Quick-start guide for first-principles modelling of point defects in crystalline materials

Sunghyun Kim,[1, 2] Samantha N. Hood,[1] Ji-Sang Park,[3] Lucy D. Whalley,[1] and Aron Walsh[1, 2, a]

[1)] *Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, UK*

[2)] *Department of Materials Science and Engineering, Yonsei University, Seoul 03722, Korea*

[3)] *Department of Physics, Kyungpook National University, Daegu 41566, Korea*

(Dated: 8 May 2020)

Point defects in crystalline materials participate in electronic (e.g. carrier generation and recombination) and optical (e.g. absorption and emission) processes. First-principles calculations of defects based on density functional theory have been widely used to complement, and even validate, experimental observations. In this 'quick-start guide', we discuss the best practice in how to calculate the formation energy of point defects in crystalline materials and analysis techniques appropriate to probe changes in structure and properties.

'The perfect crystal is one of the idealizations commonly found in theoretical physics and science fiction'.[1] Defects are present in all crystals – they can be detrimental to device performance or can be beneficial for use in a wide variety of processes, including:

- *optical*, e.g. the colour centre V_{Br} in CsBr;

- *chemical*, e.g. the catalytic centre Li_{Mg} in MgO;

- *mechanical*, e.g. hardening of Fe using C;

- *electrical*, e.g doping by V_O in ZnO.

Computational techniques allow us to investigate properties of point defects at a level of detail that is often difficult to access via experiments. It is possible to isolate the behaviour of particular defects and predict their spectral signatures and physical effects. One challenge is that much of the infrastructure for materials modelling is built on translational symmetry (e.g. Bloch wave functions). Defects break the periodicity of a crystal, and their accurate description is a continuing endeavour for materials modelling. Following our quick-start guide on interfaces,[2] this is a primer for researchers starting to work on first-principles defect simulations.

1. DEFECT NOTATION

A defect is often referred to according to its spectroscopic signature. An anion vacancy with a trapped electron in an ionic crystal may absorb light in the visible range, making the transparent host material colourful; an *F* center (*Farbe* means colour in German). In electrical measurements, such as deep level transient spectroscopy (DLTS), defects are labeled in order of their energy levels. For example, $D1$ and $D2$ for electron donors and $A1$ and $A2$ for electron acceptors, as shown for the case of the
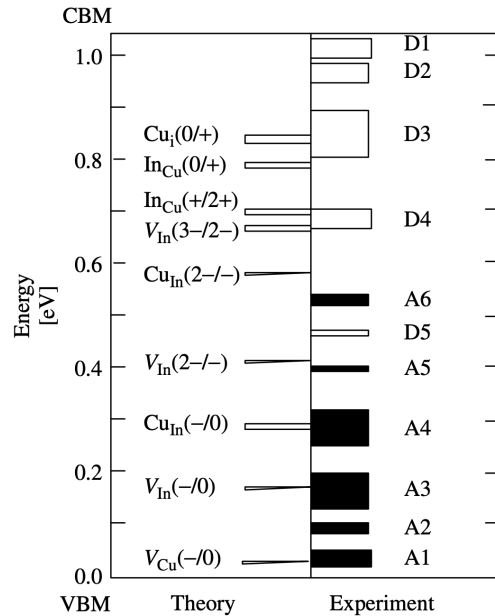

FIG. 1. Intrinsic defect levels in $CuInSe_2$ with respect to the valence (VBM) and conduction (CBM) bands that have been measured and calculated from first-principles. The height of the histogram columns on the right side represents the spread in experimental data. Reproduced with permission from Ref. 3.

chalcopyrite semiconductor $CuInSe_2$ in Fig. 1. Matching atomic models with spectroscopic signals is of considerable importance in physics and chemistry of materials.

A widely used notation for the atomic defect models is Kröger-Vink.[5] A defect is represented by X_Y, where X is the species occupying the atomic site Y. A vacancy and an interstitial site are denoted by V and i, respectively. The relative charge of a defect can be represented by a superscript of $^\times$, $^\bullet$ and $'$ for a neutral, positive and negative charges, respectively. A numerical notation[6] of the charge state (q) is popular in the recent literature. For example, the negatively-charged B-on-Si is represented

$$\Delta H_{\mathrm{D,q}}(E_F, \mu) \;=\; [E_{\mathrm{D,q}} - E_{\mathrm{H}}] \;+\; \sum_i n_i \mu_i \;+\; q E_F \;+\; E_{\mathrm{corr}}$$

