Phys. Status Solidi B **251**, No. 1, 97–129 (2014) / **DOI** 10.1002/pssb.201350155

physica status solidi b
www.pss-b.com
basic solid state physics

Part of Topical Section on
Computational Thermodynamics

Feature Article

Perspectives on point defect thermodynamics

Jutta Rogal[*,1], Sergiy V. Divinski[2], Mike W. Finnis[3], Albert Glensk[4], Jörg Neugebauer[4], John H. Perepezko[5], Sergej Schuwalow[1], Marcel H. F. Sluiter[6], and Bo Sundman[7]

[1] Interdisciplinary Centre for Advanced Materials Simulation, Ruhr-Universität Bochum, 44780 Bochum, Germany
[2] Institute of Materials Physics, Westphalien-Wilhelms University of Münster, 48149 Münster, Germany
[3] Department of Physics and Department of Materials, Thomas Young Centre, Imperial College London, Exhibition Road, London SW7 2AZ, UK
[4] Department of Computational Materials Design, Max-Planck-Institut für Eisenforschung GmbH, Max-Planck-Straße 1, 40237 Düsseldorf, Germany
[5] Department of Materials Science and Engineering, University of Wisconsin-Madison, 1509 University ave., Madison, WI 53706, USA
[6] Department of Materials Science and Engineering, Delft University of Technology, Mekelweg 2, 2628CD Delft, The Netherlands
[7] Department of Materials Science and Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden

Received 28 August 2013, revised 15 November 2013, accepted 19 November 2013
Published online 27 December 2013

Keywords CALPHAD, defects, vacancies

[*] Corresponding author: e-mail jutta.rogal@rub.de, Phone: +49 234 3229317, Fax: +49 234 3214977

We review and discuss methods for including the role of point defects in calculations of the free energy, composition and phase stability of elements and compounds. Our principle aim is to explain and to reconcile, with examples, the perspectives on this problem that are often strikingly different between exponents of CALPHAD, and others working in the overlapping fields of physics, chemistry and materials science. Current methodologies described here include the compound energy formalism of CALPHAD, besides the rather different but related canonical and grand-canonical formalisms. We show how the calculation of appropriate defect formation energies should be formulated, how they are included in the different formalisms and in turn how these yield equilibrium defect concentrations and their contribution to free energies and chemical potentials. Furthermore, we briefly review the current state-of-the-art and challenges in determining point defect properties from first-principles calculations as well as from experimental measurements.

1 Introduction When modelling thermodynamic properties of materials, it is important to include the contribution of point defects to the Gibbs energy. Defect properties can be obtained from experimental measurements and more recently also from first-principles calculations. It is desirable that the most reliable defect properties are available to enter the databases used in thermodynamic modelling and thereby to expand existing databases with experimental data as well as with complementary and accurate point defect data from first-principles calculations.

The approaches to including the contributions of point defects, in particular vacancies, to the Gibbs energy is quite different in the CALPHAD (thermodynamic modelling) and physics communities. The CALPHAD community has developed an expression for the Gibbs energy in which vacancies are an additional component. This is advantageous, because the vacancies become a member of a set of reference states of elements or compounds ('end-members') within an established formalism, in which the properties of the reference states are independent of the multicomponent phases whose Gibbs energy is modelled. It is an approach very naturally suited to the standard CALPHAD concept of a phase being defined by its sublattices and the set of constituents (real elements or vacancies) that are allowed to occupy the sites of each sublattice.

On the other hand, while physicists and crystal chemists generally also start from an ansatz for the Gibbs energy [1], their approach to modelling the contribution of point defects usually starts from the defect *formation energies*, and not from any hypothetical materials-independent energy with which they may be endowed. This is because the intention is quite different, namely to investigate a single material rather

than to deduce the properties of a multicomponent system over an extensive range of compositions, temperatures and pressures, from a database of the properties of each of its pure components. Such a 'defect-centric' approach, featuring small deviations from stoichiometry, is, however, inadequate to describe the Gibbs energy of phases over a wide range of compositions. Here, more empirical modelling approaches like CALPHAD are indispensable, and should be supported as far as possible by experimental and theoretical data.

There are now a growing amount of more reliable theoretical data on point defects, however, to be able to include it in CALPHAD databases it is imperative to understand how the two approaches are connected. In this paper, we aim to draw these communities together by explaining and illustrating some of the similarities and differences, so that they may better understand each others' language and purpose. We explain with the help of detailed examples, including pure metals, NiAl and GaN, how the compound energy formalism (CEF) often used by exponents of CALPHAD, can be related to the canonical and grand canonical formalisms favoured by physicists and chemists, and applied in the theory of point defect concentrations and their effect on the chemical potentials of the components. The hope is that this will speed the application of modern first-principles methods of calculation of defect properties where they could be helpful in the development of the CALPHAD methodology. This is the purpose of Section 2.

