Electronic Structure

PAPER • **OPEN ACCESS**

AiiDA-defects: an automated and fully reproducible workflow for the complete characterization of defect chemistry in functional materials

To cite this article: Sokseiha Muy *et al* 2023 *Electron. Struct.* **5** 024009

View the article online for updates and enhancements.

You may also like

- ON THE POSSIBILITY OF HOLOMORPHIC EXTENSION, INTO A DOMAIN, OF FUNCTIONS DEFINED ON A CONNECTED PIECE OF ITS BOUNDARY. II
  L A Azenberg and A M Kytmanov
- INTEGRAL REPRESENTATIONS OF DIFFERENTIAL FORMS ON CAUCHY-RIEMANN MANIFOLDS AND THE THEORY OF CR-FUNCTIONS. II
  R A Arapetyan and G M Khenkin
- Condensation of the stimulated emission spectrum of broad-band lasers using narrow gain lines
  S A Kovalenko and S P Semin

This content was downloaded from IP address 128.179.166.202 on 17/07/2023 at 15:12

IOP Publishing Electron. Struct. **5** (2023) 024009 https://doi.org/10.1088/2516-1075/ace014

Electronic Structure

CrossMark

OPEN ACCESS

RECEIVED
2 March 2023

REVISED
2 June 2023

ACCEPTED FOR PUBLICATION
20 June 2023

PUBLISHED
30 June 2023

Original Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

AiiDA-defects: an automated and fully reproducible workflow for the complete characterization of defect chemistry in functional materials

Sokseiha Muy[1,3], **Conrad Johnston**[1,2,3] and **Nicola Marzari**[1,*]

[1] Theory and Simulations of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
[2] Catalysis Science, Pacific Northwest National Laboratory PNNL, Richland, WA 99354, United States of America
[3] These two authors contributed equally to this work.
[*] Author to whom any correspondence should be addressed.

E-mail: nicola.marzari@epfl.ch

Keywords: point defects, sutomated workflow, AiiDA, defect formation energy, Li-ion conductors

Abstract

Functional materials that enable many technological applications in our everyday lives owe their unique properties to defects that are carefully engineered and incorporated into these materials during processing. However, optimizing and characterizing these defects is very challenging in practice, making computational modelling an indispensable complementary tool. We have developed an automated workflow and code to accelerate these calculations (AiiDA-defects), which utilises the AiiDA framework, a robust open-source high-throughput materials informatics infrastructure that provides workflow automation while simultaneously preserving and storing the full data provenance in a relational database that is queryable and traversable. This paper describes the design and implementation details of AiiDA-defects, the models and algorithms used, and demonstrates its use in an application to fully characterize the defect chemistry of the well known solid-state Li-ion conductors $LiZnPS_4$. We anticipate that AiiDA-defects will be useful as a tool for fully automated and reproducible defect calculations, allowing detailed defect chemistry to be obtained in a reliable and high-throughput way, and paving the way toward the generation of defects databases for accelerated materials design and discovery.

1. Introduction

Point defects in semiconductors, insulators and metals have enormous influences on the mechanical, transport, electronic, and optical properties of materials [1]. Many materials which have found important technological applications largely owe their functionalities to the defects that were carefully selected and introduced into the pure materials during their synthesis. There is a vast literature on defect engineering to achieve specific functionalities in various functional materials, such as light emitting diodes [2], photovoltaics [3, 4], thermoelectrics [5], solid-state Li-ion conductors [6, 7] and the defect centers for quantum sensing [8, 9] to name just a few.

Several experimental techniques have been developed to characterize point defects in materials, such as electron paramagnetic resonance [10] photoluminescence spectroscopy [11], positron annihilation spectroscopy [12], electron microscopy [13] and deep-level transient spectroscopy [14]. However, these techniques are very time-consuming and challenging to carry out in day-to-day research, making computational characterizations a valuable complementary approach. Within the realm of computational materials science and computational chemistry, there are a growing number of packages and libraries offering some degree of workflow automation, such as ASE [15], AFLOW [16], Fireworks [17], QMflows [18] as part of domain-specific subset of the larger family of automation and parallel computing frameworks of interest to a general scientific computing audience [19–21]. There are also a number of packages addressing different aspects of defect chemistry, especially the computation of defect formation energy

© 2023 The Author(s). Published by IOP Publishing Ltd

corrections, and sometimes for the case of charged defects. These includes packages such as PyCDT [22], pylada-defects [23], CoFEEE [24], and PyDEF [25] and DASP [26]. Typically, these employ post-, and sometimes pre-, processing scripts to address specific aspects of defect chemistry calculations and are often designed with a particular quantum engine in mind, interacting with the chosen one either in a manual or automatic way. The number of high-quality efforts in this direction demonstrates the demand to have these complex tasks automated and abstracted away from computational scientists' workload.

AiiDA [27, 28] is a computational infrastructure designed for high-throughput computation and automation. Like some of the other workflow frameworks, it is an open-source project written in Python; crucially, the automation tasks are abstracted using a scalable workflow engine which runs and monitors workflow steps concurrently, both locally and on high-performance computing resources, while simultaneously capturing all of the results and the calculation provenance in directed acyclic graphs to ensure full reproducibility of results. A key strength of AiiDA is its plugin system. This allows for any software or system to be interfaced with the core AiiDA functionalities, with each plugin serving as an interface to the API provided by the AiiDA framework. Thanks to an enthusiastic community of plugin developers, a considerable number of scientific codes, databases and workflows are already supported including 'common' workflows to calculate the same property with different quantum engines [29].

AiiDA-defects has been developed as a plugin for AiiDA with the goal of making the analysis of defect chemistry as seamless and reliable as possible, and to help accelerate the discovery and design of new engineered functional materials. Indeed, one of the strengths of AiiDA-defects consists in providing the automatic characterization of the complete defect chemistry of a material in a thermodynamically consistent framework, rather than just a single defect calculation, as well as providing full reproducibility of the results by leveraging the AiiDA infrastructure.

To illustrate the capabilities of AiiDA-defects, we apply to study defects in the Li-ion conductor $LiZnPS_4$, which has been proposed as a solid electrolyte for the next generation of all-solid-state Li-ion batteries [30, 31].Defects in materials used for solid-state electrolytes are particularly interesting as in their pure stoichiometric form these materials typically exhibit very low ionic conductivity. In order to reach appreciable and useful ionic conductivity, many Li-conductors require careful optimization strategies to introduce vacancies or interstitial Li into the lattice. This is typically achieved via aliovalent doping, where non-mobile ions are replaced with other ions of different charge, and it is assumed that charge compensation is driven by changing the concentration of Li to maintain the overall charge neutrality. However, for a given charge of the aliovalent dopant, there is no strategy to decide which ion should be chosen from the periodic table. Moreover, it is not guaranteed that changing Li concentration is always the main charge compensating mechanism, as there are potentially other defects—anti-sites, vacancies on the non-Li sites, . . .—that can also lead to a charge neutral system [32]. In order to optimize the type of defects and their concentrations, one needs to have access to the formation energies of various types of defects under given experimental conditions, such as temperature, pressure and chemical potentials of the different species. The required calculations are often resource demanding and tedious because of the large number of pre- and post-processing steps involved and the large number of calculations that are required to fully characterise the defect chemistry of the material, highlighting the need for automated workflows that can seamlessly and robustly handle these complex protocols.

