ROYAL SOCIETY
OF CHEMISTRY

EDGE ARTICLE

Point defect formation at finite temperatures with machine learning force fields†

Irea Mosquera-Lois,[a] Johan Klarbring[ab] and Aron Walsh ⓘD *[a]

Point defects dictate the properties of many functional materials. The standard approach to modelling the thermodynamics of defects relies on a static description, where the change in Gibbs free energy is approximated by the internal energy. This approach has a low computational cost, but ignores contributions from atomic vibrations and structural configurations that can be accessed at finite temperatures. We train a machine learning force field (MLFF) to explore dynamic defect behaviour using Te_i^{+1} and V_{Te}^{+2} in CdTe as exemplars. We consider the different entropic contributions (*e.g.*, electronic, spin, vibrational, orientational, and configurational) and compare methods to compute the defect free energies, ranging from a harmonic treatment to a fully anharmonic approach based on thermodynamic integration. We find that metastable configurations are populated at room temperature and thermal effects increase the predicted concentration of Te_i^{+1} by two orders of magnitude — and can thus significantly affect the predicted properties. Overall, our study underscores the importance of finite-temperature effects and the potential of MLFFs to model defect dynamics at both synthesis and device operating temperatures.

1. Introduction

Point defects make or break material functionality.[1] They limit photovoltaic efficiency by acting as non-radiative recombination centres, control ionic conductivity in batteries, provide active sites for catalytic reactions, and platforms for quantum information technologies. Despite their profound effect on the macroscopic properties of crystals, they are present in dilute concentrations and thus render experimental characterisation challenging. As a result, a combination of experiment and theory is needed to understand defect behaviour.

The key factor when modelling defects is their concentration, which is determined by the free energy of defect formation, g_f, at the synthesis or annealing temperature. Calculating g_f is however computationally challenging, and is thus typically approximated by the formation internal energy, $u_f(0\ K)$, *i.e.* $g_f(T_{synthesis}) \approx u_f(0\ K)$.[2,3] Inherent in this approximation is the assumption of a static framework, where most studies only consider the defect ground state structure at 0 K, and thus neglect metastable configurations that may be populated at the device operating temperature. Since the properties of a defect strongly depend on its geometry,[4–9] the predicted behaviour can be significantly affected when ignoring thermally accessible metastable configurations.[10–16]

With the development of better computational resources, more accurate studies that go beyond this static 0 K approximation are becoming possible using *ab initio* methods. In the last decades, many investigations have modelled entropic contributions for defects in elementary solids.[17–30] Thermal effects are harder to model in multinary semiconductors due to the higher number of possible intrinsic charged defects and required level of theory,[31–61] but have been included for specific defects.[62–80] However, most of these studies adopt several approximations: (i) they only account for vibrational entropies, thereby neglecting other degrees of freedom (*e.g.*, electronic, spin, orientational and configurational), and (ii) they adopt the (quasi)harmonic approximation to model vibrational effects (*e.g.*, assuming a quadratic potential energy surface for the interatomic bonds). The limitations of these approximations are system-dependent and not well investigated, and demonstrate the lack of a reliable and affordable approach to model thermal effects for defects.

In this study, we target these limitations by considering all relevant entropic contributions and systematically comparing the different methods to calculate the defect formation free energy, ranging from a harmonic to a fully anharmonic approach based on thermodynamic integration (TI). To reduce the cost of these simulations, we use machine learning force fields as a surrogate model, which successfully map the defect energy surfaces (Section 2.1). We choose Te_i^{+1} and V_{Te}^{+2} in CdTe as exemplar systems since they display potential energy surfaces of

[a]Thomas Young Centre & Department of Materials, Imperial College London, London SW7 2AZ, UK. E-mail: a.walsh@imperial.ac.uk

[b]Department of Physics, Chemistry and Biology (IFM), Linköping University, SE-581 83, Linköping, Sweden

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4sc08582e

different complexity (*e.g.*, presence *versus* lack of low-energy metastable configurations[12]) (Section 2.2). By modelling the impact of thermal effects on their predicted defect concentrations, we find that these dominate when the defect undergoes symmetry-breaking structural reconstructions and has low-energy metastable configurations, thereby demonstrating the limitations of the idealised 0 K description (Section 2.3).

2. Results and discussion

2.1 Machine learning force fields for defects

Calculating the defect formation free energy, g_f, requires modelling the defect formation reaction by computing the free energy difference between products and reactants at the synthesis temperature.[2,3] For example, the formation of the positively charged tellurium interstitial, Te_i^{+1}, can be described by the defect reaction

$$Te \xrightarrow{CdTe} Te_i^{+1} + e^-. \qquad (1)$$

The corresponding formation free energy is defined from the sum of products (charged defect with an electron in the conduction band) minus the sum of reactants (pristine CdTe host and reservoir of Te). In the standard first-principles supercell formalism, this formation energy is given by

$$g_f = g\left(Cd_nTe_{n+1}^{+1}\right) - g(Te) - g\left([CdTe]_n\right) + E_F + E_{corr} \qquad (2)$$

where Te represents the phase that acts as the external source of atoms during synthesis, E_F denotes the Fermi level and E_{corr} is the correction energy for charged defects. From these terms, the defect free energy, $g(Cd_nTe_{n+1}^{+1})$, is the most challenging to compute due to the low symmetry and large supercells required to model defects (*e.g.*, many force calculations in the (quasi) harmonic method). This is exacerbated when going beyond the harmonic approximation since computing the anharmonic free energy with TI requires many and long molecular dynamics runs.[81,82]

To reduce the cost of free energy calculations, we employ machine learning force fields and train a separate MACE model[83] for each system involved in the defect formation reaction, focussing on a single charge state of each defect. We target temperatures ranging from 100 K to the typical CdTe anneal temperature of 840 K.[84] All models show good accuracies with low mean and root mean square errors on the test set (see Table 1 and further discussion in Methods and ESI†). The accuracy of the defect models is further confirmed by mapping the one-dimensional path between the stable defect structures, which shows good agreement despite the small energy difference between the distinct configurations of Te_i^{+1} (Fig. 1 and ESI Fig. S9†).

2.2 Defect dynamics at room temperature

We first investigate the limitations of the static framework by comparing the behaviour of the defects at 0 K and around the typical operating temperature for a solar cell ($T = 300$ K). The potential energy surface calculated at 0 K shows that Te_i^{+1} is a bistable defect with two accessible structures: a split configuration with either one or two Te–Te bonds,[12] which have C_{2v} and C_s site symmetries, respectively, and an energy difference of $\Delta E(C_s - C_{2v}) = 18$ meV (Fig. 1). In contrast, V_{Te}^{+2} only has one accessible structure at the device operating conditions since the metastable C_{3V} configuration is 1.8 eV above the T_d ground state ($\gg k_BT = 25$ meV at 300 K).

To validate these predictions, we perform molecular dynamics under the *NPT* ensemble (300 K, 1 atm, 1 ns), revealing three distinct motions for Te_i^{+1} (Fig. 2). The fastest process corresponds to changes in *configuration* between the C_{2v} and C_s geometries, which is reflected by variations in the distances between Te_i^{+1} and its neighbouring Te atoms as it alternates between forming 1 and 2 Te–Te bonds. On a slower timescale, there are changes in the defect position (*i.e.*, hopping between lattice sites) as well as changes in the *orientation* of the Te–Te bond, indicated by variation in the angle between the Te–Te bond(s) and the lattice axes. These three motions occur rapidly on the nanosecond timescale due to their low energy barriers relative to the thermal energy ($E_b = 28 - 100$ meV) with rates on the order of 10^{10} s^{-1} (configurational and hopping) and 10^8 s^{-1} (rotation) — and highlight the configurational, orientational and migration degrees of freedom that contribute to the defect formation and migration entropies, respectively. In contrast, this dynamic behaviour of Te_i^{+1} differs significantly from that of V_{Te}^{+2}, which remains stable in its T_d ground state configuration, as depicted in ESI Fig. S12,† and is thus well-described by its static 0 K structure.