The plan of the rest of our paper is as follows. In Section 3, we describe the treatment of charged point defects, exemplified here by GaN. This topic merits special attention, since it entails several conceptual and computational difficulties, and is of major importance in the physics of semiconductors, where the effect of electron trapping on charge transport is critical, but also in ceramics, notably when dealing with oxides and their ionic conductivity, of importance in modelling fuel cells, gas sensors and corrosion. In Section 4 we review the current state-of-the-art and challenges in determining point defect properties for elements and disordered alloys from first-principles calculations. Finally, a brief review of the experimental techniques and measurements is given in Section 5, before we conclude in Section 6.

2 Connecting *ab initio* atomistic and CALPHAD formulations

Historically, there are several ways to model the thermodynamics of defects in crystalline phases. As will be shown most of these models can be cast mathematically in a form that makes them special cases of the CEF. This will be explained after introducing a classic treatment of vacancy defects, which might be more familiar to the physics or crystal chemistry communities.

2.1 The single vacancy: Canonical formalism

Here and in the following sections we choose as the fundamental thermodynamic quantity the Gibbs energy, $G(T, P, \{n_j\})$, which will refer to a system that can exchange heat and volume with its environment, maintained at constant temperature T and pressure P, but in which the number of atoms of each species, n_j, is constant. The Gibbs energy is useful for its property that, in such a system if any internal constraints, real or notional, are removed G will spontaneously decrease in value until it is minimized when the system is in equilibrium again. This well known property is central to the way we model the energy and concentration of vacancies or other point defects.

For studying vacancies, the atoms and vacancies are assumed to occupy the same sites in a crystal lattice. As we shall see, in applying the Gibbs energy minimization, variation in the number of vacancies implies that the number of sites is not conserved. Initially, we shall treat a system of one component, labelled A, containing n_A atoms and n_{Va} vacancies. Let us first adopt the Gibbsian viewpoint that vacancies are *not* themselves a component. In this case, we can think of the number of vacancies as the constraint that is relaxed until the Gibbs energy is minimized. In the process the number of lattice sites has to be adjusted, since the number of atoms is kept fixed. Within the ideal solution model, the Gibbs energy can be written in the form:

$$G = n_A G_A + n_{Va} G^f_{Va} - (n_A + n_{Va}) T S_{mix}\left(y_A, y_{Va}\right). \qquad (1)$$

The first term $n_A G_A(T, P)$ is the Gibbs energy of a hypothetical defect-free crystal of pure A, in which $G_A(T, P)$ would be its chemical potential. The second term $n_{Va} G^f_{Va}(T, P)$ is the contribution of vacancies in the dilute limit, including their vibrational free energy, and the thermal energy and entropy of electrons. As we shall see, $G^f_{Va}(T, P)$ turns out to be the Gibbs energy of vacancy formation, or what is generally known as the *vacancy formation energy* at constant pressure. Fortunately, the timescale for diffusion is very much slower than a period of lattice vibration, so the diffusional contribution of vacancies to the entropy can be decoupled from their vibrational and other contributions. Thus, the so-called configurational entropy of vacancies, that is to say the part contributed by their random positions, is contained in S_{mix}, given by

$$S_{mix}\left(y_A, y_{Va}\right) = -k_B[y_{Va}\ln\left(y_{Va}\right) + \left(1 - y_{Va}\right)\ln\left(1 - y_{Va}\right)]. \qquad (2)$$

The S_{mix} term assumes the ideal solution model, and is therefore independent of temperature and pressure. Any corrections to this assumption will have to be subsumed later in an additional empirical term within the regular solution model.

The vacancy concentration per site is defined by

$$y_{Va} = \frac{n_{Va}}{n_A + n_{Va}}. \qquad (3)$$

To economize on notation for future use, we let $\{y_i\}$ stand for the set of constituent fractions, or site occupancies, of all the species present. The constituent fraction of species i is y_i. We use the term 'constituent' here and elsewhere in the

paper to embrace both the physical components and lattice vacancies. In general, the ideal configurational entropy per lattice site is given by

