10.7	7.9	7.3
$\overline{\Delta\omega}$	22.2	14.1	15.1	10.5	10.6	5.7	5.2
$ \Delta\omega _{\max}$	45.4	44.3	35.4	38.8	28.7	27.1	24.8
$ \Delta\omega ^*$	8.1	8.9	6.7	7.3	6.2	6.9	6.2
$ \Delta I $	34	25	22	23	16	25	24

Statistical error estimators given in  $\text{cm}^{-1}$  for frequencies  $\omega$  and in percentage for the Raman intensities  $I$  (the intensities are compared for TO modes only; see Fig. 3 and Supplementary Discussion). The estimators are:  $|\Delta\omega| = \sum_{\nu=1}^N |\Delta\omega_{\nu}|/N$  is the mean absolute error;  $\overline{\Delta\omega} = \sum_{\nu=1}^N \Delta\omega_{\nu}/N$  is the mean error;  $|\Delta\omega|_{\max}$  is the absolute max difference;  $|\Delta\omega|^* = \sum_{\nu=1}^N |\Delta\omega_{\nu} - \overline{\Delta\omega}|/N$  is the mean absolute error for frequencies shifted by their mean difference;  $|\Delta I| = 100 \cdot \text{min}_f |\Delta(f)|$  represents the percentage error of the optimal scaled intensities;  $N$  is the number of considered modes and  $\Delta\omega_{\nu}$  is the difference between experimental and theoretical functional mode frequency.

**Table 2 | Ordinary and extraordinary dielectric components of  $\text{LiNbO}_3$  from experiments and from seven different exchange-correlation functionals**

	LDA	PBEsol+ $U_{sc}$	PBEsol	PBE	PBEsol + $U_{sc}$ + $V_{sc}$	HSE06	PBE0	Exp. <sup>51</sup>	Exp. <sup>66</sup>
$\epsilon_0^o$	48.1	45.3	44.6	43.1	39.4	43.2	43.0	40.2	41.5
$\epsilon_0^e$	29.6	31.8	27.7	26.3	24.3	25.6	25.7	23.7	26

Ordinary and extraordinary static dielectric components,  $\epsilon_0^o$  and  $\epsilon_0^e$ , as obtained with seven different functionals (ordered from left to right according to the mean average error shown in Table 1), and compared to the experimental results<sup>51,66</sup>. Other experimental values are reported in ref. 51.

good agreement with the experimental measurements<sup>51,62</sup>. For a more quantitative comparison, here we compare in Table 2 the static dielectric tensor  $\epsilon_{ij}(\omega = 0)$  for the ordinary and extraordinary components, which can be measured from absorption measurements. The theoretical predictions are close to the experimental values. We note that it is difficult here to assess accurately the theoretical errors, as the experimental evaluation of this quantity is usually subject to deviations, as it can be seen in Table 2. Nonetheless, once again, PBE, PBEsol +  $U_{sc}$  +  $V_{sc}$  and the two hybrids show a remarkable agreement with the experimental results reported<sup>51,66</sup>.

## Discussion

In this paper, we have presented a comprehensive and efficient approach that utilizes finite displacements and finite fields to calculate seamlessly various vibrational spectra, ranging from absorption and normal reflectivity spectra in the infrared regime, to Raman spectra, including non-centrosymmetric crystals and fine-grained materials (spherical-average method), with any modern exchange-correlation functional and pseudo-potential formulation (norm-conserving, ultrasoft, PAW). This approach requires 12 self-consistent electric field calculations for the prediction of Raman tensors, as opposite to  $6 \cdot N_{at}$  when taking the derivative of the dielectric tensor with respect to atomic displacements<sup>8,9</sup>, which remains impractical in large or low symmetry systems. The approach we employ becomes extremely useful when generating coupling tensors in low-symmetry supercells, Eq. (6), to incorporate temperature and quantum effects<sup>64,65,67</sup>, or to train a tensorial machine-learning potential for molecular dynamics simulations. The formalism has been implemented as a highly optimized, automatic workflow package for the prediction of vibrational spectra. The technical and practical challenges have been successfully addressed with the help of the AiiDA infrastructure<sup>22</sup>, which empowered the design of fully reproducible, reusable, user-friendly workflows for one of the most important class of spectroscopic techniques for materials characterization. As such, we believe it will be broadly useful to both the computational community and the experimental community at large. Furthermore, we have shown that the workflows can effectively operate with state-of-the-art functionals and can be employed for a comprehensive analysis of both single crystals and powder spectra. Importantly, all the code, workflows and the documentation are open and available to the community (see Code Availability). In conclusion, we believe that the present approach will be helpful for the computational and experimental community and will pave