2.3 Impact of defect entropy on predicted concentrations

The dynamic behaviour of Te_i^{+1} suggests that its formation entropy, s_f, will be significant at the CdTe annealing temperature (≈ 840 K)[84] and will thus affect the predicted equilibrium concentration. To verify this, we calculate s_f and g_f by considering the different degrees of freedom that change upon forming the defect at a fixed lattice site: electronic, spin, vibrational, orientational and structural (also referred to as configurational).[2] While the first two terms can be estimated with analytical expressions (see Methods), the vibrational contribution is more challenging and typically requires approximations. By assuming a quadratic energy surface for the interatomic bonds, the harmonic vibrational g_f can be calculated, which can be extended to account for thermal expansion with the

Table 1 Mean absolute errors and root mean square errors (shown in parentheses) of the test sets for energies, forces and stresses. The relatively high errors observed for Te are caused by including its liquid phase ($T_{melt} \approx 704$ K). Distributions of the absolute errors and the learning curve for the Te_i^{+1} model are shown in the ESI

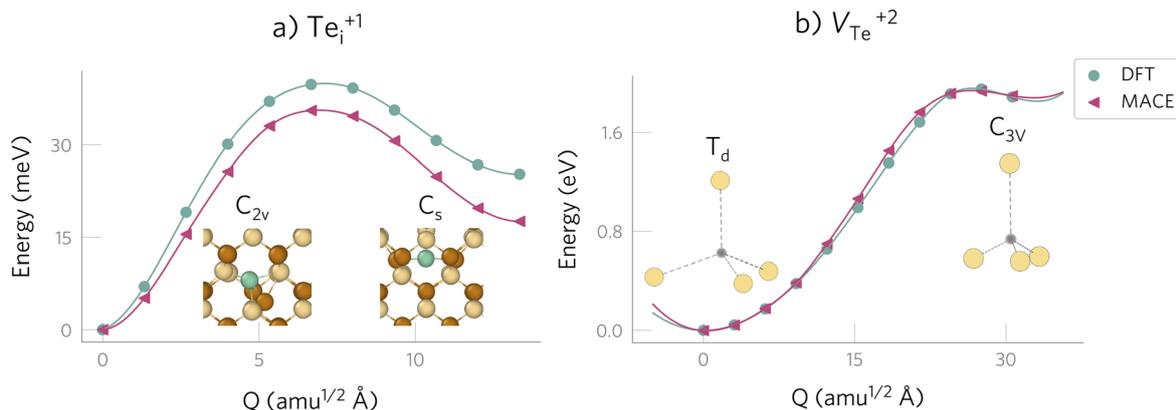| System	Energy (meV per atom)	Force (meV Å^{-1})	Stress (meV Å^{-3})
CdTe	0.3 (0.4)	13 (17)	0.2 (0.3)
Te_i^{+1}	0.5 (0.7)	21 (30)	0.2 (0.3)
V_{Te}^{+2}	0.4 (0.6)	18 (24)	0.2 (0.3)
Te	1.6 (2.3)	73 (102)	0.9 (1.3)

Fig. 1 Potential energy surfaces illustrating the defect configurations identified with defect structure searching,[85] calculated with DFT (green circles) and MACE (pink triangles). The variable Q represents a mass-weighted displacement coordinate that tracks the structural change along the pathway between the defect configurations, as described in the Methods. (a) Te_i^{+1} is a bistable defect since the metastable configuration (split Te–Cd with C_s symmetry) is only 18 meV higher than the ground state (split Te–Te with C_{2v} symmetry). Note that the differences between DFT and MACE are in meV per supercell, with the error in $\Delta E(C_{2v} - C_s)$ only accounting to 0.1 meV per atom. (b) V_{Te}^{+2}, where the metastable configuration (C_{3v}) is significantly higher in energy (1.8 eV above the ground state structure of T_d symmetry). Te in brown, Cd in yellow, Te_i^{+1} in green and V_{Te}^{+2} in shaded grey.
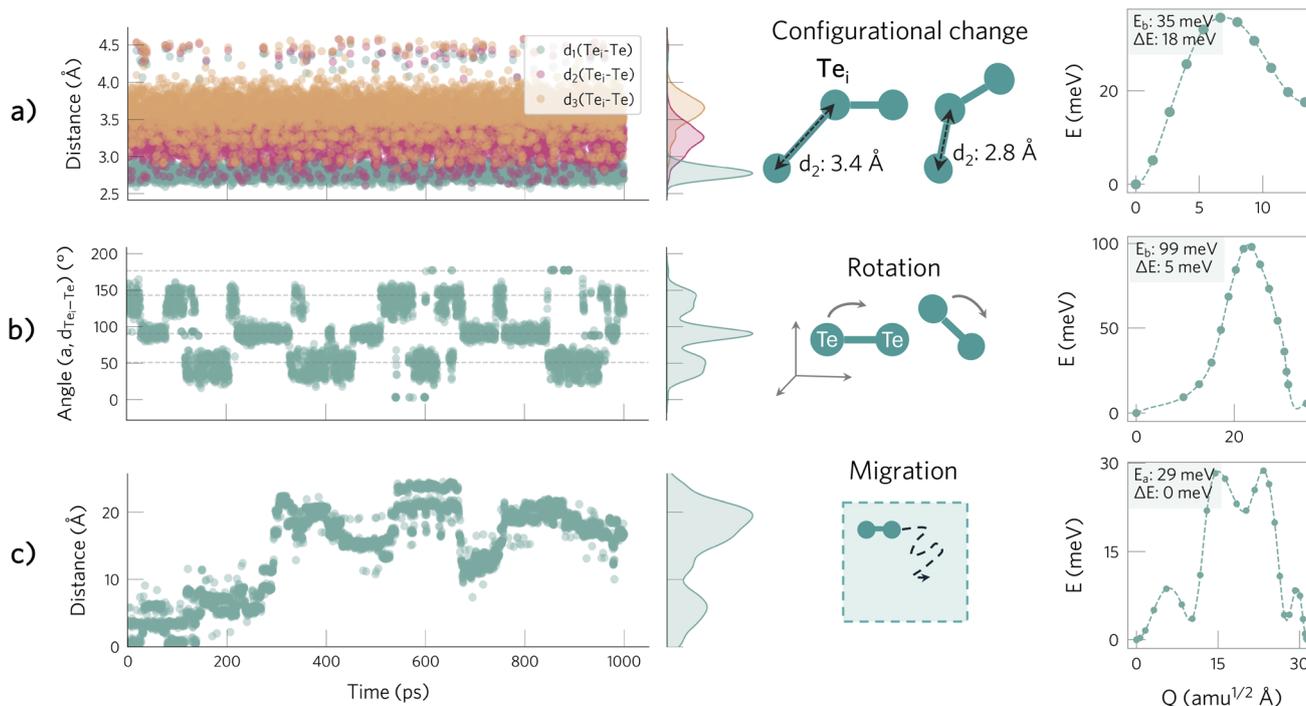

Fig. 2 Active degrees of freedom of Te_i^{+1} at 300 K and associated energy barriers. (a) Configurational change, reflected by changes in the Te_i–Te distances. The narrow green peak illustrates the shortest Te_i–Te bond, which shows little variation at around 2.8 Å, while the wider pink and orange distributions demonstrate the wide variation in the second and third shortest Te–Te distances (see ESI Fig. S11†). (b) Changes in the orientation of the Te_i–Te bond with respect to the [100] direction. (c) Migration, illustrated by tracking the distance between the current and original position of the interstitial.

quasiharmonic approximation.[2,86] However, this harmonic assumption might be limited for defects at high temperatures, where anharmonic effects seem to be important – as suggested by the high anharmonicity scores[87] observed for Te_i^{+1} relative to pristine CdTe ($\sigma(840\ K) = 4.5$ *versus* 0.8, respectively, and σ typically ranging between 0–1; ESI Fig. S13†).