2. Background and methods

A common need in the design of materials is a knowledge of the thermodynamic defects' concentrations. To estimate these, one needs first to obtain the formation energy of a given defect type. The formalism for the defect formation energy (in the dilute limit) is well established [33, 34]. Conventionally, it can be calculated using the following expression [34]:

$$E^f[X^q] = E_{\mathrm{tot}}[X^q] - E_{\mathrm{tot}}[\mathrm{bulk}] - \sum_i n_i\mu_i + qE_{\mathrm{F}} + E_{\mathrm{corr}} \quad (1)$$

where $E^f[X^q]$ is the formation energy of defect X in the charge state q, $E_{\mathrm{tot}}[X^q]$ is the energy of a supercell containing the defect (typically, the energy obtained from density functional theory (DFT) calculations), $E_{\mathrm{tot}}[\mathrm{bulk}]$ is the energy of the pristine supercell without the defect, n_i are the number of atoms of type i added ($n_i > 0$) or removed ($n_i < 0$) from the supercell to create the defect, μ_i is the chemical potential of atomic species i, E_{F} is the Fermi level relative to the maximum of the valence band and E_{corr} is the post-processing correction term added to remove the spurious long-range electrostatic interactions between the charged defect and its periodic images as well as with the uniform background charge [35–37]. In the AiiDA-defects package, each of these terms is computed automatically in a workflow, called a workchain' in AiiDA terminology, which is controlled by the AiiDA workflow management engine [27, 28]. The standard steps of running a DFT calculation are handled seamlessly and automatically—from the preparation of input files to the retrieval of results from a remote resource—with the provenance being captured in a fully queryable database.

In the following sections, we will describe in more detail the computational and practical considerations when computing defect formation energies, and detail the implementation of each workchain inside the package, and illustrate its use in the determination of the defect chemistry of the Li-ion conductor $LiZnPS_4$.

2.1. Periodic-image correction

In equation (1), we stated how the defect formation energy for a given defect could be calculated. The most important assumption underling this equation is that the defect is in the 'infinitely dilute limit', such that the interactions between periodic images of the defect can be neglected and the defect can be considered as isolated. However, most DFT codes used periodic-boundary conditions, which leads to spurious interactions between the defect and its periodic images. To simulate an isolated defect, one can use a supercell that is large enough such that any interaction between the defect and its periodic images has become negligible. However, a problem that remains is that generally DFT suffers from very unfavourable scaling, typically be $\mathcal{O}(n^3)$, making this this approach impractical for realistic systems. To overcome this, it was recognised that the interaction between defects images could be thought of as arising from different contributions—that from the electronic wavefunctions overlapping, that from electrostatic interactions and that from the long-range elastic fields of the defect. Assuming these latter have become negligible in the supercell chosen, and that the electronic density of the defect state is well localised, it can be assumed that the interaction is solely from electrostatics and so model systems can be constructed with which to estimate the size of this interaction to provide an *a posteriori* correction.

The energy of such a model is given by:

$$E = -\frac{1}{2}\int_{\mathrm{cell}} d^3\boldsymbol{r}\rho(\boldsymbol{r})\nabla^2\nu(\boldsymbol{r}) \tag{2}$$

where the electrostatic potential $\nu(\boldsymbol{r})$ for a model system with a model charge distribution $\rho(\boldsymbol{r})$ is obtained by solving the Poisson equation [37, 38]:

$$\nabla^2\nu(\boldsymbol{r}) = -\frac{4\pi}{\epsilon}\rho(\boldsymbol{r}) \tag{3}$$

where ϵ is the permittivity of the host material.

For the case of periodic boundary conditions, the potential is evaluated in reciprocal space as in:

$$\nu(\boldsymbol{G}) = 4\pi\sum_{\boldsymbol{G}\neq 0}\frac{\rho(\boldsymbol{G})}{\epsilon|\boldsymbol{G}^2|}. \tag{4}$$

To compute the electrostatic correction for a given supercell size, the difference in energy between that of model system of identical size to the DFT supercell, and an infinitely extended supercell (either extrapolated from multiple models of differing size under periodic boundary conditions, or a model with open boundary conditions) needs to be calculated. The remaining choice is which charge distribution to use for the model. A number of schemes have been proposed and these can be classified into three groups [37, 39, 40]: point countercharge corrections, such as the Makov-Payne scheme, that use a point charge in place of the charged defect [41], Gaussian countercharge corrections, that use a Gaussian type function as the model charge [35–37], and density countercharge corrections, that use the charge density difference derived from DFT [37, 42]. It is also possible to directly apply these corrections in a DFT self-consistent cycle, rather than *a posteriori* by adding a correction term to Kohn–Sham equation in the procedure to optimise the DFT wavefunctions [39, 43, 44]. In AiiDA-defects we use an *a posteriori* correction as this allows us to remain agnostic to the choice of any particular quantum code; in fact, recent efforts [45] demonstrated how AiiDA can be utilized in code-agnostic form, using the example of the computation of the equation of state to show how a common workflow can be coupled to many different codes through a set of simple interfaces. Here we also embrace this model, and recognise that the application of a charge correction scheme *a posteriori* could support any code through simple interfaces, thanks also to the fact that many solid-sate codes already have support in AiiDA for the required single-point energy calculations [29].

We implement and make available two options to specify the parameters of the Gaussian charge density. The default scheme is to a fit a multivariate Gaussian to the defect charge density, defined as the difference between the DFT charge density of the material with a defect in a charge state q and that of the pristine bulk:

$$\rho_{\text{defect}}(\boldsymbol{r}) = \rho^q_{\text{defect}}(\boldsymbol{r}) - \rho_{\text{bulk}}(\boldsymbol{r}). \tag{5}$$

The probability density function for a multivariate Gaussian in three-dimensions is given by:

$$f(x) = \frac{1}{\sqrt{(2\pi)^3 \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right) \tag{6}$$

where Σ is the symmetric positive definite covariance matrix, and μ is the mean. The fitting minimises the difference between the charge density difference, adjusting the six parameters of the covariance matrix and the three-coordinate position of the mean, representing the centre of the charge density difference distribution. The fitting automates the choice of the Gaussian width, which is a crucial consideration to enable materials and defects to be be studied in a high throughput mode. Using a multivariate Gaussian allows for anisotropy of the model charge distribution and a better fit to the DFT charge difference. If a defect charge distribution is complex, e.g. displaying degenerate peaks, the charge fitting may not converge well and so the user can opt for the 2nd option and specify the parameters of the covariance matrix directly. This also allows for an isotropic Gaussian to be specified, similar to the Freysoldt-Neugebauer-Van de Walle type corrections [35, 36].

In equation (4), the charge density is screened by the relative permittivity ε. The user may specify this value or it will be automatically calculated using density functional perturbation theory [46]. In practice, the full tensorial value is used and so equation (4) is re-written as:

$$\nu(\boldsymbol{G}) = 4\pi \sum_{\boldsymbol{G}\neq 0} \frac{\rho(\boldsymbol{G})}{\boldsymbol{G}\cdot\boldsymbol{\epsilon}\cdot\boldsymbol{G}} \tag{7}$$

where $\boldsymbol{\epsilon}$ is a 3×3 matrix describing the relative permittivity in three dimensions. The user may specify a full matrix or a single value.

With these ingredients, the energy for a given size of the model system can be computed and compared to the energy extrapolated to infinity from a series of increasingly large model systems. In the case of the model charge being represented by an isotropic Gaussian, this extrapolation can be avoided since the analytical solution of the Poisson's equation of a Gaussian charge density with open-boundary condition is known and can be used directly to obtain the isolated model energy:

$$E = \frac{q^2}{2\epsilon\sigma\sqrt{\pi}}. \tag{8}$$

This shortcut is applied automatically if an isotropic Gaussian is supplied by the user or is found by fitting. We note in passing that these considerations apply most straightforwardly to bulk 3d materials; in the case of low-dimensional structures, extra care needs to be paid in reaching the thermodynamic limit across the poorly screened vacuum regions.

2.2. Potential alignment

In section 2.1, we showed how the interaction between periodically repeated images of the charged defect can be corrected. However, there is another issues arising from the implicit use of a neutralizing background when simulating a system with an overall net charge in periodic-boundary condition. This background charge compensates for the system charge to ensure overall charge neutrality, but also introduces a further spurious interaction with the defect charge [35, 47].