the way for accelerated materials characterization of materials displaying complex, challenging chemistry.

## Methods

### Computational details

The infrared and Raman calculations were carried out using the workflows developed, relying on aiida-phonopy<sup>68</sup> for pre- and post-processing and aiida-quantumpresso<sup>22,23</sup>, having Quantum ESPRESSO<sup>46–48</sup> as DFT quantum engine. Before performing the vibrational spectra calculations, the positions of the atoms in the  $\text{LiNbO}_3$  primitive cell are optimized at fixed experimental lattice geometry<sup>69</sup> for all functionals until the total energy and forces acting on atoms are less than  $10^{-4}$  Ry/atom and  $5 \cdot 10^{-5}$  Ry/Bohr, respectively. Norm-conserving pseudopotentials from the Pseudo-Dojo library<sup>70</sup> are used for the LDA and hybrid functionals, employing PBE pseudopotentials for the latter, and pseudopotentials from the SSSP PBE(-sol) precision (version 1.1)<sup>70–76</sup> library for all other cases. A cutoff of 90 Ry is used for the wavefunctions; 4 and 12 times greater cutoffs are used for the charge density (respectively for Pseudo-Dojo and SSSP based pseudopotentials). The Brillouin zone is sampled uniformly with a  $0.3 \text{ \AA}^{-1}$  k-points distance, leading to a  $5 \times 5 \times 5$  mesh for the primitive cell containing 10 atoms. The Hubbard parameters are computed following the self-consistent procedure exploiting density-functional perturbation theory, as explained in refs. 77–79, using a q-point distance of  $0.4 \text{ \AA}^{-1}$ , orthogonalized atomic orbitals<sup>78,80</sup> for the occupation matrices, and with onsite (intersite) parameters converged within 0.01 (0.005) eV (a very conservative threshold). The hybrid calculations are performed using a 90 Ry cutoff and a single q-point mesh for the exact-exchange operator, and the adaptively compressed exchange (ACE) technique<sup>81</sup> is used to speed up the convergence of the calculations. Finally, finite differences are carried out using a  $0.01 \text{ \AA}$  displacement distance and a denser k-point distance of  $0.15 \text{ \AA}^{-1}$  for the phonon calculations (in the PhononWorkChain), and an electric field step of  $5 \cdot 10^{-4}$  Ry a.u., a 2nd-order numerical accuracy, and a k-point distance of  $0.15 \text{ \AA}^{-1}$  in the parallel direction of the applied electric field for the electric-field derivatives (as carried out in the DielectricWorkChain).

### Data availability

The data produced in this work can be found on the Materials Cloud Archive<sup>82</sup>.

## Code availability

The source code is made available open-source on GitHub (<https://github.com/bastonero/aiida-vibrosopy>). It is also distributed as an installable package through the Python Package Index (<https://pypi.org/project/aiida-vibrosopy/>). A live demonstration of the code is hosted on <https://aiida-vibrosopy-demo.readthedocs.io/en/latest/index.html>, whereas the full documentation with tutorials can be found at <https://aiida-vibrosopy.readthedocs.io/en/latest/>.