To assess these limitations, we use non-equilibrium thermodynamic integration to account for anharmonic effects and compare the g_f calculated with each approach. We do this by starting from the Einstein crystal (independent harmonic oscillators), integrating to the anharmonic crystal at 100 K, and finally integrating with respect to temperature up to 840 K (see
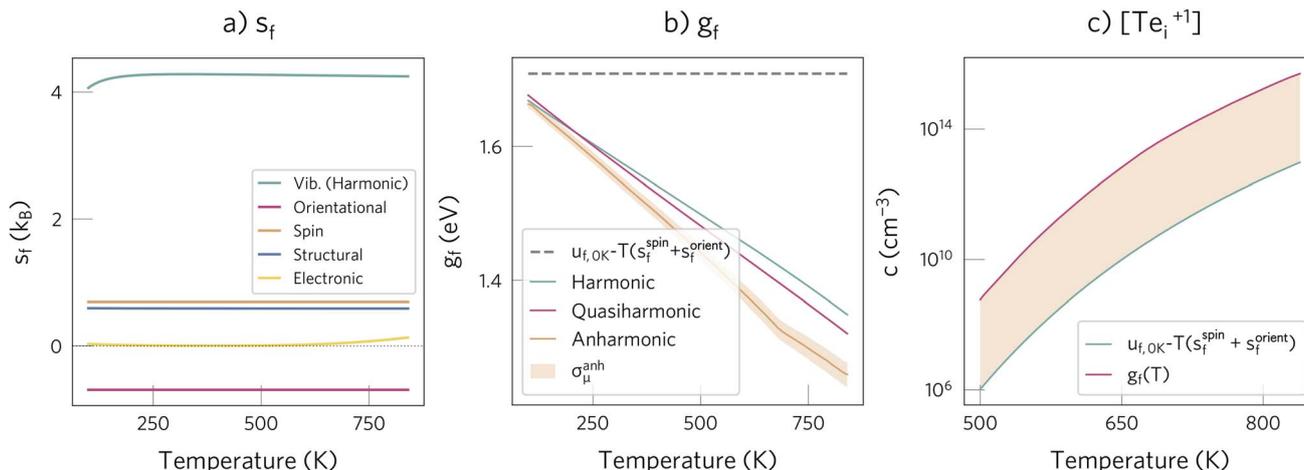
Fig. 3 (a) Contribution of the different degrees of freedom to the formation entropy of Te_i^{+1}. Note that s_f^{orient} is negative since the symmetry increases when going from the initial to the relaxed interstitial structure (see Methods). (b) Comparison of approximations for calculating the defect formation free energy of Te_i^{+1}, $g_f(T)$. The shaded orange area illustrates the estimated error in the thermodynamic integration simulations, defined as the standard error of the mean free energy, σ_μ^{anh} (details in Methods). For comparison, the formation internal energy with the spin and orientational entropies, $u_f(0\ K) - T(s_f^{spin} + s_f^{orient})$, typically used in most defect studies, is shown with a dashed grey line. Note that for Te_i^{+1}, s_f^{spin} and s_f^{orient} cancel each other, thus leading to the grey line having zero slope. (c) Effect of including the entropic contribution for predicting the defect concentration.

Methods for further details). Since TI calculates the change in the total free energy described by the MACE potential (*i.e.*, the ionic degrees of freedom), it already includes the vibrational, orientational and structural contributions, and thus we only have to add the electronic and spin terms to $g_f^{TI}(T)$. We define three defect formation free energies with increasing accuracy:

$$g_f^{harm}(T) = g_f^{vib,harm}(T) - Ts_f^{orient} - Ts_f^{struc}(T) - Ts_f^{spin} - Ts_f^{elec}(T)$$
$$g_f^{quasi}(T) = g_f^{vib,quasi}(T) - Ts_f^{orient} - Ts_f^{struc}(T) - Ts_f^{spin} - Ts_f^{elec}(T)$$
$$g_f^{anh}(T) = g_f^{TI}(T) - Ts_f^{spin} - Ts_f^{elec}(T).$$

$$(3)$$

where the (quasi)harmonic vibrational free energy, $g_f^{vib,harm}$, refers to the ground state structure (further details in Methods). Here, the (quasi)harmonic approach decouples all the degrees of freedom, and thus assumes that the timescales for these processes are sufficiently different to avoid significant mixing,[2] while the anharmonic formalism only decouples the electronic from the ionic motions.

We follow the standard convention in defect chemistry and define g_f as the change in free energy for forming a defect at a fixed lattice site (*i.e.*, excluding entropic contributions from the mixing or site entropy‡). We calculate the equilibrium defect concentration with

$$[c] = \frac{N_{sites}}{V} \exp\left(\frac{-g_f}{k_B T}\right) \qquad (4)$$

where V denotes the crystallographic unit cell volume and N_{sites} the number of symmetry-equivalent sites where the defect can form in the unit cell.

As demonstrated in Fig. 3b, for Te_i^{+1} thermal effects are significant at annealing temperature, with $g_f(840\ K)$ differing by 0.5 eV from $u_f(0\ K) - T(s_f^{spin} + s_f^{orient})$. All methods are in good agreement, indicating that the harmonic approximation gives

a reasonable estimate of g_f, since anharmonic effects approximately cancel out between the bulk and the defect in this temperature range. This agreement also validates the decoupling approximation used to separate the different degrees of freedom (*e.g.*, $g_f^{TI} \approx g_f^{vib,harm} - Ts_f^{orient} - Ts_f^{struc}$). Using this approximation, we find that their relative entropic contributions follow the expected trend, with the vibrational one dominating, followed by the structural, spin, orientational and electronic terms (with $s_f(840\ K)$ of 4.2, 0.7, 0.7, −0.7 and 0.1 k_B, respectively; Fig. 3a). However, we note that the structural term can become larger for defects that have many low-energy metastable configurations.[25]

Overall, the total entropic contribution is not negligible and significantly affects g_f, increasing the predicted concentration by a factor of 500 (Fig. 3c). This importance of entropic effects contrasts with their role in V_{Te}^{+2}, where they are almost negligible ($g_f(840\ K) - u_f(0\ K) = 0.08$ eV; ESI Fig. S14†) due to (i) smaller magnitude of the vibrational entropy and (ii) lack of spin, orientational and structural entropies. As a result, we expect thermal effects to be important for defects which (i) introduce strong structural distortions (high s_f^{vib}), (ii) break the host site symmetry (high s_f^{orient}), and (iii) have low-energy metastable configurations (high s_f^{struc}).

3. Discussion

Overall, we have illustrated how to model thermal effects for point defects in crystals, demonstrating the dynamic character of Te_i^{+1} in CdTe, which rapidly changes between configuration, orientation and position at room temperature. These degrees of freedom increase the entropy upon defect formation, and can be computed from standard defect calculations as illustrated in this study. The computationally challenging term is the vibrational entropy. We have found that the harmonic

approximation gives a reasonable and affordable description of the vibrational formation entropy at 840 K for a dynamic defect with a high anharmonicity score σ^{anh}. While this suggests the validity of the harmonic approximation for 'simpler' defects with single configurations that do not diffuse in this temperature regime, the impact of anharmonicity varies across different defects and host materials, as observed in several metals.[18–20,22,23,26] This highlights the need for further calculations of g_f^{anh} and σ^{anh} on additional defects and host crystals to establish more general guidelines.

By combining the different entropic contributions, we find that thermal effects increase the predicted concentration of Te_i^{+1} by two orders of magnitude, and can thus significantly affect the predicted behaviour by shifting the relative defect populations. Thermal effects will play a significant role for defects that undergo structural reconstructions, break the site symmetry of the host and have low-energy metastable configurations (high s_f^{vib}, s_f^{orient} and s_f^{struc}), as illustrated by comparing two defects with energy surfaces of different complexity. Beyond defect related factors, hosts with a soft and dynamic lattice or compositional disorder will also be more sensitive to thermal effects, since their defects typically lead to stronger reconstructions (*e.g.*, rebonding or local octahedral rotations in perovskite structures[88]) and display many low-energy metastable configurations (*e.g.*, Sb_2Se_3[8] or alloys[89] like $CdSe_x$-$Te_{(1-x)}$[90]). A special case is materials where the phase relevant for applications is only stable at finite temperatures. Here, it can be key to model defects at the device operating temperature since their behaviour can be sensitive to their surrounding structure — as illustrated by the discrepancies when modelling the carrier capture behaviour of I_i in different phases of $CsPbI_3$.[91]