For an isolated defect in open-boundary condition, the electrostatic potential of defect $V_{q/b} = V_{\text{defect}} - V^q_{\text{bulk}}$ decays to zero far from the defect. However, in periodic-boundary condition, $V_{q/b}$ will plateau at a value different from zero. The offset of this plateau from zero has to be determined to reproduce the absolute position of the defect potential and constitutes the alignment term $\Delta V_{q/b-m}$ introduced by Freysoldt, Neugebauer and van de Walle [35, 36]:

$$E_{\text{corr}} = E_{\text{lat}} - q\Delta V_{q/b-m} \tag{9}$$

where E_{lat} is the correction for the spurious electrostatic interaction between the images. To determine the value of this plateau, one separates the defect potential $V_{q/b}$ into a long-range and a short-range contribution. The long-range part is approximated by a screened Coulomb potential created by a Gaussian charge distribution V^q_{model} as described in the previous section:

$$V_{q/b} \equiv V_{lr} + V_{\text{sr}} \approx V^q_{\text{model}} + V_{\text{sr}}. \tag{10}$$

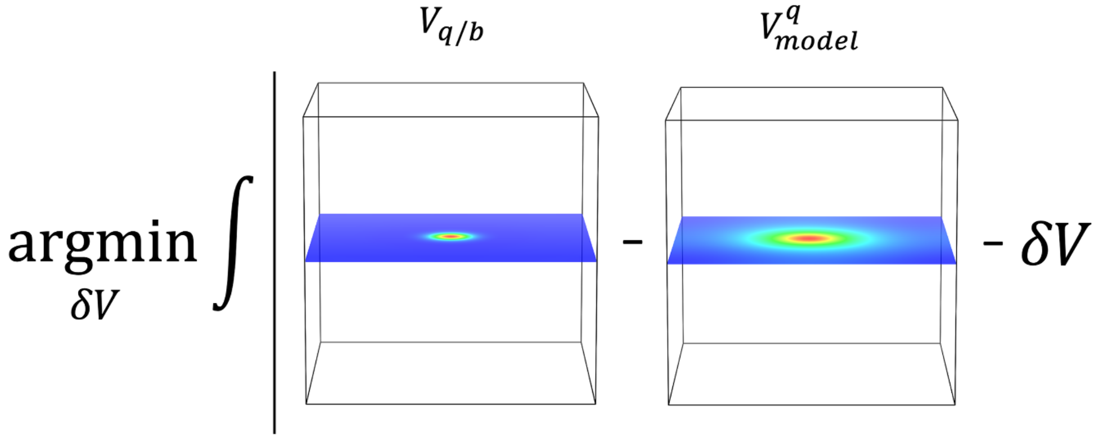

Figure 1. In AiiDA-defects, the alignment term $\Delta V_{q/b-m}$ is obtained by computing the offset the defect short-range potential $V_{\mathrm{sr}} \equiv V_{q/b} - V^q_{\mathrm{model}}$ directly on the 3D grid and not from the planar-average. The values of the potential in the plane containing the defect site (at the center of the cell) are shown.

Since $V_{q/b}$ is known and V^q_{model} can be easily computed by solving the Poisson's equation, V_{sr} can be obtained and its value far from the defect is the offset $\Delta V_{q/b-m}$ that we seek to determine.

Previously, this 'alignment' was done 'by eye' in one dimension, the typical procedure being that of computing a planar average over two cell vectors, so that a three dimensional potential can be reduced to a single dimension. The two potentials $V_{q/b}$ and V^q_{model} are then plotted and their difference will reach a plateau at some distance sufficiently far away beyond the defect's range of interaction. The potential alignment term is then read from the value of that plateau. This procedure can be prone to error and subjectivity; to allow for automation and to avoid visual inspection, we approach this differently. Rather than computing a planar average, we retain the 3-dimensional nature of the potentials for the supercell. Since we want to compute the alignment at as large a distance as possible from the defect, grid points which are closer than half the distance from the defect to its closest periodic image, as per the minimum image convention, are considered to be too close to the defect center and are discarded. With the remaining points, the alignment term is computed by minimizing the following objective function:

$$\Delta V_{q/b-m} = \underset{\delta V}{\arg\min} \int_{\bar{\Omega}} \left| \left[V_{q/b}(\boldsymbol{r}) - V^q_{\mathrm{model}}(\boldsymbol{r}) \right] - \delta V \right| \mathrm{d}\boldsymbol{r} \tag{11}$$

where $\bar{\Omega}$ corresponds to the region in the supercell where $|\boldsymbol{r} - \boldsymbol{r_d}| > L/2, \boldsymbol{r_d}$ being the position of the defect in the supercell. This is fast and automatic, and avoids any ambiguities that may occur from averaging and fluctuations in the data (figure 1). If needed, one could also consider to smooth this sharp boundary, substituting for step function a smoothed one [48].

2.3. Chemical potential of ions

The chemical potential is associated with the energy cost for exchanging an atom between the materials and the surrounding thermodynamic reservoir. Its value reflects the synthesis conditions under which the materials made, and the value of the chemical potential μ_i of atom i in a given compound is constrained not only by the stability of that compound but also that of all the competing phases in the phase diagram. To illustrate this idea, following Buckeridge *et al* [49], we consider a compound $A_{m_1}B_{n_1}C_{p_1}$ in an A-B-C ternary system. We have the following conditions:

$$\mu_A \leqslant 0; \quad \mu_B \leqslant 0 \quad \text{and} \quad \mu_C \leqslant 0 \tag{12}$$

$$m_1\mu_A + n_1\mu_B + p_1\mu_C = \mu_{A_{m_1}B_{n_1}C_{p_1}} \simeq \Delta H_f(A_{m_1}B_{n_1}C_{p_1}). \tag{13}$$

The conditions (12) ensure that A, B and C form the compound $A_{m_1}B_{n_1}C_{p_1}$ instead of precipitating to pure element A, B and C. Equation (13) is simply the definition of the Gibbs formation energy of the compound and where we have neglected the entropic contribution and approximate $\mu_{A_{m_1}B_{n_1}C_{p_1}}$ by its formation enthalpy

ΔH_f. For solids, a further approximation can be made by neglecting the PV enthalpic contribution, which is typically small for condensed phases, and replacing formation enthalpies by formation energies, which are computed from DFT. In practice, care must be taken to ensure that these formation energies are computed with the same level of theory and numerical accuracy as those of $E_{\mathrm{tot}}[X^q]$ and $E_{\mathrm{tot}}[\mathrm{bulk}]$ in equation (1), i.e, using the same exchange-correlation functionals and plane-wave cutoffs and ensuring that k-point sampling is sufficient to converge these energy differences across different phases (e.g. metal vs insulator).

For other compounds in the A-B-C system, for example $A_{m_2}B_{n_2}C_{p_2}$, we will also have:

$$m_2\mu_A + n_2\mu_B + p_2\mu_C \leqslant \Delta H_f(A_{m_2}B_{n_2}C_{p_2}). \tag{14}$$

This inequality ensures that the formation of compound under consideration i.e. $A_{m_1}B_{n_1}C_{p_1}$ is favored over the formation of competing phase $A_{m_2}B_{n_2}C_{p_2}$. If other compounds exist in the A-B-C systems, one can generate other inequalities similar to (14) for each of those compounds. Together with equations (12) and (14), these equations and inequalities form a system of linear constraints in μ_A, μ_B and μ_C which can be solved to determine the stability region of $A_{m_1}B_{n_1}C_{p_1}$ from which the chemical potentials of A, B and C can be chosen for the calculation of the defect formation energy in equation (1). This example with ternary system can be generalized straightforwardly to any binary or multinary compound. As a concrete example, we will determine the stability region of $\mathrm{Li_3PO_4}$, out of which the chemical potential of Li μ_{Li} (which is needed in equation (1)) can be used for the calculation of formation energy of Li-related defects, such as lithium vacancies or interstitials. Applying equations (12) and (14) to $\mathrm{Li_3PO_4}$, we obtain:

$$\mu_{\mathrm{Li}} \leqslant 0; \quad \mu_{\mathrm{P}} \leqslant 0 \quad \text{and} \quad \mu_{\mathrm{O}} \leqslant 0 \tag{15}$$

$$3\mu_{\mathrm{Li}} + \mu_{\mathrm{P}} + 4\mu_{\mathrm{O}} = -22.164\ \mathrm{eV\,fu^{-1}} \tag{16}$$