$$S_{\mathrm{mix}}(\{y_i\}) := -k_{\mathrm{B}} \sum_i y_i \ln y_i \tag{4}$$

$$\text{where} \qquad \sum_i y_i = 1. \tag{5}$$

Equation (1) is also the basis of a strategy to calculate the vacancy formation energy. The procedure for calculation would run as follows. Consider a crystal of N lattice sites. These usually define a single supercell in a computer model with periodic boundary conditions. Compare two such supercells, the first defect-free and the second containing just one vacancy located on just one of the sites. The Gibbs energy of the first system G_{perfect} is simply $N\,G_{\mathrm{A}}(T, P)$. And in the second system, on the timescales before diffusion is significant our knowledge of the position of the vacancy is complete, so there is no mixing entropy, thus the Gibbs energy $G_{1\mathrm{Va}}$ of the second system according to (1) is

$$G_{1\mathrm{Va}}(T, P) = (N-1)G_{\mathrm{A}}(T, P) + G^f_{\mathrm{Va}}(T, P). \tag{6}$$

The vacancy formation energy is identified as

$$G^f_{\mathrm{Va}} := G_{1\mathrm{Va}}(T, P) - \frac{N-1}{N} G_{\mathrm{perfect}}(T, P) \tag{7}$$

$$\equiv G_{1\mathrm{Va}}(T, P) - N\,G_{\mathrm{A}}(T, P) + G_{\mathrm{A}}(T, P). \tag{8}$$

There are several points to be made about Eq. (7). Firstly, the system, usually a supercell with periodic boundary conditions although possibly a large cluster, should be big enough for this result to be independent of N, otherwise surface effects, or spurious image interactions, would spoil the result. Secondly, although to be formally correct we have included temperature and pressure dependence, most calculations in the literature until recently assumed $T = 0\,\mathrm{K}$ due to the computational difficulty associated with the calculation of the vibrational contribution to the Gibbs energy at $T > 0\,\mathrm{K}$. Thirdly, it has recently been discovered by careful density-functional theory (DFT) calculations, as described in Section 4.4, that the harmonic or quasi-harmonic approximation to the effect of $T > 0\,\mathrm{K}$ is probably inadequate in many cases at high temperatures, at which the vacancy concentrations can actually be measured experimentally. Finally, although in principle a regular solution term could be added to Eq. (7) in order to account for vacancy–vacancy interactions, in practice we do not expect this to be important, since even at the melting point the vacancy concentrations in elemental metals are below 10^{-2} per lattice site.

Returning to Eq. (1), we now have to vary the number of vacancies in order to minimize G while keeping the system closed. In this process, we must allow the number of lattice sites $N = n_{\mathrm{A}} + n_{\mathrm{Va}}$ to vary. That is:

$$\left(\frac{\partial G}{\partial n_{\mathrm{Va}}}\right)_{T,P,n_{\mathrm{A}}} = 0. \tag{9}$$

The result, after differentiating Eq. (1), which involves applying the chain rule to Eq. (3), is

$$y^*_{\mathrm{Va}} = \exp\left(-\frac{G^f_{\mathrm{Va}}}{k_{\mathrm{B}}T}\right). \tag{10}$$

For clarity, here and subsequently, we distinguish the variables evaluated at equilibrium, i.e. corresponding to a minimized Gibbs potential, by an asterisk. The result in Eq. (10) reduces to the simple Arrhenius relation for vacancy concentration unless the temperature dependence of G^f_{Va} is anything other than linear. The assumption of a constant entropy of formation of the defect would introduce linear temperature dependence in G^f_{Va} and hence a constant prefactor to the Arrhenius formula. Within the harmonic approximation, in the high temperature limit, the free energy of each mode of vibration is simply $3k_{\mathrm{B}}T$, with or without a vacancy, which therefore does not affect the Arrhenius exponential.

The corresponding equilibrium value of the Gibbs energy is obtained by substituting y^*_{Va} into Eq. (1), and it can be simplified to the familiar form:

$$G^* = n_{\mathrm{A}} G_{\mathrm{A}} + n_{\mathrm{A}} k_{\mathrm{B}} T \ln(1 - y^*_{\mathrm{Va}}). \tag{11}$$

We can also express G^*, as for any homogeneous phase, by its Euler equation, which for just one component is simply

$$G^* = n_{\mathrm{A}} \mu_{\mathrm{A}}. \tag{12}$$

The identity of Eqs. (11) and (12) tells us that the chemical potential per atom of pure A is given in equilibrium by

$$\mu_{\mathrm{A}} = G_{\mathrm{A}} + k_{\mathrm{B}} T \ln(1 - y^*_{\mathrm{Va}}). \tag{13}$$

When the vacancy concentration is low, Eq. (13) is an expression of Raoult's Law for the solvent in an ideal dilute solution (see, e.g. p. 230 of Ref. [2]), where the vacancies are the 'solute'. Deviations from this ideal behaviour would mean that vacancy–vacancy interactions are important, possibly indicating that divacancies are significant.