Despite the importance of including thermal effects for accurate defect predictions, the current limitation is the computational cost. While the orientational, spin and electronic terms can be calculated from standard defect calculations using the doped Python package,[92] the configurational term requires considering the different thermally-accessible structures of a defect, which can be identified through defect-structure searching methods like ShakeNBreak.[85] More challenging is the vibrational term as it requires going beyond standard static defect calculations. In practice, accounting for this contribution will only be affordable for high-accuracy studies that target the low-energy defects; especially for applications with a high synthesis or operating temperature, like industrial thermoelectrics, thermochemical water splitting, exhaust automotive catalysts or solid-state fuel cells ($T_{operation} \approx 800$–1900 K).[93–95] In these high temperature applications, thermal effects will populate metastable configurations and could also affect the predicted position of the defect charge transition level (*i.e.*, non-negligible entropic term of $T \times [s_f(q,T) - s_f(q',T)]/(q - q')$) — especially when the change in charge state leads to significant differences in the defect structure and symmetry, spin state and position of the defect level within the bandgap.[80] Finally, based on the performance of our trained models, we note the promise of machine learning force fields for describing the thermal effects of defects. Beyond the

prediction of more accurate concentrations, by learning the defect potential energy surface, they can also reduce the cost of modelling defect dynamics at the device operating temperature and on larger time and length scales, which can be key to predict complex processes like defect reactions or diffusion.[96]

4. Methodology

4.1 Density Functional Theory calculations

All reference calculations were performed with Density Functional Theory using the exchange-correlation functional PBE-sol[97] and the projector augmented wave method,[98] as implemented in the Vienna *Ab initio* Simulation Package (VASP).[99,100] We used the standard PAW PBE potentials (version 64) for Te ($5s^25p^4$) and Cd ($4d^{10}5s^2$). Although hybrid functionals are typically required to accurately model the electronic behaviour of defects, we used a more affordable GGA functional for several reasons: (i) PBEsol has been found to accurately describe the vibrational properties of crystals;[101] (ii) it correctly identifies the same defect configurations reported in a previous study using the HSE06 hybrid functional;[12] and (iii) we aimed to benchmark how to properly train defect MLFFs to reach the high accuracies required to estimate g_f, and as a result needed a functional that would allow a thorough exploration of the configurational landscape up to the CdTe synthesis temperature.

We converged the plane wave energy cutoff and Γ-centered k-point mesh to 1 meV per atom, resulting in values of 450 eV and $4 \times 4 \times 4$ for the conventional cell of CdTe. To minimise Pulay stress errors during molecular dynamics simulations, we increased the converged energy cutoff by 30% (585 eV). The threshold for electronic convergence was set to 10^{-5} eV.

4.2 Training of machine learning force fields

We used the structure similarity kernel implemented in VASP to generate the training sets of configurations using its on-the-fly molecular dynamics approach.[102–105] This involved heating runs performed under the *NPT* ensemble with a pressure of 1 atm and from an initial temperature of 100 K up to 30% above our target temperature of 840 K. In addition, we generated a series of compressed and expanded structures (0.9–1.1 of the original cell volume) to ensure that the model could be used for the quasiharmonic approximation. For bulk CdTe and its defects, we used a $2 \times 2 \times 2$ supercell of the conventional cell (13.0 Å in length and 64 atoms for bulk CdTe). For Te, we included all the low-energy phases available in the Materials Project[106] ($E_{hull} \leq k_B T_{synthesis} = k_B \times (840 \text{ K}) = 0.08$ eV), which were expanded to cubic supercells of at least 10 Å in length. For the liquid Te phase, we generated two models with the packmol code:[107] two cubic boxes of 15 Å and 17.5 Å in length, containing 95 and 220 atoms, respectively, which gave densities matching the reported values in previous studies at our target temperatures ($\rho = 0.027$ atoms per Å^3).[108]

An independent model was trained for each system (bulk CdTe, Te_i^{+1}, V_{Te}^{+2} and Te), with the defect datasets only including one charge state (*e.g.* +1 for Te_i and +2 for V_{Te}). We note that the

models for the charged defects were trained on the absolute DFT energies (*e.g.* without electrostatic corrections, which are applied *a posteriori* when calculating the defect formation energy, as described below). Training separate models for each system lead to higher accuracy models than training one *joint* model on the bulk and defect datasets. However, we have observed that training a model *only* on defect configurations should be avoided if the model will be applied to study the absolute energies of larger system sizes (*e.g.*, defect formation energies). While models selectively trained on defect configurations achieve higher accuracy for defective supercells with the same number of atoms, these models lead to a systematic error in the total energies of larger supercells than the supercells used for training, as explained in detail in ESI Section 1C.†

After generating the training sets with VASP, we trained separate MACE[83] force fields on these datasets to obtain models with higher accuracy and speed. 10% of the configurations in these datasets were used as validation sets to monitor the loss during training. We used a MACE model with Ziegler–Biersack–Littmark (ZBL) pair repulsion,[109] 2 message passing layers, 256 invariant messages, correlation order of 3, angular resolution of 3 and cutoff radius of 5 Å. The batch size was set to 2 and the Huber loss function was used, with weights of 1, 100 and 100 for the mean square errors in the energies, forces and stresses, respectively. For the last 20% of the training epochs, the weights were updated to values of 1000, 10 and 100 for energy, force and stress, respectively — following the recommended strategy of increasing the weight on the energy errors during the final training epochs. The models were trained until the validation loss converged, which required around 150–200 epochs. The reference energies were defined as the potential energies of isolated Cd and Te atoms.

4.3 Validation of machine learning force fields

To generate the test sets, we performed *NPT* molecular dynamics simulations with the trained models at three different temperatures (300, 550 and 900 K), running five independent 24 ps runs at each temperature. We then sampled 100–300 equally-spaced configurations from these trajectories, and performed DFT calculations on them, which were used to calculate the MAE and RMSE on the predicted properties (energy, forces and stresses) of each model (see distribution of sampled configurations and associated errors in Figs. ESI Fig. S1† to ESI Fig. S5†). The MAE and RMSE for the forces and stresses were calculated component wise, as defined in ref. 110. The predicted properties exhibited small absolute errors with no outliers, confirming that the models accurately describe the potential energy surfaces of each system up to 900 K. This extensive testing of the models is especially important for defective crystals, which exhibit more complex energy surfaces than the pristine hosts. In addition, we also validated that the model successfully described the phonons and vibrational free energy of bulk CdTe (see ESI Fig. S6†).

4.4 Defect calculations

Defect calculations were setup and analysed using doped.[92] To account for spurious finite-size supercell effects, the Kumagai–

Oba[111] (eFNV) charge correction scheme was used to calculate E_{corr}, as automated in doped. The Fermi level was assumed to be located in the middle of the band gap.

4.5 Spin degeneracy

The spin degeneracy was calculated with doped using

$$\Omega^{spin} = \frac{Z_d^{spin}}{Z_b^{spin}} = 2S + 1 \tag{5}$$

where Z denotes the partition function and S the total spin angular momentum. For example, Te_i^{+1} has one unpaired electron, resulting in $\Omega^{spin} = 2 \times (1/2) + 1 = 2$. This degeneracy factor Ω can be converted into its respective formation entropy using $s_f = k_B \ln(\Omega)^2$.

4.6 Orientational degeneracy

The orientational degeneracy was also calculated with doped using

$$\Omega^{orient} = \frac{Z_d^{orient}}{Z_b^{orient}} = \frac{N_b}{N_d} \tag{6}$$

where N is the number of symmetry operations of the defect site in the bulk (b) and defective (d) supercells.[2] As discussed in the doped documentation,[92] for vacancies and substitutions there is a clear definition of the defect site in the pristine supercell (*e.g.*, the lattice site where the vacancy/substitution forms). In contrast, for interstitials, the definition of the lattice site in the bulk can be ambiguous, which affects the *partition* between orientational degeneracy and site multiplicity. Here, we follow the definition adopted by Kavanagh *et al.*[92] where the interstitial site in the bulk is defined as the relaxed site of the interstitial but with all other atoms fixed in their bulk (unrelaxed) positions, while the interstitial site in the *defect* supercell corresponds to the relaxed position of *both* the interstitial and all other atoms. Accordingly, the site multiplicity is determined for the lattice site that the interstitial occupies after relaxation and with all other atoms fixed in their bulk positions. Note that other definitions can be adopted and will lead to the same total prefactor ($\Omega^{orient} \times N_{site}$) but different partitions into the orientational and site degeneracies.