where −22.164 eV is the formation energy per formula unit (fu) (not to be confused with the defect formation energy) of $\mathrm{Li_3PO_4}$. Note that thanks to equation (16) only two of the three chemical potentials are independent. We can therefore choose one of the elements, e.g. phosphorus as the 'dependent' element whose chemical potential is fixed once the chemical potential of the 'independent' elements are chosen. In addition to $\mathrm{Li_3PO_4}$, other stable phases in the Li–P–O system consist of $\mathrm{LiPO_3}$, $\mathrm{Li_4P_2O_7}$, $\mathrm{P_2O_5}$, $\mathrm{Li_2O}$, $\mathrm{LiO_8}$, $\mathrm{Li_2O_2}$, LiP, $\mathrm{Li_3P}$, $\mathrm{LiP_7}$ and $\mathrm{Li_3P_7}$. For each of these competing phases, we can write down a constraint similar to equation (14) which will give respectively:

$$\mu_{\mathrm{Li}} + \mu_{\mathrm{P}} + 3\mu_{\mathrm{O}} \leqslant -13.614\ \mathrm{eV\,fu^{-1}} \tag{17}$$

$$4\mu_{\mathrm{Li}} + 2\mu_{\mathrm{P}} + 7\mu_{\mathrm{O}} \leqslant -35.901\ \mathrm{eV\,fu^{-1}} \tag{18}$$

$$\ldots \tag{19}$$

$$3\mu_{\mathrm{Li}} + 7\mu_{\mathrm{P}} \qquad \leqslant -3.605\ \mathrm{eV\,fu^{-1}}. \tag{20}$$

By solving these constraints, the stability region of $\mathrm{Li_3PO_4}$ can be determined and is shown as the gray area in figure 2. Any chemical potential in this region can be used to compute the defect formation energy in equation (1) and by default, the chemical potential corresponding to the centroid of the stability region (the gray dot) is chosen. Note that, so far, all the chemical potentials are *relative* chemical potentials in the sense their values are referred to the reference state of each element for which the chemical potentials are set to zero. However, the chemical potentials as appear in equation (1) are *absolute* chemical potentials which can be obtained from the relative chemical potentials by adding the energy of the corresponding element in its reference state. One complication that should be mentioned is the case where the compound under consideration is predicted to be unstable; i.e. its formation energy is above the convex hull. In this instance, the formation energy of that compounds is simply shifted downward until it reaches the convex hull and a warning message is issued.

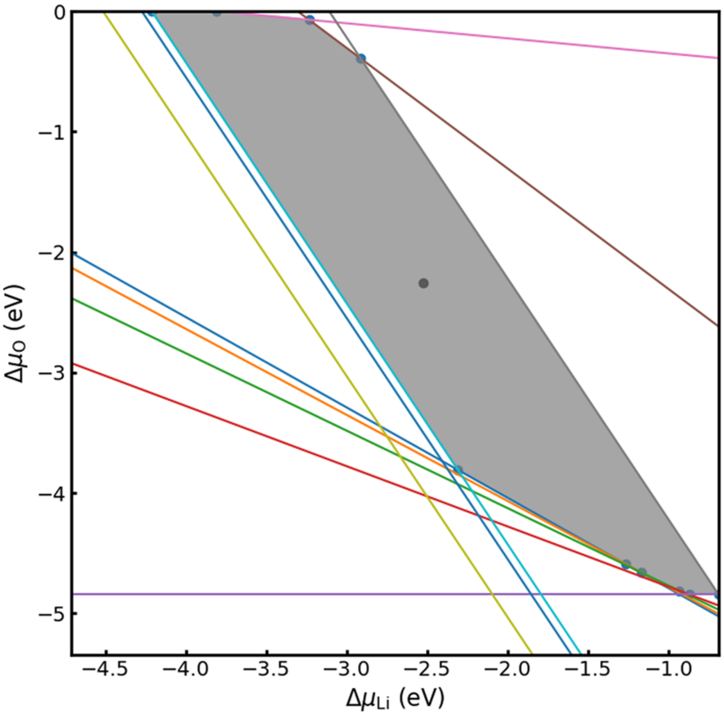

Figure 2. Computed stability region of Li_3PO_4 shown in gray as well as the centroid of the stability region (shown as the central black dot) which used by default in the calculation of the defect formation energies.

2.4. Self-consistent Fermi level

At thermodynamic equilibrium, in the presence of multiple defects with possibly multiple charge states, the Fermi level E_F can be determined self-consistently by imposing a condition of charge neutrality [50, 51]. The algorithm to compute this self-consistent Fermi level is shown schematically in figure 3. More precisely, if we denote n_o as the concentration of free electrons, p_o the concentration of free holes and C_{X^q} the concentration of defect X in the charge state q, the charge neutrality condition requires that:

$$n_o - \sum_X \sum_q qC_{X^q} = p_o \tag{21}$$

where:

$$n_o = \int_{E_\mathrm{F}}^{\infty} f_e(E)\rho(E)\mathrm{d}E \tag{22}$$

$$p_o = \int_{-\infty}^{E_\mathrm{F}} f_h(E)\rho(E)\mathrm{d}E \tag{23}$$

$$C_{X^q} = N_\mathrm{site}\exp\left(-\frac{E^f[X^q]}{kT}\right) \tag{24}$$

and where f_e is the Fermi–Dirac distribution for the electrons, $f_e(E) = [1+\exp((E-E_\mathrm{F})/kT)]^{-1}$, $f_h = 1 - f_e$ is that for the holes and $\rho(E)$ the density of states per unit volume of the pristine host. Each term n_o, p_o and C_{X^q} depends on E_F and therefore the condition (21) constitutes a non-linear equation in E_F which can be solved numerically to determine the value of Fermi level that corresponds to a neutral system. Notice that, by using the density of states $\rho(E)$ of the perfect bulk, we implicitly used the so-called rigid band approximation which essentially states that the presence of the defects does not affect the electronic bandstructure of the host but simply shifts its Fermi level.

Having this algorithm implemented allows also for the possibility to fix the defect concentrations at certain values and compute the corresponding Fermi level. This feature is useful in a situation where we would like to simulate the ‘quenching’ of the system from high temperature, in which the defect concentrations are not the equilibrium concentrations at room temperature, but are rather those that are quenched from high temperature. In this scenario, one first computes the self-consistent Fermi level, and the concentration of defects and free electrons and holes self-consistently at a higher temperature and then fixes

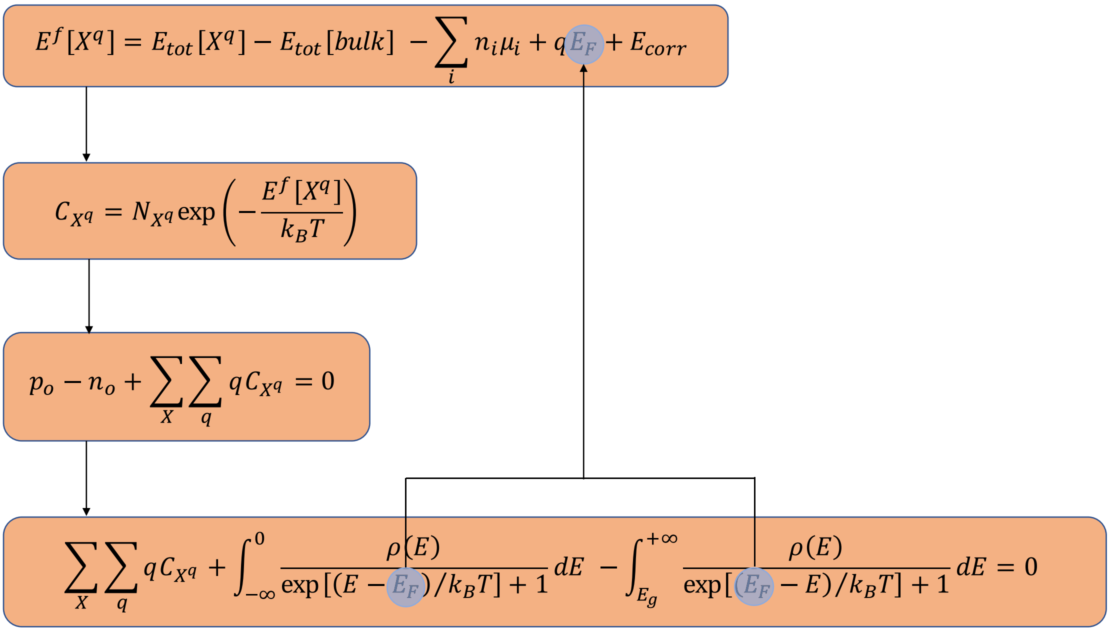
$E^f[X^q] = E_{tot}[X^q] - E_{tot}[bulk] - \sum_i n_i\mu_i + qE_F + E_{corr}$
$C_{X^q} = N_{X^q}\exp\left(-\frac{E^f[X^q]}{k_BT}\right)$
$p_o - n_o + \sum_X\sum_q qC_{X^q} = 0$
$\sum_X\sum_q qC_{X^q} + \int_{-\infty}^{0}\frac{\rho(E)}{\exp[(E-E_F)/k_BT]+1}dE - \int_{E_g}^{+\infty}\frac{\rho(E)}{\exp[(E_F-E)/k_BT]+1}dE = 0$