Although the detailed discussion here has been about vacancies, the reader should recognize that interstitials can be treated in a precisely analogous way, and the derivation with the addition of an interstitial term to the fundamental equation for G is equally straightforward, although the accurate computation of a formation energy will be trickier because the local atomic displacements are greater and the consequent strain fields tend to be higher in magnitude. In addition, there are more possible local configurations that can be adopted by an interstitial, such as the ⟨111⟩-oriented or

⟨110⟩-oriented dumbells that are exhibited by different metals in the bcc structure, and each of these has n orientational degrees of freedom (e.g. $n = 4$ orientations for the ⟨111⟩-oriented dumbell), which contribute an extra term $k_B \log n$ to the entropy per defect. Because of their high formation energies, however, at least in metallic systems self-interstitials will not play a part in the determination of phase diagrams.

A feature of the above treatment, as mentioned at the start, is that vacancies are not dealt with as if they are a second component. Nevertheless, the problem can be formulated as if they were, and although the reasoning is somewhat different, the result Eq. (11) turns out to be the same, because the 'chemical potential' of the vacancies must be zero. This is the approach used in CALPHAD. In order to achieve their equilibrium concentration, vacancies are unlikely to be spontaneously generated or annihilated within a perfect crystal lattice, which would require simultaneous creation of more energy costly interstitials; this is a process we only expect in radiation damage events. Under normal conditions, they are created or absorbed at sources and sinks, which are generally believed to be located on surfaces, grain boundaries or the cores of edge dislocations. The processes of vacancy emission is thermally activated, as is the process of vacancy diffusion, so the equilibration time will depend on the activation energies but is generally within experimental timescales above room temperature. We shall revisit the problem of vacancy concentration in a unary phase from the CALPHAD perspective after we have discussed the CEF in Section 2.2.

2.2 The compound energy formalism The CEF is a concept used by much of the CALPHAD community for modelling the Gibbs energy of crystalline phases, which may also contain point defects. Only a brief summary of this will be given here, since a full description can be found in the book by Lukas et al. [3]. Another useful reference is the original paper by Hillert [4]. The constituents on each sublattice of a CEF model of a crystalline phase are listed within parenthesis and the number of sites on the sublattice is given as a subscript. This is best illustrated by a simple example, which is a special case of two sublattices and four constituents:

$$(\mathrm{A,\,B})_m(\mathrm{C,\,D})_n, \tag{14}$$

where A and B are the constituents on the first sublattice with m sites and C and D are constituents on the second sublattice with n sites. Notice the use of the term *constituent* rather than *component* here. In the most general case, a constituent can be an element, molecule, ion or vacancy, and the same constituent may exist on several sublattices. The ratios of the numbers of sites, $m, n, \cdots$ in a CEF model are always constant, and for computational purposes one mole comprises $m + n + \ldots$ sites multiplied by Avogadro's number.

The vacancy has the constraint that its chemical potential is always zero. In this simple case, we see there are two sublattices, and the same constituent is not allowed to occupy more than one of them.

The Gibbs energy of the CEF model per mole of formula unit is defined here slightly differently to how you might find it in a chemistry textbook, namely for a fixed number of lattice sites rather than for a fixed number of atoms. In this simple example, it is written as:

$$\begin{aligned} G_\mathrm{m} = {} & y_\mathrm{A}^{(1)} y_\mathrm{C}^{(2)}\, {}^\circ G_\mathrm{A:C} + y_\mathrm{A}^{(1)} y_\mathrm{D}^{(2)}\, {}^\circ G_\mathrm{A:D} \\ & + y_\mathrm{B}^{(1)} y_\mathrm{C}^{(2)}\, {}^\circ G_\mathrm{B:C} + y_\mathrm{B}^{(1)} y_\mathrm{D}^{(2)}\, {}^\circ G_\mathrm{B:D} \\ & + RT\left[m(y_\mathrm{A}^{(1)} \ln(y_\mathrm{A}^{(1)}) + y_\mathrm{B}^{(1)} \ln(y_\mathrm{B}^{(1)})\right. \\ & \left. + \; n(y_\mathrm{C}^{(2)} \ln(y_\mathrm{C}^{(2)}) + y_\mathrm{D}^{(2)} \ln(y_\mathrm{D}^{(2)}))\right] + {}^E G_\mathrm{m} \end{aligned} \tag{15}$$