For Te_i^{+1}, the C_{2v} configuration has an initial site symmetry of C_s which becomes C_{2v} when the atoms around the interstitial relax due to the formation of the Te–Te dimer. Similarly, the (metastable) C_s configuration has an initial C_1 site symmetry which becomes C_s after the relaxation. As a result, for both configurations the orientational degeneracy Ω^{orient} is 0.5 (*e.g.* the site symmetry increases upon relaxation of the atoms around the interstitial). The site multiplicities per primitive cell are 12 and 24 for the C_{2v} and C_s configurations, respectively. These degeneracies are accounted for when predicting the defect concentration (eqn (4)) where the orientational entropy is included in g_f.

4.7 Electronic entropy

The electronic entropy was calculated using the fixed density of states (DOS) approximation,[56,112–115] that assumes a temperature-independent DOS. Since the electronic entropy is sensitive to the bandgap, and PBEsol significantly underestimates it, we performed self-consistent field HSE06 ($\alpha = 0.345$ (ref. 13)) calculations on the 0 K structures optimised with PBEsol. The electronic entropy is then calculated using

$$S^{elec} = -\gamma k_B \int_{-\infty}^{\infty} D(E)(f(E,T)\ln(f(E,T))$$
$$+ (1 - f(E,T))\ln(1 - f(E,T)))dE \quad (7)$$

where γ equals 1 for spin-polarized systems and 2 for spin-unpolarized systems.[113] $D(E)$ is the electronic density of states at energy E (calculated at 0 K) and $f(E)$ is the occupation of the energy level E given by Fermi–Dirac occupation statistics

$$f(E,T) = \left(\exp\left(\frac{E - E_F}{k_B T}\right) + 1\right)^{-1} \quad (8)$$

with E_F denoting the Fermi level. We define the formation electronic entropy as the entropy change in the reaction $(CdTe)_{32} + Te \rightarrow Cd_{32}Te_{33}^{+1} + e^-$ (ESI Fig. S17†), with E_F for each of these terms calculated as follows:

• Te, CdTe and $Cd_{32}Te_{33}^{+1}$: The Fermi level is assumed to be located mid-gap between the highest occupied and lowest unoccupied state. Note that, in theory, the Fermi level of CdTe should correspond to the *self-consistent* value determined for a set of defects and charge states.[116,117] Determining E_F would require calculating the formation energies for all possible intrinsic defects in CdTe, which is beyond the scope of our study. Accordingly, we assume the Fermi level to be midgap.[38] Te_i^{+1} introduces an empty state 0.75 eV above the valence band maximum (VBM), and thus E_F is set to $0.75/2 = 0.37$ eV above the VBM. Note that the electronic entropy is sensitive to the energy difference between E_F and the lowest unoccupied state, thus requiring an accurate electronic structure.

• Extra electron: It is defined as the excess electronic entropy when one extra electron is added to the conduction band minimum (CBM) of the bulk system, i.e. $s^{elec}(N + 1) - s^{elec}(N)$, where $s^{elec}(N)$ denotes the electronic entropy of bulk CdTe with N and $N + 1$ electrons. To calculate the electronic entropy of the $N + 1$ system, we determine E_F with[118]

$$E_F = E_{CBM} + k_B T \ln\left(\frac{n}{N_C}\right) \quad (9)$$

where N_C denotes the effective density of states in the conduction band, given by

$$N_C = 2\left(\frac{2\pi m_e^* k_B T}{h^2}\right)^{3/2} \quad (10)$$

with m_e^* and h representing the electron effective mass and Planck constant, respectively. For the concentration of excess electrons donated by the defect, n, we assume a value of $n = 10^{15}$ cm^{-3}, which equals our predicted value for the equilibrium concentration of Te_i^{+1} at 840 K (thus making the reaction $(CdTe)_{32} + Te \rightarrow Cd_{32}Te_{33}^{+1} + e^-$ charge neutral). We note that if

all possible point defects were considered, the value of n should be set to the excess electron concentration, as determined self-consistently.[116]

Finally, we note that the electronic entropy can become significant at elevated temperatures ($T \geq 1000$ K, ESI Fig. S17†) if the defect i) introduces an (occupied) empty state close to the (CBM) VBM or ii) changes the occupation of localised d/f bands of nearby cations (e.g., V_O^{+2} reducing two Ce^{+4} to Ce^{+3} in CeO_2), as demonstrated in previous studies.[93,119]

4.8 Structural entropy

The Python code ShakeNBreak was used to identify the defect ground state and metastable configurations.[4,5,85] From these configurations, the structural or configurational entropy can be estimated using

$$s_f^{struc} = -k_B \sum_i p_i \ln(p_i) \quad (11)$$

where p_i denotes the Boltzmann probability of configuration i, given by

$$p_i = \frac{e^{-\frac{G_i}{k_B T}}}{\sum_j e^{-\frac{G_j}{k_B T}}} = \frac{Q_i^{elec} Q_i^{spin} Q_i^{orient} Q_i^{vib} e^{-\frac{U_i}{k_B T}}}{\sum_j Q_j^{elec} Q_j^{spin} Q_j^{orient} Q_j^{vib} e^{-\frac{U_j}{k_B T}}} \quad (12)$$

where G_i and U_i are the Gibbs free energy and internal energy of configuration i, and we have included the degeneracy factors Q since these can be configuration dependent. In practice, the main degrees of freedom that change between configurations are the orientational, spin and vibrational, thus simplifying eqn (12) to

$$p_i = \frac{Q_i^{spin} Q_i^{orient} e^{-\frac{U_i - TS_i^{vib}}{k_B T}}}{\sum_j Q_j^{spin} Q_j^{orient} e^{-\frac{U_j - TS_j^{vib}}{k_B T}}} \quad (13)$$

Beyond applying this analytical approach, we also calculated the structural entropy using the 'inherent structures' method (IS).[25] Within this formalism, we performed *NPT* MD trajectories at 30 temperatures (equally spaced ranging from 100 to 840 K), then sampled 1600 equally-spaced configurations and performed conjugate gradient optimisations to quench the structures to the nearest local minima in the 0 K PES (ESI Fig. S16†). The configurational entropy was then calculated with

$$s_f^{struc} = s_f^{struc}(T_i) + \int_{T_i}^{T_f} \frac{1}{T} \frac{\partial \langle e_{IS}^V(T) \rangle}{\partial T} dT. \quad (14)$$

where $\langle e_{IS}^V(T) \rangle$ denotes the average potential energy of the inherent structures sampled at T, with the volume fixed to the optimal value for the 0 K ground state structure, and T_i and T_f set to 100 and 840 K, respectively. This method resulted in $s_f^{struc}(840 \text{ K}) = 1.05\ k_B$, in the same order of the value of 0.6 k_B obtained with the analytical method. The slightly larger value of s_f^{struc} estimated with the IS formalism likely stems from the additional intermediate configurations sampled with quenching.

Finally, a third approach to account for the structural entropy involves coarse-graining the configurational degree of freedom, and calculating the total defect concentration as a sum over the different configurations i using[2]

$$c = \sum_i c_i = \sum_i \frac{N_i}{V} \exp\left(-\frac{g_{f,i}}{k_B T}\right) \quad (15)$$

where N_i and $g_{f,i}$ denote the number of symmetry-equivalent sites and formation free energy of configuration i (with $g_{f,i} = u_{f,i} - T(s_{f,i}^{\text{vib,harm}} + s_{f,i}^{\text{orient}} + s_{f,i}^{\text{spin}} + s_{f,i}^{\text{elec}})$). From this expression, an effective formation free energy can be obtained by

$$g_{\text{eff}} = -k_B T \ln\left(\frac{V}{N} c\right) \quad (16)$$

where N denotes the number of symmetry-equivalent sites for the ground state structure. By comparing g_{eff} with the g_f calculated using eqn (11), we verified that these values agree (see ESI Fig. S18†).

4.9 Vibrational entropy

The harmonic and quasiharmonic vibrational free energies were calculated using phonopy.[86,120] Within the quasiharmonic framework, which includes the effect of thermal expansion on the phonon frequencies, we generated 11 structures by scaling the supercell volume by factors ranging from 0.9 to 1.10 with a 0.02 increment.