Received: 28 July 2023; Accepted: 29 February 2024;

Published online: 15 March 2024

## References

- Giannozzi, P., de Gironcoli, S., Pavone, P. & Baroni, S. Ab initio calculation of phonon dispersions in semiconductors. *Phys. Rev. B* **43**, 7231–7242 (1991).
- Windl, W. et al. Second-order raman spectra of diamond from ab initio phonon calculations. *Phys. Rev. B* **48**, 3164–3170 (1993).
- Gonze, X. Adiabatic density-functional perturbation theory. *Phys. Rev. A* **52**, 1096–1114 (1995).
- Gonze, X. & Lee, C. Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B* **55**, 10355–10368 (1997).
- Baroni, S., de Gironcoli, S., Corso, A. D. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Modern Phys.* **73**, 515–562 (2001).
- Lazzeri, M. & Mauri, F. First-principles calculation of vibrational Raman spectra in large systems: signature of small rings in crystalline. *Phys. Rev. Lett.* **90**, 036401 (2003).
- Veithen, M., Gonze, X. & Ghosez, P. Nonlinear optical susceptibilities, Raman efficiencies, and electro-optic tensors from first-principles density functional perturbation theory. *Phys. Rev. B* **71**, 125107 (2005).
- Baroni, S. & Resta, R. Ab initio calculation of the low-frequency Raman cross section in silicon. *Phys. Rev. B* **33**, 5969–5971 (1986).
- Liang, Q., Dwaraknath, S. & Persson, K. A. High-throughput computation and evaluation of Raman spectra. *Sci. Data* **6**, 135 (2019).
- Calzolari, A. & Nardelli, M. B. Dielectric properties and Raman spectra of ZnO from a first principles finite-differences/finite-fields approach. *Sci. Rep.* **3**, 2999 (2013).
- Popov, M. N. et al. Raman spectra of fine-grained materials from first principles. *npj Comput. Mater.* **6**, 121 (2020).
- Souza, I., Íñiguez, J. & Vanderbilt, D. First-principles approach to insulators in finite electric fields. *Phys. Rev. Lett.* **89**, 117602 (2002).
- Umari, P. & Pasquarello, A. Ab initio molecular dynamics in a finite homogeneous electric field. *Phys. Rev. Lett.* **89**, 157602 (2002).
- Umari, P. & Pasquarello, A. Polarizability and dielectric constant in density-functional supercell calculations with discrete k-point samplings. *Phys. Rev. B* **68**, 085114 (2003).
- Umari, P. & Pasquarello, A. Density functional theory with finite electric field. *Int. J. Quantum Chem.* **101**, 666–670 (2005).
- Togo, A. & Tanaka, I. First principles phonon calculations in materials science. *Scr. Mater.* **108**, 1–5 (2015).
- Zwanziger, J. et al. Finite homogeneous electric fields in the projector augmented wave formalism: applications to linear and nonlinear response. *Comput. Mater. Sci.* **58**, 113–118 (2012).
- Umari, P. & Pasquarello, A. Infrared and Raman spectra of disordered materials from first principles. *Diamond Relat. Mater.* **14**, 1255–1261 (2005).
- Umari, P. & Pasquarello, A. Fraction of boroxol rings in vitreous boron oxide from a first-principles analysis of Raman and NMR spectra. *Phys. Rev. Lett.* **95**, 137401 (2005).
- Niu, L. et al. First-principles calculation of vibrational Raman spectra of tetrahedral amorphous carbon. *Phys. B: Condens. Matter* **403**, 3559–3562 (2008).
- Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
- Huber, S. P. et al. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* **7**, 300 (2020).