with

$$1 = y_\mathrm{A}^{(1)} + y_\mathrm{B}^{(1)} \tag{16}$$

and

$$1 = y_\mathrm{C}^{(2)} + y_\mathrm{D}^{(2)}, \tag{17}$$

where $y_i^{(s)}$ is the fraction of constituent i on sublattice s and ${}^\circ G_\mathrm{A:C}$ the Gibbs energy of formation of the compound $\mathrm{A}_m\mathrm{C}_n$, also known as one of the four 'end-members' of the model. Note the use of a colon in the subscript in order to delimit the constituents belonging to different sublattices. The term multiplied with RT is the configurational entropy assuming random mixing in each sublattice and ${}^E G_\mathrm{m}$ is the excess Gibbs energy. If, as in this case, each constituent enters only a single sublattice the superscript on y could be omitted.

For the above (A,B)(C,D) system the excess Gibbs energy is written, omitting the sublattice superscript:

$$\begin{aligned} {}^E G_\mathrm{m} = {} & y_\mathrm{A} y_\mathrm{B} y_\mathrm{C} L_\mathrm{A,B:C} + y_\mathrm{A} y_\mathrm{B} y_\mathrm{D} L_\mathrm{A,B:D} \\ & + y_\mathrm{A} y_\mathrm{C} y_\mathrm{D} L_\mathrm{A:C,D} + y_\mathrm{B} y_\mathrm{C} y_\mathrm{D} L_\mathrm{B:C,D} \\ & + y_\mathrm{A} y_\mathrm{B} y_\mathrm{C} y_\mathrm{D} L_\mathrm{A,B:C,D}. \end{aligned} \tag{18}$$

The $L_\mathrm{A,B:C}$, etc are called interaction parameters and can themselves also depend on the fractions of A, B and C. The $L_\mathrm{A,B:C,D}$ term is called a reciprocal parameter as it depends on two constituents in two different sublattices. The structure of this formula for the Gibbs energy which, given the two constraints in Eqs. (16) and (17), has just two degrees of freedom, is illustrated schematically as a constitutional square in Fig. 1.

It is possible to have more sublattices, constituents and parameters as well as contributions from magnetic transitions as described in Refs. [3, 5].

Note that the Gibbs energy expression is for a fixed number of sites. If the constituents are molecules or a constituent represents a vacant site, the number of real atoms per formula unit can vary. Note that where there is no danger of confusion the same symbols have sometimes been used for quantities per mole and per formula unit. The only difference to the thermodynamic formulae will be whether the gas constant R or Boltzmann's constant k_B appears (see Appendix).

For a CEF model with sublattices, it is straightforward to calculate the chemical potential or partial Gibbs energy of

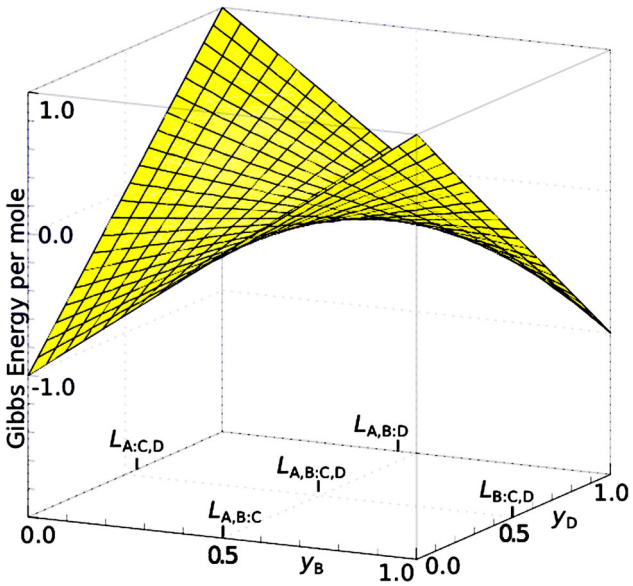

Figure 1 Schematic illustration of the form of the Gibbs energy for the (A,B)(C,D) system described in the text. The Gibbs energy at the four corners, i.e. the end-members energies ${}^{\circ}G_{A:C}$ etc., are the most important. The symbols $L_{A,B:C}$, etc. in Eqs. (15) and (18) are placed where these parameters have most influence on the magnitude of the Gibbs energy.