Defect Formation Energy · Defect Supercell · Host Supercell · Chemical potentials from phase stability · Electron chemical potential · Finite size corrections
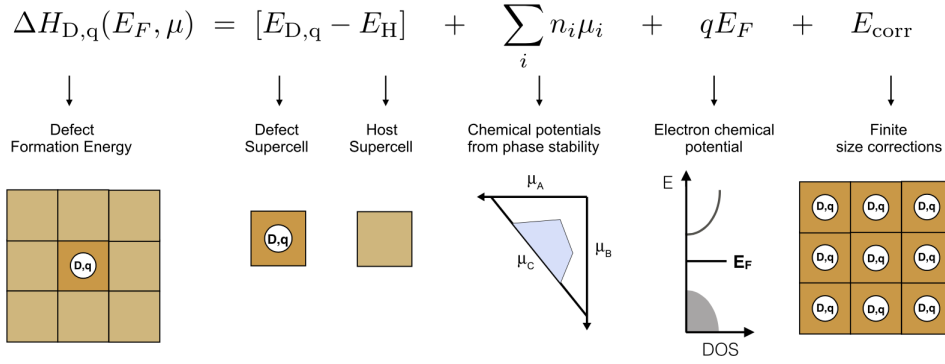
FIG. 2. Illustration of the terms required to compute the charged defect formation energy as function of the atomic and electronic chemical potentials. Reproduced with permission from Ref. 4.

by B_{Si}^{-1}, and the neutral Si self-interstitial is represented by Si_i^0. When more than one symmetrically inequivalent site exists, an additional subscript of the Wyckoff position or the point group symmetry of the site can be added to distinguish distinct species.

2. EQUILIBRIUM CONCENTRATIONS

The equilibrium concentration of defects (n_d) at a fixed temperature and pressure is given by the density that minimises the free energy:

$$n_d = N_{\mathrm{site}} g \exp\left(-\frac{\Delta G_f}{k_{\mathrm{B}} T}\right), \tag{1}$$

where N_{site} and g denote the number of available sites of the defect in the unit volume and the degeneracy of the defect, respectively. ΔG_f is the Gibbs free energy of formation of the defect, k_{B} is the Boltzmann constant, and T is temperature. This can be further decomposed into contributions from enthalpy (H) and vibrational entropy (S) as follows:

$$n_d = N_{\mathrm{site}} g \exp\left(-\frac{\Delta H_f}{k_{\mathrm{B}} T}\right) \exp\left(\frac{\Delta S}{k_{\mathrm{B}}}\right), \tag{2}$$

The enthalpy change dominates under standard conditions and is easier to compute, therefore vibrational terms are often neglected.

The formation energy of a defect is given as:

$$\Delta H_f = \Delta E + \sum n_i \mu_i + q E_F + E_{\mathrm{corr}}, \tag{3}$$

where ΔE is a change in the total energy due to the formation of the defect. The second term takes into account the energy cost to exchange n_i atoms of kind i with their reservoir chemical potential μ_i (e.g. this term for X_Y^q is $\mu_X - \mu_Y$). q and E_F denote the charge state of the defect and the Fermi level, respectively. The correction term (E_{corr}) will be discussed in the next section. These terms are illustrated in Fig. 2. The typical magnitudes

of first, second, and third terms in Equation 3 are a few eV, while the value of the forth term is usually < 1 eV.

Since a macroscopic crystal should be charge neutral overall, the concentrations of electrons n_0, holes p_0, positively-charged donors $n_{D_i}^{q_i}$, and negatively-charged acceptors $n_{A_j}^{q_j}$ must satisfy electroneutrality:[7]

$$p_0 + \sum_i q_i n_{D_i}^{q_i} = n_0 + \sum_j q_j n_{A_j}^{q_j}. \tag{4}$$

The equilibrium population of charge carriers are given by:

$$n_0 = N_C e^{-\frac{E_C - E_F}{k_{\mathrm{B}} T}}, \tag{5}$$

$$p_0 = N_V e^{-\frac{E_F - E_V}{k_{\mathrm{B}} T}}, \tag{6}$$

where N_C and N_V are the effective density of states of the conduction (E_C) and valence (E_V) bands, respectively. Since the formation energy of charged defects depends on the Fermi level (Equation 3), which in turn depends on the population of charged defects, the concentrations must be solved self-consistently as illustrated in Fig. 3.