- Uhrin, M., Huber, S. P., Yu, J., Marzari, N. & Pizzi, G. Workflows in AiiDA: engineering a high-throughput, event-based engine for robust and modular computational workflows. *Comput. Mater. Sci.* **187**, 110086 (2021).
- El-Batanouny, M. & Wooten, F. *Symmetry and Condensed Matter Physics* (Cambridge University Press, 2008).
- Max Born, K. H. *Dynamical Theory of Crystal Lattices* (Clarendon, Oxford, 1954).
- Placzek, G. Rayleigh-streuung und Raman-effekt. in *Hanbuch der Radiologie*, vol. VI, 209–374 (ed Marx, E.) (Akademische Verlagsgesellschaft, Leipzig, 1934).
- Brüesch, P. *Phonons: Theory and Experiments II* (Springer Berlin Heidelberg, 1986).
- Porezag, D. & Pederson, M. R. Infrared intensities and Raman-scattering activities within density-functional theory. *Phys. Rev. B* **54**, 7830–7836 (1996).
- Gonze, X. First-principles responses of solids to atomic displacements and homogeneous electric fields: implementation of a conjugate-gradient algorithm. *Phys. Rev. B* **55**, 10337–10354 (1997).
- Tschannen, C. D. et al. Anti-stokes Raman scattering of single carbyne chains. *ACS Nano* **15**, 12249–12255 (2021).
- Sohier, T., Gibertini, M., Calandra, M., Mauri, F. & Marzari, N. Breakdown of optical phonons' splitting in two-dimensional materials. *Nano Lett.* **17**, 3758–3763 (2017).
- Rivano, N., Marzari, N. & Sohier, T. Infrared-active phonons in one-dimensional materials and their spectroscopic signatures. *npj Comput. Mater.* **9**, 194 (2023).
- Royo, M., Hahn, K. R. & Stengel, M. Using high multipolar orders to reconstruct the sound velocity in piezoelectrics from lattice dynamics. *Phys. Rev. Lett.* **125**, 217602 (2020).
- Nunes, R. W. & Gonze, X. Berry-phase treatment of the homogeneous electric field perturbation in insulators. *Phys. Rev. B* **63**, 155107 (2001).
- Vanderbilt, D. & King-Smith, R. D. Electric polarization as a bulk quantity and its relation to surface charge. *Phys. Rev. B* **48**, 4442–4455 (1993).
- Resta, R. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Rev. Modern Phys.* **66**, 899–915 (1994).
- Teter, M. P., Payne, M. C. & Allan, D. C. Solution of Schrödinger's equation for large systems. *Phys. Rev. B* **40**, 12255–12263 (1989).
- Marzari, N., Vanderbilt, D. & Payne, M. C. Ensemble density-functional theory for ab initio molecular dynamics of metals and finite-temperature insulators. *Phys. Rev. Lett.* **79**, 1337–1340 (1997).
- Souza, I., Íñiguez, J. & Vanderbilt, D. Dynamics of Berry-phase polarization in time-dependent electric fields. *Phys. Rev. B* **69**, 085106 (2004).
- Umari, P. & Pasquarello, A. Finite electric field in density functional calculations with periodic boundary conditions. *Comput. Mater. Sci.* **30**, 116–119 (2004).
- Umari, P., Williamson, A. J., Galli, G. & Marzari, N. Dielectric response of periodic systems from quantum monte carlo calculations. *Phys. Rev. Lett.* **95**, 207602 (2005).
- Johnston, W. D. Nonlinear optical coefficients and the Raman scattering efficiency of LO and TO phonons in acentric insulating crystals. *Phys. Rev. B* **1**, 3494–3503 (1970).
- Fornberg, B. Generation of finite difference formulas on arbitrarily spaced grids. *Mathe. Comput.* **51**, 699–706 (1988).