Figure 3. Algorithm used in this work to compute the Fermi level via a self-consistency cycle. Starting with an estimate of the Fermi level, the defect formation energy can be computed and, using this, a defect concentration can be estimated. By enforcing the charge neutrality condition, the concentration of electrons and holes can be computed, and from this the Fermi energy can be updated. The cycle is closed and new defect formation energies are computed. This continues until the change in E_F is sufficiently small and self-consistency is achieved.

the concentration of defects but allows the Fermi level and the free electron/hole concentrations to adjust at lower temperature. This is useful for example to assess the influence of the cooling rate in high-temperature synthesis on the electronic conductivity of the materials at lower temperatures [52, 53]. Another situation in which this type of self-consistent calculation of the Fermi level can be used is to study the response of the system in terms of the change in the concentration of its native defects as a result of an *effective* aliovalent doping where only the charge and the concentration of the dopants are relevant. In this approximation, the defect $\mathrm{Li}_{\mathrm{Mg}}^{x}$ or $\mathrm{S}_{\mathrm{Cl}}^{x}$ will produce exactly the same effects as a generic defect with $a-1$ charge at the same concentration.

3. Package implementation and structure

As previously introduced in section 1, a key feature of the AiiDA framework is the concept of workchains, representing a specific workflow, as a unit of computational work. Similarly to a subroutine, a workchain accepts inputs, follows an internal logic, and returns outputs. However, unlike standard Python scripting, the inputs and outputs are captured and stored in the AiiDA database, as are the relations between these data nodes. Also, if a particular computational step in the workflow requires a calculation done by external codes, e.g. computing the DFT energy, this can also be implemented as a workchain with defined inputs and outputs via one of the numerous AiiDA plugins that interface with some of the most commonly used quantum engines [29]. Currently, we use Quantum ESPRESSO as our quantum engines for DFT calculations. However, thanks to the recent effort to create a common, code-agnostic workflow to compute well-defined material properties, automated and robust AiiDA protocols for DFT calculations were implemented for eleven quantum engines [45], making AiiDA-defects easily extendable to other DFT and quantum chemistry packages. A step in a workchain may also invoke another workchain (a child workchain) and so workchains may be specialised to a specific modular task and then nested to create complex logic. The AiiDA workflow engine is able to run all the workchains, managing each task, concurrently where appropriate, and storing intermediate results so that even if the machine running AiiDA is rebooted, the project can be resumed from the last successful step.

To best leverage this sophisticated functionality, AiiDA-defects has been implemented with this modularity in mind. Each of the main steps in computing a defect formation energy is done by its own workchain. Where there is a choice of specific method for a given task, we have implemented a parent workchain for the task which is generic, and a child workchain which implements a specific method. This design choice allows for any new method to be 'plugged in' by adding a workchain self-contained in a single

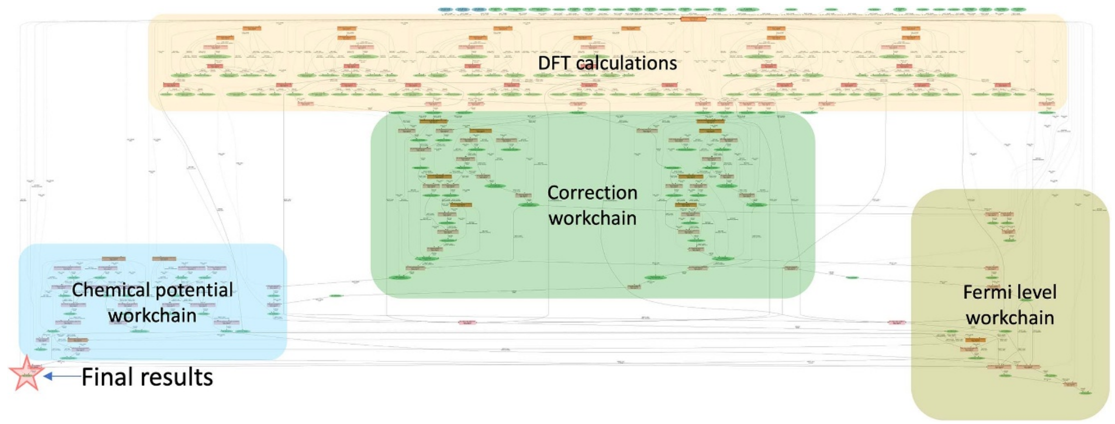

Figure 4. A provenance graph of the *DefectChemistry* workchain as implemented in the AiiDA-defects package. The groups of nodes belonging to different sub-workchains which compute each term in the definition of defect formation energy are shown in boxes with the corresponding name of the workchain.

file, and changing a minimal amount of code in the parent workchain in order to make it available as an option. At the heart of the package is the *FormationEnergy* workchain which takes as input defect configurations generated by users either manually or by means of other packages such as pymatgen [54] or ShakeNBreak [55] just to name a few. The latter allows users to generate plausible low-symmetry configurations of defects which can be missed from conventional local minimization techniques by applying targeted bond distortions and rattling. This workchain computes the formation energy of a given defect from the components discussed in section 1. The total energy of the supercells with and without the defect are computed using DFT as implemented in Quantum ESPRESSO package [56, 57]. For the charged defect, because of the use of periodic boundary condition, a background charge is implicitly introduced to maintain charge neutrality and avoid the divergence of the total energy. The spurious interaction between the charged defect and this background charge is then removed using the correction scheme described in section 2.2. In turn, it relies on the *Correction* workchain to be able to correct the spurious electrostatic interactions. This workchain is in itself generic and a specific choice of correction scheme must be used. Currently, only Guassian CounterCharge method with the potential alignment functionality described in sections 2.1 and 2.2 are implemented, but thanks to the modularity of the package, other correction schemes can be easily implemented as child workchains and will be added in the future. The chemical potential of the defect is required to compute the formation energy, so this was implemented as a separate workchain from the *FormationEnergy* workchain, and its result is passed to the *FormationEnergy* workchain. The final component is the Fermi level. However, as discussed previously, this must be self-consistent across all defects of interest to be meaningful. To compute a self-consistent-Fermi level, we must have knowledge of all of the defects we want to study and be able to compute their formation energies, and so it follows that the *FormationEnergy* workchain must be called repeatedly by the *FermiLevel* workchain. To manage all of these interdependent workchains, we built a master workchain called the *DefectChemistry* workchain. While in principle a user may interact with any workchain directly, it is intended that this top-most workchain be the main interface. With this workchain, the full gamut of defect chemistry calculations can be carried out by AiiDA-defects automatically. The resulting provenance graph from an example run is shown in figure 4 and serves to show how much automation is achieved. Each node in the provenance graph represents a data or a function performing intermediate steps in the workchain while the edges between nodes indicate whether a data is an input to a function node or is produced as an output of that function. The nodes corresponding to different child workchains are grouped and highlighted in the corresponding boxes. The provenance graph can capture and explicitly display the inter-relations between data that required or produced at any steps of the workflow. Moreover, it should be stressed that the reproducibility is insured; without such automatic provenance capture, it would be near impossible to describe specifically what steps and calculations are behind the final data for all but the most simplistic cases. This shows the potential of AiiDA-defects as a versatile and powerful tool for the study of defects, and how it can be used to study such defects in a high-throughput manner. While high-throughput generally refers to high-throughput of materials to achieve some computational screening for a desired property, here we refer to it as high-throughput with respect to specific defects and charge states, allowing one to rapidly and completely characterize a material's defect chemistry. Finally, while using the package, users should keep in mind the limitations of the underlying set up and theory. The first one is the requirement that there is no explicit interaction between defects i.e. the