3. PRACTICAL CALCULATIONS

Several approaches exist for calculating defect formation energies. Those based on embedding potentials (a dilute defect in a host matrix) offer several advantages but remain technically challenging to setup and analyse.[8,9] For first-principles approaches, including density functional theory (DFT), the supercell method is by far the most widely employed.[10]

With the supercell method, periodic boundary conditions remain in place, but an expanded repeat unit is employed. This repeat unit should be large enough so that the host material is well described and the periodic interactions between repeating defects can be corrected. For example, if a cubic unit cell with lattice spacing (a)
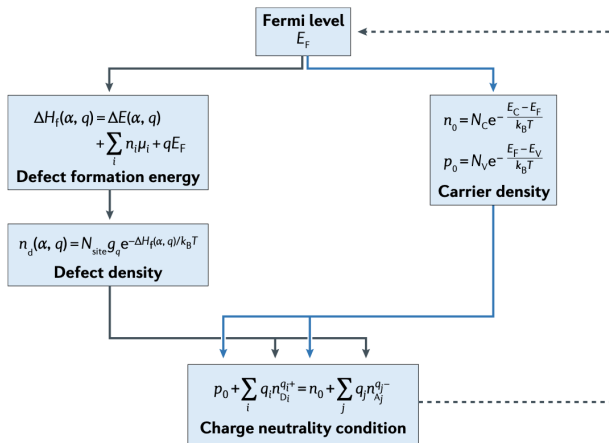
FIG. 3. Illustration of a self-consistent treatment of charged defect populations in a crystalline host material. Reproduced with permission from Ref. 7.

is expanded using a matrix of:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad (7)$$

there will be a spacing of $2a$ between repeating defects along each axis.

Vacancies and substitutions can be introduced by simply removing atoms and replacing atoms, respectively. Interstitial sites require further care because of the larger configurational space. Candidate sites for interstitials can be assigned in a way that resembles well-known structure motifs such as tetrahedra or octahedra.[11] `PyLada` adapts a scheme based on Voronoi tessellation.[4] The range of possible charge states is inferred from the oxidation states of atoms. For example, the charge state of Sn_{Zn} can be 0 or +2 which correspond to Sn(II) and Sn(IV), respectively.

We will now breakdown Equation 3 into its components and discuss how to compute the four terms in turn.

A. Term 1: Raw defect formation energy

To obtain ΔE, the standard operating procedure is as follows:

1. Fully optimise the crystal structure of the host material in its *primitive unit cell*.

2. Create a *supercell* expansion that is as close to cubic as possible, which minimises anisotropy in defect-defect interactions.

3. The total energy of the pristine *supercell* becomes the E_H reference for defect formation energies.

4. Introduce a defect of your choosing and optimise the structure under *constant volume* conditions keeping the lattice vectors fixed.

5. The total energy of the optimised defective *supercell* is $E_{D,q}$.

6. The raw defect formation energy in Equation 3 is the difference of two total energy calculations, $\Delta E = E_{D,q} - E_H$.

One common mistake is to assume an incorrect spin state. Even if a host material is non-magnetic, the ground state of a defect may require spin-polarisation, e.g. in an open-shell singlet or triplet configuration.[12]

B. Term 2: Atomic chemical potentials

The raw defect formation energies are not meaningful as they do not represent balanced reactions (atoms may have been created or annihilated). The chemical potentials μ_i account for the exchange of species with their environment. In practice, a range of environments are possible. In some cases we may try to mirror an experiment, e.g. an environment of oxygen gas at a particular temperature or pressure. More generally, we treat μ_i as a parameter than can vary over the accessible phase space of the material. The standard operating procedure is as follows:

1. Calculate the total energy of the standard states of each element found in your host material (e.g. Zn metal and oxygen gas for ZnO).

2. Calculate the total energy of all possible secondary phases that could form (e.g. ZnO and Al_2O_3 for a $ZnAl_2O_4$ host).

3. Solve the accessible chemical potential region considering these boundaries. Any point in this stability field can be used safely in Equation 3.