44. Marzari, N. & Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **56**, 12847–12865 (1997).
45. Huber, S. P. et al. Common workflows for computing material properties using different quantum engines. *npj Comput. Mater.* **7**, 136 (2021).
46. Giannozzi, P. et al. Quantum ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condensed Matter* **21**, 395502 (2009).
47. Giannozzi, P. et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys.: Condensed Matter* **29**, 465901 (2017).
48. Giannozzi, P. et al. Quantum ESPRESSO toward the exascale. *J. Chem. Phys.* **152**, 154105 (2020).
49. Veithen, M. & Ghosez, P. First-principles study of the dielectric and dynamical properties of lithium niobate. *Phys. Rev. B* **65**, 214302 (2002).
50. Hermet, P., Veithen, M. & Ghosez, P. First-principles calculations of the nonlinear optical susceptibilities and Raman scattering spectra of lithium niobate. *J. Phys.: Condensed Matter* **19**, 456202 (2007).
51. Margueron, S. et al. Resolved E-symmetry zone-centre phonons in LiTaO<sub>3</sub> and LiNbO<sub>3</sub>. *J. Appl. Phys.* **111**, 104105 (2012).
52. Sanna, S. et al. Raman scattering efficiency in LiTaO<sub>3</sub> and LiNbO<sub>3</sub> crystals. *Phys. Rev. B* **91**, 224302 (2015).
53. Himmetoglu, B., Wentzcovitch, R. M. & Cococcioni, M. First-principles study of electronic and structural properties of CuO. *Phys. Rev. B* **84**, 115108 (2011).
54. Himmetoglu, B., Floris, A., de Gironcoli, S. & Cococcioni, M. Hubbard-corrected DFT energy functionals: the LDA+U description of correlated systems. *Int. J. Quantum Chem.* **114**, 14–49 (2013).
55. Campo Jr, V. L. & Cococcioni, M. Extended DFT+U+V method with on-site and inter-site electronic interactions. *J. Phys.: Condensed Matter* **22**, 055602 (2010).
56. Cococcioni, M. & Marzari, N. Energetics and cathode voltages of LiMPO<sub>4</sub> olivines (M=Fe, Mn) from extended Hubbard functionals. *Phys. Rev. Mater.* **3**, 033801 (2019).
57. Timrov, I., Aquilante, F., Binci, L., Cococcioni, M. & Marzari, N. Pulay forces in density-functional theory with extended Hubbard functionals: from nonorthogonalized to orthogonalized manifolds. *Phys. Rev. B* **102**, 235159 (2020).
58. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.* **118**, 8207–8215 (2003).
59. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Erratum: “Hybrid functionals based on a screened coulomb potential”. *J. Chem. Phys.* **124**, 219906 (2006).
60. Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982–9985 (1996).
61. Ridah, A., Bourson, P., Fontana, M. D. & Malovichko, G. The composition dependence of the Raman spectrum and new assignment of the phonons in LiNbO<sub>3</sub>. *J. Phys.: Condensed Matter* **9**, 9687–9693 (1997).
62. Kojima, S., Kanehara, K., Hoshina, T. & Tsurumi, T. Optical phonons and polariton dispersions of congruent LiNbO<sub>3</sub> studied by far-infrared spectroscopic ellipsometry and Raman scattering. *Japanese J. Appl. Phys.* **55**, 10TC02 (2016).
63. Monacelli, L., Errea, I., Calandra, M. & Mauri, F. Pressure and stress tensor of complex anharmonic crystals within the stochastic self-consistent harmonic approximation. *Phys. Rev. B* **98**, 024106 (2018).
64. Monacelli, L. & Mauri, F. Time-dependent self-consistent harmonic approximation: anharmonic nuclear quantum dynamics and time correlation functions. *Phys. Rev. B* **103**, 104305 (2021).
65. Monacelli, L. et al. The stochastic self-consistent harmonic approximation: calculating vibrational properties of materials with full quantum and anharmonic effects. *J. Phys.: Condensed Matter* **33**, 363001 (2021).
66. Barker, A. S. & Loudon, R. Dielectric properties and optical phonons in LiNbO<sub>3</sub>. *Phys. Rev.* **158**, 433–445 (1967).
67. Siciliano, A., Monacelli, L., Caldarelli, G. & Mauri, F. Wigner Gaussian dynamics: simulating the anharmonic and quantum ionic motion. *Phys. Rev. B* **107**, 174307 (2023).
68. Bastonero, L. AiIDA-Phonopy code v1.0.0. <https://github.com/aiida-phonopy/aiida-phonopy> (2022).
69. Weigel, T. et al. X-ray diffraction using focused-ion-beam-prepared single crystals. *J. Appl. Crystallogr.* **53**, 614–622 (2020).
70. van Setten, M. et al. The PseudoDojo: training and grading a 85 element optimized norm-conserving pseudopotential table. *Comput. Phys. Commun.* **226**, 39–54 (2018).
71. Prandini, G., Marrazzo, A., Castelli, I. E., Mounet, N. & Marzari, N. Precision and efficiency in solid-state pseudopotential calculations. *npj Comput. Mater.* **4**, 72 (2018).
72. Garrity, K. F., Bennett, J. W., Rabe, K. M. & Vanderbilt, D. Pseudopotentials for high-throughput DFT calculations. *Comput. Mater. Sci.* **81**, 446–452 (2014).
73. Schlipf, M. & Gygi, F. Optimization algorithm for the generation of ONCV pseudopotentials. *Comput. Phys. Commun.* **196**, 36–44 (2015).
74. Willand, A. et al. Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations. *J. Chem. Phys.* **138**, 104109 (2013).
75. Corso, A. D. Pseudopotentials periodic table: from H to Pu. *Comput. Mater. Sci.* **95**, 337–350 (2014).
76. Topsakal, M. & Wentzcovitch, R. Accurate projected augmented wave (PAW) datasets for rare-earth elements (RE=La-Lu). *Comput. Mater. Sci.* **95**, 263–270 (2014).
77. Timrov, I., Marzari, N. & Cococcioni, M. Hubbard parameters from density-functional perturbation theory. *Phys. Rev. B* **98**, 085127 (2018).
78. Timrov, I., Marzari, N. & Cococcioni, M. Self-consistent Hubbard parameters from density-functional perturbation theory in the ultrasoft and projector-augmented wave formulations. *Phys. Rev. B* **103**, 045141 (2021).
79. Timrov, I., Marzari, N. & Cococcioni, M. HP - A code for the calculation of Hubbard parameters using density-functional perturbation theory. *Comput. Phys. Commun.* **279**, 108455 (2022).
80. Floris, A. et al. Hubbard-corrected density functional perturbation theory with ultrasoft pseudopotentials. *Phys. Rev. B* **101**, 064305 (2020).
81. Lin, L. Adaptively compressed exchange operator. *J. Chem. Theory Comput.* **12**, 2242–2249 (2016).
82. Bastonero, L. & Marzari, N. Automated all-functionals infrared and Raman spectra. *Mater. Cloud Archive* 2023.118 (2023).