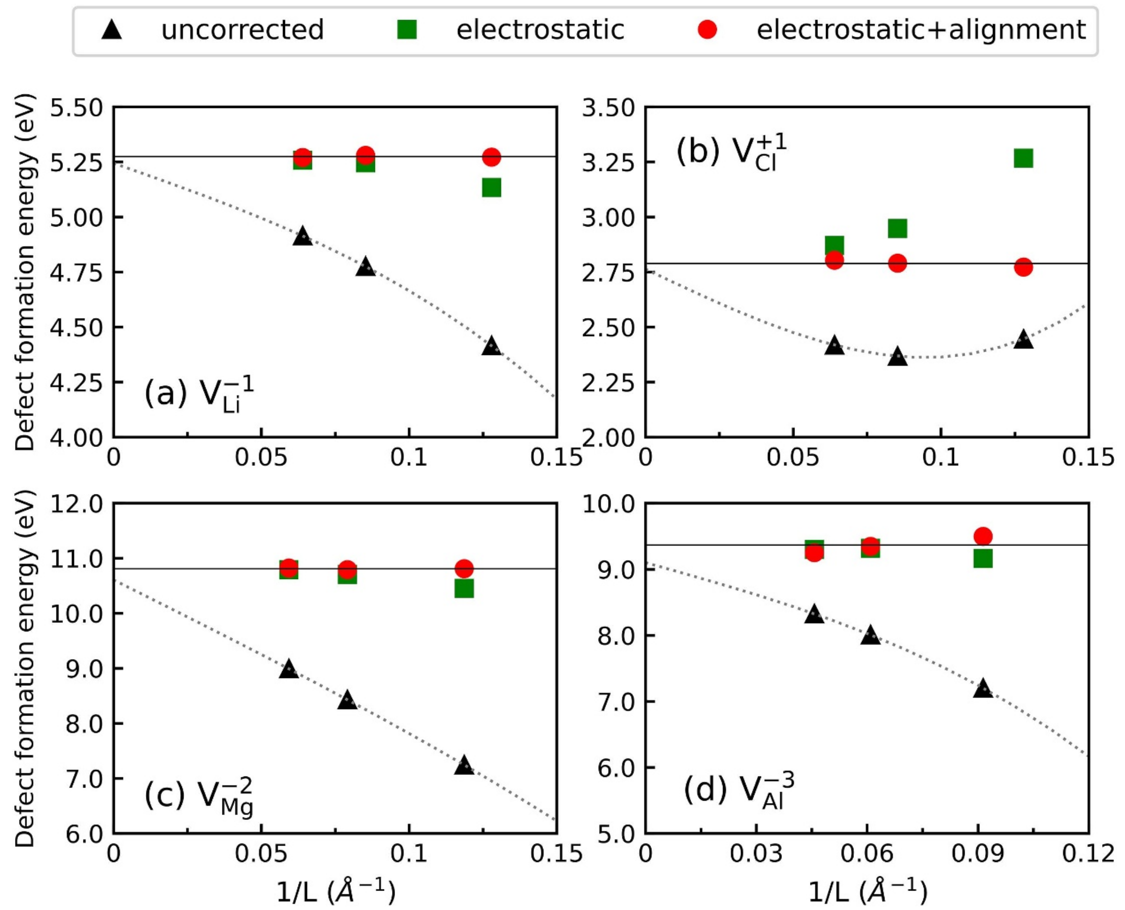

Figure 5. Formation energies of (a) v_{Li}^{-1} (b) v_{Cl}^{+1} in the antiperovskite Li_3ClO (c) v_{Mg}^{-2} in MgO and (d) v_{Al}^{-3} in AlP as a function of the size of the supercell.

need to reach the infinitely dilute limit. The second limitation lies in the so-called rigid band approximation, which is implicitly assumed in the calculation of the Fermi level. In this approximation, one assumes that the presence of the defect does not alter the electronic band structure (or density of states) of the host materials, but simply shifts the Fermi level—this can also be verified *a posteriori* to holds broadly and is consistent with the infinitely dilute limit assumption. Finally, and related to the calculations of chemical potentials, when the host material is a high-temperature phase which is only metastable at 0 K (i.e. its energy above the convex hull is positive), we simply downshift its formation energy by the same amount so that it is now on the convex hull and its stability region can be determined. AiiDA-defects will issue a warning, and in such case, users may wish compute the chemical potentials from a finite-temperature free energy of the compound of interest instead of using its 0 K DFT energy.

4. Use cases

4.1. Simple defects

To validate this implementation of the correction workchain, we have computed the defect formation energies of 4 different defects including v_{Li}^{-1}, v_{Cl}^{+1} in the antiperovskite Li_3ClO, v_{Mg}^{-2} in MgO and v_{Al}^{-3} in AlP as a function of the supercell size as shown in figure 5. As expected, the uncorrected formation energies show strong variations with the supercell size. However, upon correction, the formation energies become essentially independent of the supercell size and are all within 5% of the their values in the dilute limit.

4.2. Applications

To demonstrate the capabilities of AiiDA-Defect in a (semi) high-throughput environment, we choose to study the defect chemistry the Li-ion conductors $LiZnPS_4$, which has been computationally predicted to exhibit extremely high Li-ion conductivity at room temperature, if the material can be synthesized with excess Li [30, 31]. From the defect point of view, one way to obtain a Li-rich composition is to have partial substitution on the Zn sublattice by lithium (Li_{Zn}) and the creation of lithium interstitial Li_i to maintain

Table 1. All native and extrinsic defects in their various charge states which were included in the calculation of defect chemistry of $LiZnPS_4$.

Native defects		Extrinsic defects	
Defect types	Charge states	Defect types	Charge states
v_{Li}	$0, -1$	Mg_{Li}	$0, +1$
v_{Zn}	$0, -2$	Mg_{Zn}	0
v_P	$0, -5$	Ca_{Li}	$0, +1$
v_S	$0, +2$	Ca_{Zn}	0
Li_i	$0, +1$	B_{Li}	$0, +2$
Zn_i	$0, +2$	B_{Zn}	$0, +1$
Li_{Zn}	$0, -1$	B_P	$0, -2$
Zn_{Li}	$0, +1$	Al_{Li}	$0, +2$
		Al_{Zn}	$0, +1$
		Al_P	$0, -2$
		Si_P	$0, -1$
		Sn_P	$0, -1$