4.10 Thermodynamic integration

Fully anharmonic free energies were calculated using non-equilibrium TI in LAMMPS,[121,122] as implemented in the code calphy.[81] This involved two thermodynamic integration paths: first, we integrate from the Einstein crystal to the anharmonic one at 100 K (Frenkel Ladd method) and then we calculate the temperature variation of the free energy at constant pressure by integrating from 100 K to 840 K (commonly known as reversible scaling). These simulations were performed fixing the center of mass, as done in previous studies.[18,81,122] For each path, we performed 10 independent TI runs to estimate the error, defined as the standard error of the mean free energy ($\sigma_\mu = \sigma/\sqrt{N}$, where σ denotes the standard deviation in the individual free energies calculated from the $N = 10$ independent TI runs). The switching time was adjusted for each system until the error converged to an acceptable value (see ESI Fig. S15†), which must be very low ($\sigma_\mu < 0.25$ meV per atom ≈ 20 meV per supercell) to get an accurate g_f. Since g_f is the difference between two large and similar numbers (the free energies of the bulk and defect supercells), small relative errors in either of these quantities can lead to a large error in g_f. The convergence tests resulted in the switching times reported in Table 2.

We note that during the temperature scaling runs of the interstitial, defect diffusion occurs within the simulation time-scale. This migration, which arises from the shape of the potential energy surface, contributes to the anharmonic free energy through the sampling of intermediate structures during site hopping. We expect that their contribution is small as the defects spend more time around their local minima configurations.

Table 2 Equilibration and switching times in picoseconds used for the thermodynamic integration paths of each system involved in the defect formation reaction. The timestep was set to 2 fs

Path	CdTe	Te_i^{+1}	V_{Te}^{+2}	Te
Einstein → anharmonic	25, 100	25, 200	15, 150	25, 70
100 K → 840 K	25, 170	25, 1000	25, 1000	25, 70
840 K → 500 K	—	—	—	25, 70

Finally, calculating the temperature dependence of the free energy for tellurium is slightly more challenging since it melts at 722 K.[123] Accordingly, we performed two TI simulations:[81] (i) Einstein crystal (100 K) → anharmonic crystal (100 K) → anharmonic crystal (840 K) and (ii) Uhlenbeck–Ford model (840 K) → liquid Te (840 K) → liquid Te (500 K). By comparing the free energies from both simulations, we determined the phase transition temperature and the variation of the free energy with temperature (ESI Fig. S19†). The calculated phase transition temperature is 704 K, which is in good agreement with the experimentally reported value of 722 K.[123]

4.11 Molecular dynamics

To model the behaviour of the defects at room temperature, we performed *NPT* molecular dynamics with LAMMPS[121] using both a 65-atom ($a = 13$ Å) and 513-atom cubic supercells ($a = 26$ Å) to properly capture the dynamics and diffusion of the interstitial. The Nosé–Hoover thermostat and barostat were used (1 atm, 300 K and timestep of 2 fs) with equilibration and production times of 300 ps and 1 ns, respectively. These trajectories were analysed with Python using tools from the ase,[124] pymatgen-analysis-defects,[125,126] pymatgen,[127–129] dscribe,[130,131] umap,[132] direct,[133] matplotlib,[134] and seaborn[135] packages, and visualised with Ovito[136] and CrystalMaker.[137]

The energy barriers for the changes in configuration, orientation and position of Te_i^{+1} were calculated with the Nudged Elastic Band method,[138,139] as implemented in ase.[124] The associated rates for these processes were calculated using Transition State Theory (*e.g.* $k(T) = \nu \exp(-E_b/(k_B T))$) and approximating the attempt frequency ν by the curvature of the PES at the initial state. The reaction pathways between defect configurations were represented in one dimension using a generalised mass-weighted displacement coordinate Q, calculated using

$$Q = \sqrt{\sum_i m_i (r_i^a - r_i^b)^2} \quad (17)$$

where m_i and r_i represent the mass and atomic position of atom i in two configurations a and b. The anharmonicity scores were calculated with FHI-vibes[87,140] on MD trajectories (*NPT* ensemble; 1 atm, 500 ps) at three temperatures (300, 550 and 900 K) for pristine CdTe and Te_i^{+1}.

Data availability

The datasets and trained models are available from the Zenodo repository **https://zenodo.org/records/15224961** (DOI: **https://doi.org/10.1039/d4sc08582e**).

Author contributions

Conceptualisation & project administration: A. W., I. M.-L. Investigation, methodology and formal analysis: I. M.-L. Supervision: A. W. Writing – original draft: I. M.-L. Writing – review & editing: all authors. Resources and funding acquisition: A. W. These author contributions are defined according to the CRediT contributor roles taxonomy.

Conflicts of interest

The authors declare no competing interests.

Acknowledgements

Notes and references

‡ This mixing entropy arises from the different ways in which a defect can be arbitrarily placed at the symmetry-equivalent lattice sites and depends on the equilibrium defect concentration. Due to this dependence on c, it is separated from the free energy of forming the defect at a fixed site (g_f, which is independent of c within the dilute limit) when deriving the expression for the defect concentration.[2]

1 A. M. Stoneham, *Theory of defects in solids*, Oxford University Press, 1975.
2 I. Mosquera-Lois, S. R. Kavanagh, J. Klarbring, K. Tolborg and A. Walsh, *Chem. Soc. Rev.*, 2023, **52**, 5812–5826.
3 C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti and C. G. Van De Walle, *Rev. Mod. Phys.*, 2014, **86**, 253–305.
4 I. Mosquera-Lois, S. R. Kavanagh, A. Walsh and D. O. Scanlon, *npj Comput. Mater.*, 2023, **9**, 1–11.
5 I. Mosquera-Lois and S. R. Kavanagh, *Matter*, 2021, **4**, 2602–2605.
6 X. Wang, S. R. Kavanagh, D. O. Scanlon and A. Walsh, *Phys. Rev. B*, 2023, **108**, 134102.
7 X. Wang, S. R. Kavanagh and A. Walsh, *ACS Energy Lett.*, 2025, **10**(1), 161–167.
8 X. Wang, S. R. Kavanagh, D. O. Scanlon and A. Walsh, *Joule*, 2024, 2105–2122.
9 A. G. Squires, L. Ganeshkumar, C. N. Savory, S. R. Kavanagh and D. O. Scanlon, *ACS Energy Lett.*, 2024, **9**, 4180–4187.
10 J.-H. Yang, L. Shi, L.-W. Wang and S.-H. Wei, *Sci. Rep.*, 2016, **6**, 21712.
11 A. Alkauskas, C. E. Dreyer, J. L. Lyons and C. G. Van de Walle, *Phys. Rev. B*, 2016, **93**, 201304.
12 S. R. Kavanagh, D. O. Scanlon, A. Walsh and C. Freysoldt, *Faraday Discuss.*, 2022, **239**, 339–356.
13 S. R. Kavanagh, A. Walsh and D. O. Scanlon, *ACS Energy Lett.*, 2021, 1392–1398.
14 S. R. Kavanagh, R. S. Nielsen, J. L. Hansen, R. S. Davidsen, O. Hansen, A. E. Samli, P. C. Vesborg, D. O. Scanlon and A. Walsh, *Energy Environ. Sci.*, 2025, DOI: 10.1039/D4EE04647A.
15 D. Guo, C. Qiu, K. Yang and H.-X. Deng, *Phys. Rev. Appl.*, 2021, **15**, 064025.
16 W. B. Fowler, M. Stavola, A. Venzie and A. Portoff, *J. Appl. Phys.*, 2024, **135**, 170901.
17 A. S. Bochkarev, A. van Roekeghem, S. Mossa and N. Mingo, *Phys. Rev. Mater.*, 2019, **3**, 093803.
18 B. Cheng and M. Ceriotti, *Phys. Rev. B*, 2018, **97**, 054102.
19 S. Chiesa, P. M. Derlet and S. L. Dudarev, *Phys. Rev. B*, 2009, **79**, 214109.
20 M. de Koning, C. R. Miranda and A. Antonelli, *Phys. Rev. B*, 2002, **66**, 104110.
21 M. De Koning, S. Ramos De Debiaggi and A. Monti, *Defect Diffus. Forum*, 2003, **224–225**, 59–74.
22 A. Glensk, B. Grabowski, T. Hickel and J. Neugebauer, *Phys. Rev. X*, 2014, **4**, 011018.
23 B. Grabowski, L. Ismer, T. Hickel and J. Neugebauer, *Phys. Rev. B*, 2009, **79**, 134106.
24 G. Lucas and R. Schäublin, *Nucl. Instrum. Methods Phys. Res.*, 2009, **267**, 3009–3012.
25 J. Luo, C. Zhou, Y. Cheng, Q. Li, L. Liu, J. F. Douglas and T. Sinno, *Phys. Rev. Mater.*, 2022, **6**, 064603.
26 T. A. Mellan, A. I. Duff, B. Grabowski and M. W. Finnis, *Phys. Rev. B*, 2019, **100**, 024303.
27 E. V. Safonova, Y. P. Mitrofanov, R. A. Konchakov, A. Y. Vinogradov, N. P. Kobelev and V. A. Khonik, *J. Phys.: Condens. Matter*, 2016, **28**, 215401.
28 A. Satta, F. Willaime and S. de Gironcoli, *Phys. Rev. B*, 1998, **57**, 11184–11192.
29 G. S. Smirnov and V. V. Stegailov, *J. Phys.: Condens. Matter*, 2019, **31**, 235704.
30 D. Shin and C. Wolverton, *Acta Mater.*, 2012, **60**, 5135–5142.
31 Y. Gong, B. Grabowski, A. Glensk, F. Körmann, J. Neugebauer and R. C. Reed, *Phys. Rev. B*, 2018, **97**, 214106.
32 D. Smirnova, S. Starikov, G. D. Leines, Y. Liang, N. Wang, M. N. Popov, I. A. Abrikosov, D. G. Sangiovanni, R. Drautz and M. Mrovec, *Phys. Rev. Mater.*, 2020, **4**, 013605.
33 X. Zhang, B. Grabowski, T. Hickel and J. Neugebauer, *Comput. Mater. Sci.*, 2018, **148**, 249–259.
34 L. Mathes, T. Gigl, M. Leitner and C. Hugenschmidt, *Phys. Rev. B*, 2020, **101**, 134105.
35 T. Sinno, Z. K. Jiang and R. A. Brown, *Appl. Phys. Lett.*, 1996, **68**, 3028–3030.