For multi-component materials, the associated simultaneous equations become cumbersome to solve. One freely available package developed for this purpose is `CPLAP`.[13]

C. Term 3: Fermi level

In semiconductors and dielectrics, the electronic chemical potential (Fermi level) can vary between the valence and conduction bands, depending on the doping regime and history of a sample. Here, qE_F represents the cost of exchanging electrons or holes with the host and is therefore proportional to the defect charge. It has become standard to present defect formation energies as a function of a parametric E_F in the range $[0, E_g]$, where E_g is the bandgap of the host compound. In reality, the full range will not always be accessible. The equilibrium Fermi level can be solved conveniently using a package such as `SC-FERMI`.[14]

Code	Purpose	Correction Scheme
PyLada[4]	Automate point defect calculations	Lany-Zunger
CoFFEE[15]	Electrostatic corrections for charged defect calculations	FNV
PyDEF[16]	Defect formation energies using VASP	Lany-Zunger
PyCDT[11]	Facilitate high-throughput DFT defect calculations	FNV and KO
sxdefectalign[17]	Point defects in bulk within SPHinX[18]	FNV
sxdefectalign2d[19]	Point defects at surfaces and interfaces within SPHinX	FNV
CPLAP[13]	Calculation of stable chemical potential ranges	..
SC-Fermi[14]	Determine the Fermi level based on defect formation energies	..
CarrierCapture.jl[20]	Calculate non-radiative carrier capture by anharmonic defects	..

TABLE I. A selection of packages available to assist in processing first-principles defect calculations.

D. Term 4: Charged defect corrections

The treatment of charged defects is a longstanding issue for periodic boundary conditions due to the long-range nature of the Coulomb interaction. There are two issues to resolve. Firstly, charged defects are able to interact with their periodic images. Secondly, a homogeneous "jellium" background charge is introduced to enforce charge neutrality and a convergent Coulomb energy. These issue results in a shift to the average electrostatic potential the supercell and in $E_{D,q}$. Large supercells are optimal to minimise these errors; however, we are often limited by computational cost and available resources.

A number of correction schemes have been developed that result in the term E_{corr}. Some are discussed below and for a more complete description of these issues, we refer the reader elsewhere.[21,22]

1. The *Leslie-Gillan* correction[23] models a point charge q interacting with its periodic images through an isotropic dielectric medium. This correction takes a simple analytic form that depends on the charge state q, static dielectric constant ε, separation between images L and the Madelung constant α_m characteristic of the lattice.

2. The *Makov-Payne* correction[24] includes an additional term to account for higher-order multipoles:

$$E^{\mathrm{MP}} = \frac{q^2 \alpha_m}{2\varepsilon L} + qQL^{-3}. \tag{8}$$

An issue associated with this approach is in determining the quadrupole moment Q.

3. The *Lany-Zunger* correction[25] combines the Makov-Payne correction, including a procedure for calculating Q, with potential alignment to correct for the shift in electrostatic potential.

4. The *Freysoldt, Neugebauer and van de Walle* (FNV) method[26] models the defect charge as a Gaussian distribution. The difference between the electrostatic potential of the charged defect and perfect bulk supercells, calculated far from the defect, is aligned with the defect model potential.

5. *Kumagai and Oba* (KO) extended the FNV method using atomic site potentials combined with the Gaussian charge model for an anisotropic dielectric medium.[27]

Such schemes were initially developed for use with three-dimensional crystal with homogeneous dielectric screening. Recent work has extended these methods to two-dimensional[28,29] and one-dimensional[30] materials.

There is no standardised approach to defect charge corrections, which can lead to a spread in calculated defect formation energies in the literature, and predicted defect densities that differ by orders of magnitude. FNV and KO have become the most widely used methods. These are implemented in several computational tools such as `PyLada` and `sxdefectalign`, see Table 1. Developing an efficient scheme to account for microscopic effects and anisotropy remains an active area of research.[21,22]

As a side note, for shallow defects where the valence or conduction bands become occupied, an additional band filling correction can be required to obtain results in the dilute limit.[31]

FRONTIERS OF DEFECT MODELLING

Carrier capture and recombination

Our previous discussion was limited to an equilibrium description of defects and charge states (e.g. for a crystal in the dark at a certain temperature). Either following a pump pulse or under steady-state illumination, the kinetics of carrier (electron and hole) capture by defects becomes important.

A defect may capture a delocalised free carrier with the aid of electron-phonon coupling. As the charge is transferred from the delocalised state to the localised state around the defect, the local atomic configuration is rearranged. Such processes can be described in the framework of a configurational coordinate (CC) diagram. The CC is usually in the form of a one-dimensional pathway between two local minimum structures as shown in Fig. 4.