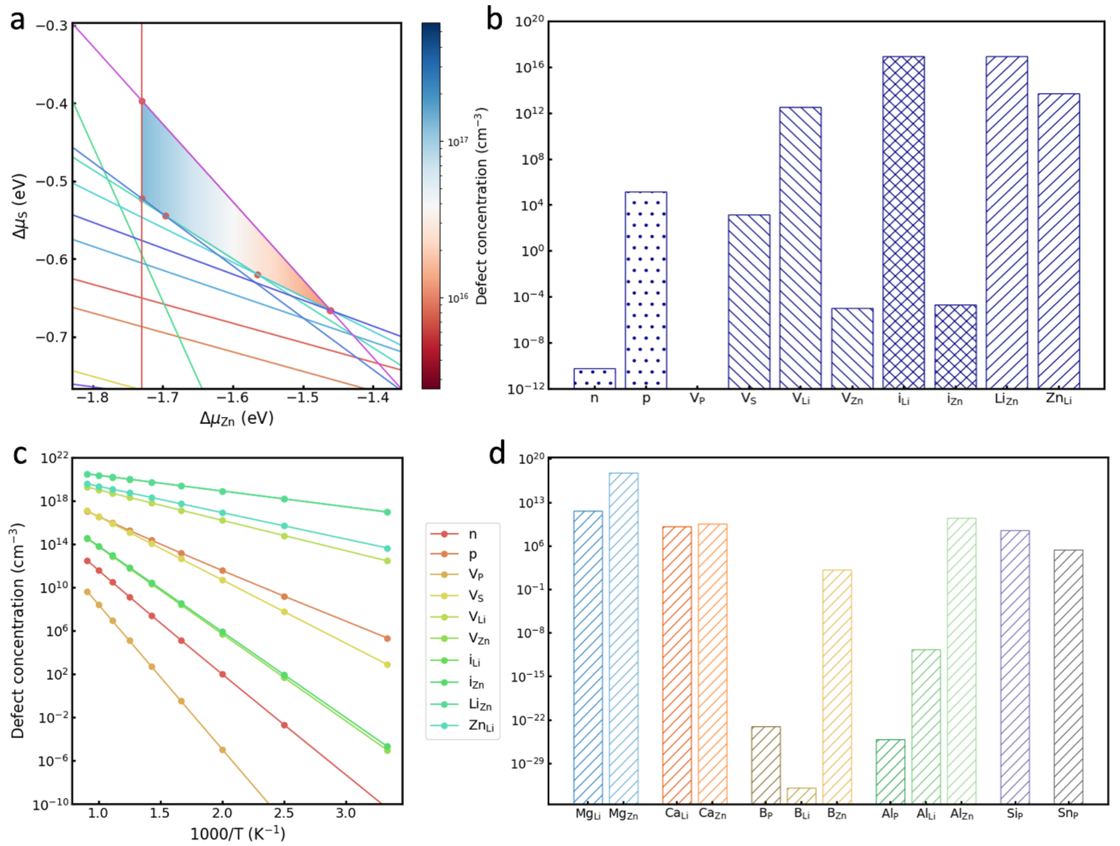

Figure 6. (a) A section of stability region of $LiZnPS_4$ with the concentration of Li_i directly plotted on top. (b) Concentration of various native defects at 300 K. (c) Arrhenius plots of concentrations of native defects. (d) Concentration of various extrinsic defects at 300 K.

charge neutrality. At thermodynamics equilibrium, whether these defects are operative or not will depend on their formation energies relative to that of other possible defects. AiiDA-defects greatly simplifies the undertaking of this kind of study. These calculations can be conveniently carried out and the results can be easily retrieved and analyzed using the functionalities implemented in the package.

As mentioned, in order to obtain a complete picture of defect chemistry of a material at thermodynamic equilibrium, several defects in various charge states have to be considered. For this materials, all the intrinsic (native) and extrinsic defects which are included in this study are listed in table 1 along with their charge states. First, we need to determine the stability region of $LiZnPS_4$ from which the chemical potential of each constitutive element can be chosen for the calculation of defect formation energy. Because $LiZnPS_4$ is a quaternary compound, its stability region is a polyhedron and a section of the polyhedron at a fixed $\Delta\mu_{Li}$ is shown in figure 6(a). Unlike in figure 2, we also directly show the concentration of the defect that we want to optimize, namely Li_i in the stability region, therefore allowing a quick assessment of chemical potentials and thus, the synthesis conditions under which the concentration of the desired defects can be maximized. The concentration of various defects at room-temperature in the intrinsic regime (at chemical potentials corresponding to the centroid of the stability region, see section 2.3) is shown in figure 6(b). One can immediately see that the dominant native defect in $LiZnPS_4$ are lithium-zinc antisite defects which are charge-compensated by the creation of interstitials lithium, leading to a lithium-excess composition as one would desire. This excess is very small under this condition and may not lead to any observable effects in ionic conductivity but can be further increased by varying the chemical potentials as shown in figure 6(a). This is in very good agreement with the results of Suzuki *et al* who, by controlling the synthesis conditions, was able to synthesize the Li-excess and Zn-deficient composition $Li_{1+2x}Zn_{1-x}PS_4$ with x up to 0.625 [31]. The variation of the defect concentrations as a function of temperature can be also be obtained as shown in figure 6(c). Finally, we can also assess the dopability of this materials by examining the concentration of various aliovalent dopants as shown in figure 6(d). The most favorable extrinsic defect is Mg_{Zn} which is an isovalent defect and will not lead to appreciable change in the concentration of lithium. Mg_{Li} and Ca_{Li} are thermodynamically favorable but will lead to Li antisite defects and reduce Li concentration. The most promising dopants are predicted to be Si_P and Sn_P which can lead to an increase of Li_i via charge compensation, but their concentration at equilibrium at room temperature is predicted to be low which implies that one might need to 'quench' the defects from high temperature so that they can exist at appreciable concentrations at room temperature to produce any observable effects on the Li-ion conductivity. This shows the robustness and user-friendliness of AiiDA-defects to perform defect calculations in a high-throughput environment, and its utility to both academic and industrial users alike.

5. Conclusions

The accurate calculation of defect formation energies for charged systems can be challenging. We summarized the theory behind these calculations and highlighted the spurious interactions and sources of errors for these calculations when conducted within the limitations of density-functional theory calculations under periodic boundary conditions. A number of corrections and techniques have been proposed to address these technicalities, especially in the form of *a posteriori* processing steps. There is an even larger menagerie of different codes and packages applying these techniques to different parts of the problem. These packages are very valuable, but tend to be applicable to a specific subset of use-cases or quantum codes, and may require some or even substantial manual computation and intervention. We briefly introduced AiiDA—a materials informatics platform providing a rich set of tools for automating computational workflows, distributing jobs to high performance computing resources, and for parsing and storing the results, and the associated provenance graph, in a high-performance queryable database. Then, we presented and described AiiDA-defects, a plugin to this framework which seeks to fully automate defect chemistry characterization and to enable this research in reliable, robust and even high-throughput modalities. In particular, our efforts have focused on identifying and solving the challenges to realizing these calculations as an automated workflow. These include the choice of charge model, the calculation of relative permittivity, the alignment of potentials, the calculation of chemical potential, the calculation and selection of an appropriate self-consistent Fermi level across a set of defects. The interaction of all of these components is taken care of automatically within a set of nested workchains. This modularity also allows for easy extension to new approximations and corrections as they are developed. The required DFT calculations are prepared and run automatically as needed, and like the rest of the package, the interfaces to the DFT codes are also modular, allowing for codes beyond Quantum ESPRESSO [56, 57] to be conveniently employed via existing AiiDA plugins. We showed that the correction procedures can be applied automatically to correct the formation energy of charged defects for the cases of the typical test systems. We also showed the potential of this package for a more elaborate application to the Li-ion conductors $LiZnPS_4$, highlighting how a complete characterization of the defect chemistry can be obtained for a range of possible defects. AiiDA-defects is powerful tool for the computational community that enables defect calculations to be performed in a robust, automated way, that is convenient and suitable for high-throughput computational studies. We hope it will be an impactful tool for new materials discovery, and a most welcome edition to the suite of community-produced packages in the AiiDA ecosystem.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://archive.materialscloud.org/.

Acknowledgments

This work is supported by the MARVEL National Centre of Competence in Research (NCCR) funded by the Swiss National Science Foundation (Grant Agreement ID 51NF40-182892) and by the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 824143 (European MaX Centre of Excellence 'Materials design at the Exascale') and Grant Agreement No. 814487 (INTERSECT project). We thank Chiara Ricca and Ulrich Aschauer for discussions and prototype implementation ideas. The authors also would like to thank the Swiss National Supercomputing Centre CSCS (Project s1073) for providing the computational ressources and Solvay for funding this project. We thank Arsalan Akhtar Lorenzo Bastonero, Luca Bursi, Francesco Libbi, Riccardo De Gennaro and Daniele Tomerini for useful discussions and feedback.