## Acknowledgements

We would like to thank (in alphabetic order) Marnik Bercx, Chiara Cignarella, Virginie de Mestral, Sebastiaan P. Huber, Michele Kotiuga, Eric Macke, Giovanni Pizzi, Norma Rivano, and Iurii Timrov for fruitful discussions. The authors gratefully acknowledge support from the Deutsche Forschungsgemeinschaft (DFG) under Germany’s Excellence Strategy (EXC 2077, No. 390741603, University Allowance, University of Bremen) and Lucio Colombi Ciacchi, the host of the “U Bremen Excellence Chair Program”, as well as computing time granted by the Resource Allocation Board and provided on the supercomputer Lise and Emmy at NHR@ZIB and NHR@Göttingen as part of the NHR infrastructure. The calculations for this research were conducted with computing resources under the project hbc00053 and hbi00059. N.M. acknowledges support by the NCCR MARVEL, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 205602).

### Author contributions

L.B. designed and developed the code, wrote the first draft of the manuscript, ran and analysed the calculations. N.M. supervised the project, discussed the results, and contributed to the final version of the manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at

<https://doi.org/10.1038/s41524-024-01236-3>.

**Correspondence** and requests for materials should be addressed to Lorenzo Bastonero.

**Reprints and permissions information** is available at

<http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024