an end-member, as shown in Appendix A. For example the end-member AC in the model above, has a chemical potential given by:

$$G_{A:C} = G_m + \left(\frac{\partial G_m}{\partial y_A}\right)_{y_{i\neq A}} + \left(\frac{\partial G_m}{\partial y_C}\right)_{y_{i\neq C}} - \sum_i y_i \frac{\partial G_m}{\partial y_i} \quad (19)$$

where the last sum is for all constituents in all sublattices. A derivation of this formula is given for completeness in Appendix A. However, the physically interesting chemical potentials of components cannot be calculated within a CEF model unless the model permits a particular adjustment of the proportion of the end-members corresponding to the addition of a pure component while holding the ratios of all the other components constant, as required by the definition of a partial molar quantity. In the above CEF model, if all of A, B, C and D are components, that is atomic species, it is not possible to add pure A without simultaneously removing B. To do so would violate the constraint of constant number of sites. To have the necessary degree of freedom to add a pure component within the CEF requires the definition of vacancies as a constituent, or the possibility of a constituent occupying more than one sublattice. For example a possible model would be (A,B)(A,Va). Note that differences between chemical potentials, for example the so-called diffusion potentials, may be calculated even if the individual chemical potentials cannot be defined.

For a model with a single set of sites, as treated in the preceding subsection, the above formula for the partial Gibbs energy reduces to

$$G_A = G_m + \frac{\partial G_m}{\partial y_A} - \sum_i y_i \frac{\partial G_m}{\partial y_i}. \quad (20)$$

As already mentioned the vacancy, denoted Va, can in most respects be treated like a real component, together with the additional constraint that the chemical potential of vacancies is always zero, $\mu_{Va} = 0$.

2.3 Thermal vacancies revisited In this section, we revisit from the point of view of the CEF, the case treated in Section 2.1 of a single sublattice including vacancies, in order to explore the connection between the different approaches.

There are always some sites that are vacant in a sublattice, but if the vacancy fraction is very small it may often be ignored. For interstitial sublattices, however, such as the one occupied by C dissolved in iron, vacant sites are the dominating fraction and will be described in Section 2.5. When vacancies are included in all sublattices of a CEF model one must consider that the model includes the possibility of a phase consisting only of vacancies, which is unrealistic, and this requires special precautions to be taken in constructing the model. The representation of vacancies on a single substitutional lattice, shared with components A, B, C, etc., is written (A, B, C, ... Va) which gives a Gibbs energy expression according to the CEF:

$$G_m = \sum_{i\neq Va} y_i \, {}^{\circ}G_i + y_{Va} \, {}^{\circ}G_{Va} + RT\left(\sum_{i\neq Va} y_i \ln(y_i) + y_{Va}\ln(y_{Va})\right) + {}^{E}G_m, \quad (21)$$

where

$${}^{E}G_m = y_{Va}\sum_{i\neq Va} y_i L_{i,Va} \quad (22)$$

with the constraint

$$1 = \sum_{i\neq Va} y_i + y_{Va}. \quad (23)$$

The mole fractions of the real atoms are given by

$$x_A = \frac{y_A}{\sum_{i\neq Va} y_i} = \frac{y_A}{1 - y_{Va}}. \quad (24)$$

We can apply the equation equivalent to Eq. (20), expressed for vacancies, namely:

$$G_{Va} = G_m + \frac{\partial G_m}{\partial y_{Va}} - \sum_i y_i \frac{\partial G_m}{\partial y_i}. \quad (25)$$

Inserting Eq. (21) in Eq. (25) gives

$$G_{\mathrm{Va}} = {}^{\circ}G_{\mathrm{Va}} + RT\ln(y_{\mathrm{Va}}) + (1 - y_{\mathrm{Va}})\sum_{i\neq \mathrm{Va}} y_i L_{i,\mathrm{Va}} \qquad (26)$$

in which we have included the simplest possible excess term. At equilibrium G_{Va} is the chemical potential of vacancies, which must be zero:

$$G^{*}_{\mathrm{Va}} = \mu_{\mathrm{Va}} = 0 \qquad (27)$$

assuming there are grain boundaries and free surfaces or other sources and sinks for vacancies. For a sufficiently small fraction of vacancies, $y_{\mathrm{Va}} \ll 1$, we make the approximation $1 - y_{\mathrm{Va}} \approx 1$ and find that the fraction of vacancies on a lattice with a number of different components is:

$$y^{*}_{\mathrm{Va}} = \exp\left(-\frac{{}^{\circ}G_{\mathrm{Va}} + \sum_{i\neq \mathrm{Va}} y_i L_{A,\mathrm{Va}}}{RT}\right). \qquad (28)$$

Let us now compare the vacancy formation energy G^{f}_{Va}, introduced in Eq. (7) with the CEF energies for a phase with a single real component A. Equating Eqs. (28) and (10) we have in the dilute limit:

$$G^{f}_{\mathrm{Va}} = {}^{\circ}G_{\mathrm{Va}} + L_{\mathrm{A,Va}}. \qquad (29)$$