Code availability

An open-source software implementation of AiiDA-defects is available at https://github.com/epfl-theos/aiida-defects.

ORCID iDs

Sokseiha Muy https://orcid.org/0000-0001-6457-9187
Nicola Marzari https://orcid.org/0000-0002-9764-0199

References

[1] Seebauer E G and Kratzer M C 2006 *Mater. Sci. Eng.* R **55** 57
[2] Lyons J L, Wickramaratne D and Van de Walle C G 2021 *J. Appl. Phys.* **129** 111101
[3] Kim J Y, Lee J-W, Jung H S, Shin H and Park N-G 2020 *Chem. Rev.* **120** 7867
[4] Nayak P K, Mahesh S, Snaith H J and Cahen D 2019 *Nat. Rev. Mater.* **4** 269
[5] Shi X-L, Zou J and Chen Z-G 2020 *Chem. Rev.* **120** 7399
[6] Bachman J C *et al* 2016 *Chem. Rev.* **116** 140
[7] Ohno S, Banik A, Dewald G F, Kraft M A, Krauskopf T, Minafra N, Till P, Weiss M and Zeier W G 2020 *Prog. Energy* **2** 022001
[8] Doherty M W, Manson N B, Delaney P, Jelezko F, Wrachtrup J and Hollenberg L C L 2013 *Phys. Rep.* **528** 1
[9] Degen C L, Reinhard F and Cappellaro P 2017 *Rev. Mod. Phys.* **89** 035002
[10] Spaeth J-M and Overhof H 2003 *Point Defects in Semiconductors and Insulators* (*Springer Series in Materials Science*) (Berlin: Springer)
[11] Sanguinetti S, Guzzi M and Gurioli M 2008 *Characterization of Semiconductor Heterostructures and Nanostructures* ed C Lamberti (Amsterdam: Elsevier) pp 175–208
[12] Krause-Rehberg R and Leipner H S 2003 *Positron Annihilation in Semiconductors: Defect Studies*(Berlin: Springer)
[13] Li W, Field K G and Morgan D 2018 *npj Comput. Mater.* **4** 1
[14] Bourgoin J and Lannoo M 1983 *Point Defects in Semiconductors II: Experimental Aspects* (*Springer Series in Solid-State Sciences*) (Berlin: Springer)
[15] Hjorth Larsen A *et al* 2017 *J. Phys.: Condens. Matter* **29** 273002
[16] Curtarolo S *et al* 2012 *Comput. Mater. Sci.* **58** 218
[17] Jain A *et al* 2015 *Concurr. Comput.: Pract. Exp.* **27** 5037
[18] Zapata F, Ridder L, Hidding J, Jacob C R, Infante I and Visscher L 2019 *J. Chem. Inf. Model.* **59** 3191
[19] Adorf C S, Dodd P M, Ramasubramani V and Glotzer S C 2018 *Comput. Mater. Sci.* **146** 220
[20] Maffioletti S and Murri R 2012 *Proc. EGI Community Forum 2012/EMI Second Technical Conf.—PoS(EGICF12-EMITC2)* (Trieste:Sissa Medialab) p 143
[21] Babuji Y *et al* 2019 *Proc. 28th Int. Symp. on High-Performance Parallel and Distributed Computing* (*HPDC '19*) (New York: Association for Computing Machinery) pp 25–36
[22] Broberg D, Medasani B, Zimmermann N E R, Yu G, Canning A, Haranczyk M, Asta M and Hautier G 2018 *Comput. Phys. Commun.* **226** 165
[23] Goyal A, Gorai P, Peng H, Lany S and Stevanović V 2017 *Comput. Mater. Sci.* **130** 1
[24] Naik M H and Jain M 2018 *Comput. Phys. Commun.* **226** 114
[25] Péan E, Vidal J, Jobic S and Latouche C 2017 *Chem. Phys. Lett.* **671** 124
[26] Huang M, Zheng Z, Dai Z, Guo X, Wang S, Jiang L, Wei J and Chen S 2022 *J. Semicond.* **43** 042101
[27] Huber S P *et al* 2020 *Sci. Data* **7** 300
[28] Uhrin M, Huber S P, Yu J, Marzari N and Pizzi G 2021 *Comput. Mater. Sci.* **187** 110086
[29] 2012 AiiDA plugin registry (available at: https://aiidateam.github.io/aiida-registry/) (Accessed 23 August 2021)
[30] Richards W D, Wang Y, Miara L J, Kim J C and Ceder G 2016 *Energy Environ. Sci.* **9** 3272
[31] Suzuki N, Richards W D, Wang Y, Miara L J, Kim J C, Jung I-S, Tsujimura T and Ceder G 2018 *Chem. Mater.* **30** 2236
[32] Squires A G, Scanlon D O and Morgan B J 2020 *Chem. Mater.* **32** 1876
[33] Zhang S B and Northrup J E 1991 *Phys. Rev. Lett.* **67** 2339
[34] Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A and Van de Walle C G 2014 *Rev. Mod. Phys.* **86** 253
[35] Freysoldt C, Neugebauer J and Van de Walle C G 2009 *Phys. Rev. Lett.* **102** 016402
[36] Freysoldt C, Neugebauer J and Van de Walle C G 2011 *Phys. Status Solidi* b **248** 1067
[37] Dabo I, Kozinsky B, Singh-Miller N E and Marzari N 2008 *Phys. Rev.* B **77** 115139
[38] Dabo I, Kozinsky B, Singh-Miller N E and Marzari N 2011 *Phys. Rev.* B **84** 159910

[39] Walsh A 2021 *npj Comput. Mater.* **7** 72
[40] Kim S, Hood S N, Park J-S, Whalley L D and Walsh A 2020 *J. Phys. Energy* **2** 036001
[41] Makov G and Payne M C 1995 *Phys. Rev.* B **51** 4014
[42] Lany S and Zunger A 2008 *Phys. Rev.* B **78** 235104
[43] Suo Z-J, Luo J-W, Li S-S and Wang L-W 2020 *Phys. Rev.* B **102** 174110
[44] Chagas da Silva M, Lorke M, Aradi B, Farzalipour Tabriz M, Frauenheim T, Rubio A, Rocca D and Deák P 2021 *Phys. Rev. Lett.* **126** 076401
[45] Hubere S P *et al* 2021 *npj Comput. Mater.* **7** 136
[46] Baroni S, de Gironcoli S, Dal Corso A and Giannozzi P 2001 *Rev. Mod. Phys.* **73** 515
[47] Komsa H-P, Rantala T T and Pasquarello A 2012 *Phys. Rev.* B **86** 045112
[48] Andreussi O, Dabo I and Marzari N 2012 *J. Chem. Phys.* **136** 064102
[49] Buckeridge J, Scanlon D, Walsh A and Catlow C 2014 *Comput. Phys. Commun.* **185** 330
[50] Buckeridge J 2019 *Comput. Phys. Commun.* **244** 329
[51] Alexander D O S, Squires G and Morgan B J 2023 *J. Open Source Softw.* **8** 4962
[52] Gorai P, Famprikis T, Singh B, Stevanović V and Canepa P 2021 *Chem. Mater.* **33** 7484
[53] Canepa P, Sai Gautam G, Broberg D, Bo S-H and Ceder G 2017 *Chem. Mater.* **29** 9657
[54] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 *Comput. Mater. Sci.* **68** 314
[55] Mosquera-Lois I, Kavanagh S R, Walsh A and Scanlon D O 2023 *npj Comput. Mater.* **9** 1
[56] Giannozzi P *et al* 2009 *J. Phys.: Condens. Matter* **21** 395502
[57] Giannozzi P *et al* 2017 *J. Phys.: Condens. Matter* **29** 465901