36 O. K. Al-Mushadani and R. J. Needs, *Phys. Rev. B*, 2003, **68**, 235205.

37 E. Rauls and T. Frauenheim, *Phys. Rev. B*, 2004, **69**, 155213.

38 P. E. Blöchl, E. Smargiassi, R. Car, D. B. Laks, W. Andreoni and S. T. Pantelides, *Phys. Rev. Lett.*, 1993, **70**, 2435–2438.

39 D. Maroudas and R. A. Brown, *Phys. Rev. B*, 1993, **47**, 15562–15577.

40 M. I. Mendelev and Y. Mishin, *Phys. Rev. B*, 2009, **80**, 144111.

41 P. J. Ungar, T. Halicioglu and W. A. Tiller, *Phys. Rev. B*, 1994, **50**, 7344–7357.

42 P. Wynblatt, *J. Phys. Chem. Solids*, 1969, **30**, 2201–2211.

43 J. H. Harding, *Physica B+C*, 1985, **131**, 13–26.

44 J. H. Harding and A. M. Stoneham, *Philos. Mag. B*, 1981, **43**, 705–713.

45 J. H. Harding, *Phys. Rev. B*, 1985, **32**, 6861.

46 J. Luo, C. Zhou, Q. Li and L. Liu, *J. Chem. Phys.*, 2022, **156**, 214113.

47 J. Luo, C. Zhou, Q. Li and L. Liu, *Materials*, 2022, **15**, 4026.

48 Y. Mishin, M. R. Sorensen and A. F. Voter, *Philos. Mag. A*, 2001, **81**, 2591–2612.

49 P. Wynblatt, *Phys. Status Solidi B*, 1969, **36**, 797–808.

50 C. Lapointe, T. D. Swinburne, L. Proville, C. S. Becquart, N. Mousseau and M.-C. Marinica, *Phys. Rev. Mater.*, 2022, **6**, 113803.

51 P. W. M. Jacobs, *J. Chem. Soc., Faraday Trans.*, 1990, **86**, 1197–1201.

52 J. Nam and R. Gómez-Bombarelli, *arXiv*, 2024, preprint, arXiv:2404.10746, DOI: 10.48550/arXiv.2404.10746.

53 S. Ramos de Debiaggi, M. de Koning and A. M. Monti, *Phys. Rev. B*, 2006, **73**, 104103.

54 K. M. Carling, G. Wahnström, T. R. Mattsson, N. Sandberg and G. Grimvall, *Phys. Rev. B*, 2003, **67**, 054101.

55 S. M. Foiles, *Phys. Rev. B*, 1994, **49**, 14930–14938.

56 S. K. Estreicher, M. Sanati, D. West and F. Ruymgaart, *Phys. Rev. B*, 2004, **70**, 125209.

57 M. Sanati and S. K. Estreicher, *Phys. B*, 2003, **340–342**, 630–636.

58 M. Sanati and S. K. Estreicher, *Solid State Commun.*, 2003, **128**, 181–185.

59 M. Sanati and S. K. Estreicher, *Phys. Rev. B*, 2005, **72**, 165206.

60 E. Smargiassi and R. Car, *Phys. Rev. B*, 1996, **53**, 9760–9763.

61 C. Catlow, J. Corish, P. Jacobs and A. Lidiard, *J. Phys. C:Solid State Phys.*, 1981, **14**, L121.

62 P. Ágoston and K. Albe, *Phys. Chem. Chem. Phys.*, 2009, **11**, 3226–3232.

63 T. Zacherle, P. C. Schmidt and M. Martin, *Phys. Rev. B*, 2013, **87**, 235206.

64 T. Smith, S. Moxon, J. S. Tse, J. M. Skelton, D. J. Cooke, L. J. Gillie, E. L. d. Silva, R. M. Harker, M. T. Storr, S. C. Parker and M. Molinari, *J. Phys.: Energy*, 2023, **5**, 025004.

65 T. Zacherle, A. Schriever, R. A. De Souza and M. Martin, *Phys. Rev. B*, 2013, **87**, 134104.

66 A. Walsh, A. A. Sokol and C. R. A. Catlow, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2011, **83**, 224105.

67 B. Baldassarri, J. He and C. Wolverton, *Phys. Rev. Mater.*, 2024, **8**, 055407.

68 S. L. Millican, J. M. Clary, C. B. Musgrave and S. Lany, *Chem. Mater.*, 2022, **34**, 519–528.

69 S. Moxon, J. Skelton, J. S. Tse, J. Flitcroft, A. Togo, D. J. Cooke, E. L. d. Silva, R. M. Harker, M. T. Storr, S. C. Parker and M. Molinari, *J. Mater. Chem. A*, 2022, **10**, 1861–1875.

70 Y. Sun, T. Liu, Q. Chang and C. Ma, *J. Phys. Chem. Solids*, 2018, **115**, 228–232.

71 G. Miceli and A. Pasquarello, *Phys. Rev. B*, 2016, **93**, 165207.

72 S. Grieshammer, T. Zacherle and M. Martin, *Phys. Chem. Chem. Phys.*, 2013, **15**, 15935–15942.

73 C. Cazorla, *Phys. Rev. Appl.*, 2017, **7**, 044025.

74 J. M. Wynn, R. J. Needs and A. J. Morris, *arXiv*, 2016, preprint, arXiv:1609.04760, DOI: 10.48550/arXiv.1609.04760.

75 M. Youssef and B. Yildiz, *Phys. Rev. B*, 2012, **86**, 144109.

76 L. N. Holtzman, P. A. Vargas, R. G. Hennig and K. Barmak, *J. Chem. Phys.*, 2024, **161**, 144105.

77 A. Gorfer, R. Abart and C. Dellago, *Phys. Rev. Mater.*, 2024, **8**, 073602.

78 R. J. Tarento and J. H. Harding, *J. Phys. C: Solid State Phys.*, 1987, **20**, L677.

79 C. Zhang, F. Gygi and G. Galli, *Phys. Rev. Mater.*, 2024, **8**, 046201.

80 T. S. Bjørheim, M. Arrigoni, D. Gryaznov, E. Kotomin and J. Maier, *Phys. Chem. Chem. Phys.*, 2015, **17**, 20765–20774.