Thus the CEF separates the energy to create a vacancy into a *structure-dependent* ${}^{\circ}G_{\mathrm{Va}}$ vacancy formation energy and a *structure- and constituent-dependent* energy $L_{\mathrm{A,Va}}$. Note that this interaction energy does not represent a vacancy–vacancy interaction, it is part of the energy to create a single vacancy for a specific phase with a single component A. Formally, it would seem to be arbitrary how the physical vacancy formation energy G^{f}_{Va} is partitioned between ${}^{\circ}G_{\mathrm{Va}}$ and $L_{\mathrm{A,Va}}$. However, this freedom allows us to make a practically useful partition of the vacancy formation energy between these two terms, which is important for the modelling of point defects in CALPHAD.

The structure-dependent energy within the context of the CEF provides an effective means to limit the maximum vacancy fraction found in the stable phase, i.e. a kind of 'penalty function'. If we assume that the maximum fraction of thermal vacancies is 0.1 before the crystal structure breaks down this gives:

$${}^{\circ}G_{\mathrm{Va}} = -RT\ln(y^{\max}_{\mathrm{Va}}) \approx 2.3RT. \qquad (30)$$

This end-member energy is independent of the actual atoms occupying the lattice, it is only related to the structure. As ${}^{\circ}G_{\mathrm{Va}}$ is assigned a fixed value for each phase, $L_{\mathrm{A,Va}}$ can be adjusted to experimental or theoretical data. Equation (28) is the simplest method to extend the equation for vacancy formation for a pure element to a multicomponent system, and it behaves well also outside the dilute range. For such cases, the CEF provides additional parameters that can be adjusted to the available data. The diagram in Fig. 2 shows the vacancy fraction of an element with the vacancy fraction equal to 0.001 at its melting temperature 1000 K. The actual fraction of vacancies in the dilute range can be fitted using the $L_{i,\mathrm{Va}}$ parameter for any real component i. If the vacancy fraction at the melting temperature T_{m} is 0.001, much lower than that causing breakdown of the crystal structure, we get

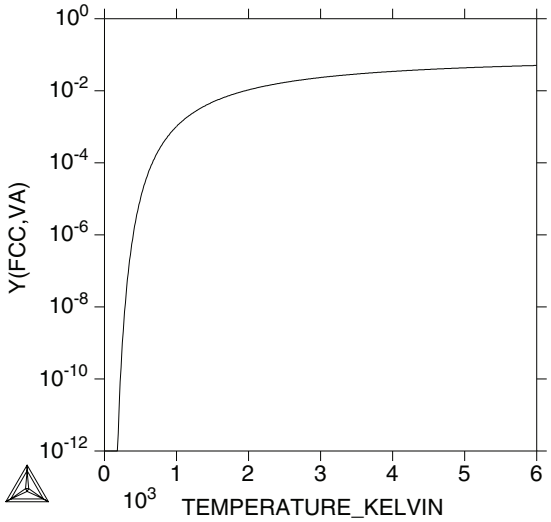

Figure 2 The variation of vacancy fractions in a phase calculated from Eq. (28) with ${}^{\circ}G_{\mathrm{Va}} = 2.3RT$ and $L_{\mathrm{A,Va}} = 4600R$ from Eq. (32) for $T_{\mathrm{m}} = 1000$ K. The vacancy fraction is 0.001 at 1000 K and it does not exceed 0.1 even at 6000 K.

$$\exp\left(-2.3 - \frac{L_{\mathrm{A,Va}}}{RT_{\mathrm{m}}}\right) = 0.001, \qquad (31)$$

$$L_{\mathrm{A,Va}} = 4.6RT_{\mathrm{m}}.$$

If the experimental vacancy fractions are higher we can in principle use higher order interaction parameters to fit them. Of course, if we have determined a physical Gibbs energy of vacancy formation as G^{f}_{Va} we can make use of it by setting

$$L_{\mathrm{A,Va}} = G^{f}_{\mathrm{Va}} - 2.3RT\,. \qquad (32)$$

2.4 Ordered alloys: Example of NiAl In the Al–Ni system, there are several ordering phenomena, the technologically most important is the formation of the Ni_3Al ordering, structure type $L1_2$, of the disordered Ni fcc phase (A1 structure), see the phase diagram in Fig. 3.