To describe such non-radiative defect processes, both the nuclear and electron wave functions must be ade-
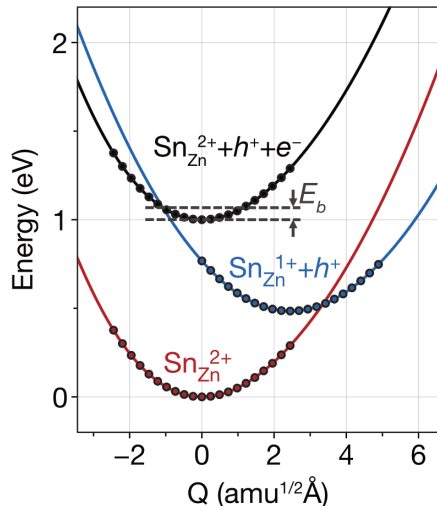
FIG. 4. Configuration coordinate diagram for Sn_{Zn} in $Cu_2ZnSnSe_4$. It describes the non-radiative electron-hole recombination process: $Sn_{Zn}^{2+} \xrightleftharpoons{h\nu} Sn_{Zn}^{2+} + e^- + h^+ \xrightleftharpoons{-\hbar\omega} Sn_{Zn}^+ + h^+ \xrightleftharpoons{-\hbar\omega} Sn_{Zn}^{2+}$. Reproduced with permission from Ref. 35.

quately described. Several schemes have been proposed to calculate the cross-sections and rates of carrier capture based on first-principles simulations.[32–34] The current schemes require a significant amount of researcher expertise and computational resource. Further developments are required to develop reliable and robust procedures suitable for general applications.

Dynamics and transport

Materials engineering often requires a certain spatial distribution of defects. To achieve a desirable profile of defect species in a sample, one needs to understand atomic diffusion in crystals. However, diffusion is a rare event compared to typical vibrational frequencies (1–10 THz). Standard *ab initio* molecular dynamics simulations can not properly capture diffusion kinetics due to the limited time and length scales that are accessible. To overcome such difficulties, statistical methods such as transition state theory (TST)[36] can be employed. The nudged elastic band (NEB)[37] method has been developed to find a minimum energy path and the energy barrier at the saddle point, which can be used to parameterise kinetic models.

Real diffusion processes may consist of multiple barriers, even in relatively simple structure types such as zincblende.[38] The role of excited states in crystals with defect-mediated mass transport, for example as found in halide perovskite solar cells[39,40], is one area that requires further development.

Automation and databases

Defect studies are demanding on human time due to the large number of individual calculations involved. There are a growing number of publicly available packages to assist with pre- or post-processing (see Table I). Powered by the rapidly expanding computational capacity and the automation frameworks such as `AiiDA`,[41] databases for first-principles calculations including modelling of defects are being developed, which will offer the opportunity for to gain insights when combined with statistical analysis and machine learning models.

CONCLUSION

We have discussed common considerations and procedures for simulating point defects in bulk materials. First-principles modelling can provide both qualitative and quantitative descriptions of properties of crystalline solids including carrier generation (dopants) and recombination (traps). Emerging renewable energy technologies – including batteries, solar cells, and thermoelectrics – can benefit from first-principles modelling of point defects, complementing experimental characterisation. The prospect of developing large defect databases with the aid of automation offers an opportunity to extract valuable insight toward *defect-engineered* materials.

ACKNOWLEDGMENTS

We thank Seán Kavanagh for a careful reading of the manuscript. The research was funded by the Royal Society and the EU Horizon2020 Framework (STARCELL, grant no. 720907). Additional support was received from the Faraday Institution (grant no. FIRG003) and the EPSRC (EP/L01551X/1).