81 S. Menon, Y. Lysogorskiy, J. Rogal and R. Drautz, *Phys. Rev. Mater.*, 2021, **5**, 103801.

82 V. Kapil, E. Engel, M. Rossi and M. Ceriotti, *J. Chem. Theory Comput.*, 2019, **15**, 5845–5857.

83 I. Batatia, D. P. Kovacs, G. Simm, C. Ortner and G. Csanyi, *Adv. Neural Inf. Process. Syst.*, 2022, **35**, 11423–11436.

84 W. K. Metzger, S. Grover, D. Lu, E. Colegrove, J. Moseley, C. L. Perkins, X. Li, R. Mallick, W. Zhang, R. Malik, J. Kephart, C.-S. Jiang, D. Kuciauskas, D. S. Albin, M. M. Al-Jassim, G. Xiong and M. Gloeckler, *Nat. Energy*, 2019, **4**, 837–845.

85 I. Mosquera-Lois, S. R. Kavanagh, A. Walsh and D. O. Scanlon, *J. Open Source Softw.*, 2022, **7**, 4817.

86 A. Togo, L. Chaput, T. Tadano and I. Tanaka, *J. Phys.:Condens. Matter*, 2023, **35**, 353001.

87 F. Knoop, T. A. R. Purcell, M. Scheffler and C. Carbogno, *Phys. Rev. Mater.*, 2020, **4**, 083809.

88 M. Choi, F. Oba, Y. Kumagai and I. Tanaka, *Adv. Mater.*, 2013, **25**, 86–90.

89 J. Park, B. Xu, J. Pan, D. Zhang, S. Lany, X. Liu, J. Luo and Y. Qi, *npj Comput. Mater.*, 2023, **9**, 1–13.

90 I. Mosquera-Lois, S. R. Kavanagh, A. M. Ganose and A. Walsh, *npj Comput. Mater.*, 2024, **10**, 1–9.

91 L. D. Whalley, *J. Phys. Chem. C*, 2023, **127**, 15738–15746.

92 S. R. Kavanagh, A. G. Squires, A. Nicolson, I. Mosquera-Lois, A. M. Ganose, B. Zhu, K. Brlec, A. Walsh and D. O. Scanlon, *J. Open Source Softw.*, 2024, **9**, 6433.

93 S. S. Naghavi, A. A. Emery, H. A. Hansen, F. Zhou, V. Ozolins and C. Wolverton, *Nat. Commun.*, 2017, **8**, 285.

94 D. Sikstrom and V. Thangadurai, *Ionics*, 2024, **1**, 1.

95 J. Gao, G. Tian, A. Sorniotti, A. E. Karci and R. Di Palo, *Appl. Therm. Eng.*, 2019, **147**, 177–187.

96 X. Zhong, F. Höfling and T. John, *Geochem., Geophys., Geosyst.*, 2025, **26**, e2024GC011951.

97 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys. Rev. Lett.*, 2008, **100**, 136406.

98 G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.

99 G. Kresse and J. Hafner, *Phys. Rev. B*, 1993, **47**, 558–561.

100 G. Kresse and J. Hafner, *Phys. Rev. B*, 1994, **49**, 14251–14269.

101 J. M. Skelton, D. Tiana, S. C. Parker, A. Togo, I. Tanaka and A. Walsh, *J. Chem. Phys.*, 2015, **143**, 064710.

102 R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse and M. Bokdam, *Phys. Rev. Lett.*, 2019, **122**, 225701.

103 R. Jinnouchi, F. Karsai and G. Kresse, *Phys. Rev. B*, 2019, **100**, 014105.

104 R. Jinnouchi, K. Miwa, F. Karsai, G. Kresse and R. Asahi, *J. Phys. Chem. Lett.*, 2020, **11**, 6946–6955.

105 P. Liu, C. Verdi, F. Karsai and G. Kresse, *Phys. Rev. B*, 2022, **105**, L060102.

106 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.

107 L. Martínez, R. Andrade, E. G. Birgin and J. M. Martínez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.

108 J. Akola, R. O. Jones, S. Kohara, T. Usuki and E. Bychkov, *Phys. Rev. B*, 2010, **81**, 094202.

109 J. F. Ziegler and J. P. Biersack, in *The Stopping and Range of Ions in Matter*, ed. D. A. Bromley, Springer US, Boston, MA, 1985, pp. 93–129.

110 J. D. Morrow, J. L. A. Gardner and V. L. Deringer, *J. Chem. Phys.*, 2023, **158**, 121501.

111 Y. Kumagai and F. Oba, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 2014, **89**, 195205.

112 O. Eriksson, J. M. Wills and D. Wallace, *Phys. Rev. B*, 1992, **46**, 5221–5228.

113 X. Zhang, B. Grabowski, F. Körmann, C. Freysoldt and J. Neugebauer, *Phys. Rev. B*, 2017, **95**, 165126.

114 A. Metsue, A. Oudriss, J. Bouhattate and X. Feaugas, *J. Chem. Phys.*, 2014, **140**, 104705.

115 F. Willaime, A. Satta, M. Nastar and O. Le Bacq, *Int. J. Quantum Chem.*, 2000, **77**, 927–939.

116 A. G. Squires, D. O. Scanlon and B. J. Morgan, *J. Open Source Softw.*, 2023, **8**, 4962.

117 J. Buckeridge, *Comput. Phys. Commun.*, 2019, **244**, 329–342.

118 A. H. M. Smets, K. Jäger, O. Isabella, R. A. v. Swaaij and M. Zeman, *Solar energy : the physics and engineering of photovoltaic conversion, technologies and systems*, UIT Cambridge Ltd, Cambridge, England, 2016.

119 S. Lany, *J. Chem. Phys.*, 2018, **148**, 071101.

120 A. Togo and I. Tanaka, *Scr. Mater.*, 2015, **108**, 1–5.

121 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Comput. Phys. Commun.*, 2022, **271**, 108171.

122 R. Freitas, M. Asta and M. de Koning, *Comput. Mater. Sci.*, 2016, **112**, 333–341.

123 F. C. Kracek, *J. Am. Chem. Soc.*, 1941, **63**, 1989–1990.

124 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *World J. Condens. Matter Phys.*, 2017, **29**, 273002.

125 J.-X. Shen and J. Varley, *J. Open Source Softw.*, 2024, **9**, 5941.

126 J.-X. Shen, L. F. Voss and J. B. Varley, *J. Appl. Phys.*, 2024, **135**, 145102.

127 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2013, **68**, 314–319.

128 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.

129 S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder and K. A. Persson, *Comput. Mater. Sci.*, 2015, **97**, 209–215.

130 L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, *Comput. Phys. Commun.*, 2020, **247**, 106949.

131 J. Laakso, L. Himanen, H. Homm, E. V. Morooka, M. O. Jäger, M. Todorović and P. Rinke, *J. Chem. Phys.*, 2023, **158**, 234802.

132 L. McInnes, J. Healy, N. Saul and L. Großberger, *J. Open Source Softw.*, 2018, **3**, 861.

133 J. Qi, T. W. Ko, B. C. Wood, T. A. Pham and S. P. Ong, *npj Comput. Mater.*, 2024, **10**, 43.

134 J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 90–95.

135 M. L. Waskom, *J. Open Source Softw.*, 2021, **(6)**, 3021.

136 A. Stukowski, *Modell. Simul. Mater. Sci. Eng.*, 2010, **18**, 015012.

137 *CrystalMaker®: a crystal and molecular structures program for Mac and Windows*. CrystalMaker Software Ltd, Oxford, England (**https://www.crystalmaker.com**.

138 G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.

139 G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9978–9985.

140 F. Knoop, T. A. R. Purcell, M. Scheffler and C. Carbogno, *J. Open Source Softw.*, 2020, **5**, 2671.