[1] A. M. Stoneham, Appl. Surf. Sci. **14**, 249 (1983).
[2] J.-S. Park, Y.-K. Jung, K. T. Butler, and A. Walsh, J. Phys. Energy **1**, 016001 (2018).
[3] S. Zhang, S.-H. Wei, A. Zunger, and H. Katayama-Yoshida, Phys. Rev. B **57**, 9642 (1998).
[4] A. Goyal, P. Gorai, H. Peng, S. Lany, and V. Stevanovic, Comput. Mater. Sci. **130**, 1 (2017).
[5] F. Kröger, The Chemistry of Imperfect Crystals (North-Holland, Amsterdam, 1974).
[6] N. G. Connelly, T. Damhus, R. M. Hartshorn, and A. T. Hutton, eds., Nomenclature of Inorg. Chem. (The Royal Society of Chemistry, 2005).
[7] J. S. Park, S. Kim, Z. Xie, and A. Walsh, Nat. Rev. Mater. , 1 (2018).
[8] Z. Xie, Y. Sui, J. Buckeridge, C. R. A. Catlow, T. W. Keal, P. Sherwood, A. Walsh, D. O. Scanlon, S. M. Woodley, and A. A. Sokol, Phys. Stat. Sol. (a) **214**, 1600445 (2017).
[9] J. Buckeridge, C. R. A. Catlow, D. Scanlon, T. Keal, P. Sherwood, M. Miskufova, A. Walsh, S. Woodley, and A. Sokol, Phys. Rev. Lett. **114**, 016405 (2015).
[10] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, Rev. Mod. Phys. **86**, 253 (2014).

[11] D. Broberg, B. Medasani, N. E. Zimmermann, G. Yu, A. Canning, M. Haranczyk, M. Asta, and G. Hautier, Comput. Phys. Commun. **226**, 165 (2018).

[12] A. A. Sokol, A. Walsh, and C. R. A. Catlow, Chem. Phys. Lett. **492**, 44 (2010).

[13] J. Buckeridge, D. O. Scanlon, A. Walsh, and C. R. A. Catlow, Comp. Phys. Commun. **185**, 330 (2014).

[14] J. Buckeridge, Comp. Phys. Commun. **244**, 329 (2019).

[15] M. H. Naik and M. Jain, Comput. Phys. Commun. **226**, 114 (2018).

[16] E. Pan, J. Vidal, S. Jobic, and C. Latouche, Chem. Phys. Lett. **671**, 124 (2017).

[17] C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).

[18] O. Marquardt, S. Schulz, C. Freysoldt, S. Boeck, T. Hickel, E. P. O'Reilly, and J. Neugebauer, Opt. and Quant. Electron. **44**, 183 (2012).

[19] C. Freysoldt and J. Neugebauer, Phys. Rev. B **97**, 205425 (2018).

[20] S. Kim, S. Hood, P. van Gerwen, L. Whalley, and A. Walsh, J. Open Source Softw. **5**, 2102 (2020).

[21] T. R. Durrant, S. T. Murphy, M. B. Watkins, and A. L. Shluger, J. Chem. Phys. **149**, 024103 (2018).

[22] D. Vinichenko, M. G. Sensoy, C. M. Friend, and E. Kaxiras, Phys. Rev. B **95**, 235310 (2017).

[23] M. Leslie and M. J. Gillan, J. Phys. C: Solid State Phys. **18**, 973 (1985).

[24] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[25] S. Lany and A. Zunger, Modell. Simul. Mater. Sci. Eng. **17**, 084002 (2009).

[26] C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).

[27] Y. Kumagai and F. Oba, Phys. Rev. B **89**, 195205 (2014).

[28] C. Freysoldt and J. Neugebauer, Phys. Rev. B **97**, 205425 (2018).

[29] H.-P. Komsa and A. Pasquarello, Phys. Rev. Lett. **110**, 095505 (2013).

[30] S. Kim, K. J. Chang, and J.-S. Park, Phys. Rev. B **90**, 085435 (2014).

[31] S. Lany and A. Zunger, Phys. Rev. B **78**, 2637 (2008).

[32] L. Shi and L.-W. Wang, Phys. Rev. Lett. **109**, 245501 (2012).

[33] A. Alkauskas, Q. Yan, and C. G. Van de Walle, Phys. Rev. B **90**, 075202 (2014).

[34] S. Kim, S. N. Hood, and A. Walsh, Phys. Rev. B **100**, 041202 (2019).

[35] S. Kim, J. A. Marquez, T. Unold, and A. Walsh, Energ. Environ. Sci. , Advance article (2020).

[36] G. H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).

[37] G. Henkelman, B. P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113**, 9901 (2000).

[38] J.-H. Yang, J.-S. Park, J. Kang, and S.-H. Wei, Phys. Rev. B **91**, 075202 (2015).

[39] G. Y. Kim, A. Senocrate, T.-Y. Yang, G. Gregori, M. Grätzel, and J. Maier, Nat. Mater. **17**, 445 (2018).

[40] A. Walsh and S. D. Stranks, ACS Energy Lett. **3**, 1983 (2018).

[41] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, Comput. Mater. Sci. **111**, 218 (2016).