However, in this case we will study the NiAl phase in the middle of the phase diagram which is a B2 ordered structure of the bcc phase (A2 structure). In the disordered bcc phase the fractions of all components on all sites are equal but in the ordered structure the atoms prefer to be surrounded by atoms of another kind, i.e. if the central atom is Al the corner atoms will be mainly Ni (or vice versa because the sites are equivalent). The order is never perfect, there is a significant

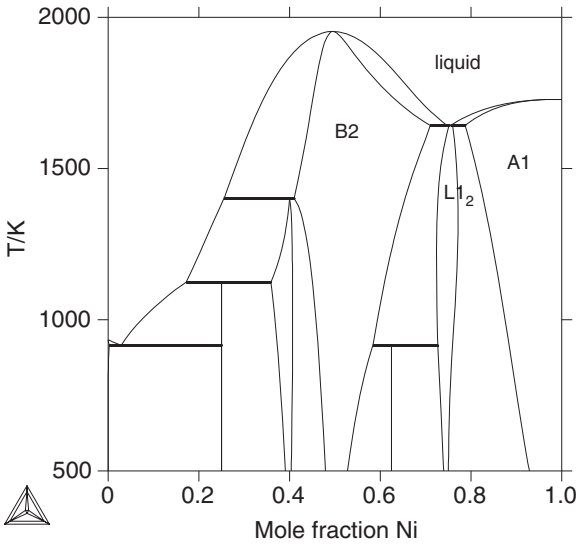

Figure 3 The Al–Ni phase diagram phase

fraction of antisite atoms on the sublattices, which explains the large composition range.

We start by discussing what might be called the 'defect-centric' approach, as outlined in Section 2.1, in which we are not concerned with transferability of parameters to other phases, but merely with the concentrations of defects in the dilute limit. This system, which has been extensively studied both theoretically and experimentally, can be treated in an analogous fashion to the single component system by writing down an expression for the Gibbs energy and minimizing it to find the equilibrium concentrations of point defects. In the *canonical* formalism, which we discuss first [6–11], the chemical potential of each component is obtained as an output of a Gibbs energy minimization, which takes the composition as a fixed input [7]. For comparison, we describe later also the *grand canonical* formalism, which takes the chemical potentials as a fixed input and must therefore generate the composition as output [9, 12, 13]. If applied consistently, both approaches must give identical results for Gibbs energies and corresponding equilibrium concentrations of point defects. In an ordered alloy, point defects arise by thermal excitation, as in elemental metals. However, they also appear in order to accommodate deviations from perfect stoichiometry, in which case they are referred to as *constitutional defects* and have a fixed limiting concentration as the temperature is reduced. An interesting feature of NiAl, which has been extensively studied and is well understood, is that the constitutional defect in a Ni-rich alloy is the antisite defect (Ni on an Al site), whereas in an Al-rich alloy it is the Ni vacancy.

The simplest model we discuss in detail is again an ideal solution model, but now we have to define two sublattices, which we label *a* and *b*. We use these labels for convenience to identify them with the majority species on the sublattice, thus A is the majority species on sublattice *a*, and B on *b*. In a stoichiometric, defect-free alloy, all the A atoms would occupy sites of the *a* sublattice and all the B atoms would occupy sites of the *b* sublattice. The method described here is easily extended to alloys of arbitrary stoichiometric ratio, e.g. A_mB_n, as shown by Hagen and Finnis [7]. It is also formally trivial to add a regular solution term to account for defect–defect interactions. However, we do not include that aspect here, since its omission enables us to obtain some helpful analytical results to illustrate the theoretical framework.

We allow the possibility of four species of point defects: n^b_{A} antisites (wrong occupancy) on the *b* sublattice, n^a_{B} antisites on the *a* sublattice, n^a_{Va} vacancies on the *a* sublattice and n^b_{Va} vacancies on the *b* sublattice. There are therefore six species altogether, with numbers n^a_i and n^b_i, where i stands for A, B or Va. Each sublattice has N sites where

$$N = n^a_{\mathrm{A}} + n^a_{\mathrm{B}} + n^a_{\mathrm{Va}} \tag{33}$$

$$= n^b_{\mathrm{A}} + n^b_{\mathrm{B}} + n^b_{\mathrm{Va}}, \tag{34}$$

while conservation of atoms requires that

$$n_{\mathrm{A}} = n^a_{\mathrm{A}} + n^b_{\mathrm{A}} \tag{35}$$

and

$$n_{\mathrm{B}} = n^a_{\mathrm{B}} + n^b_{\mathrm{B}}. \tag{36}$$

By analogy with Eq. (1) the Gibbs energy can be written as

$$\begin{aligned} G = N \{ & y^a_{\mathrm{A}} G^a_{\mathrm{A}} + y^a_{\mathrm{B}} G^a_{\mathrm{B}} + y^a_{\mathrm{Va}} G^a_{\mathrm{Va}} \\ & + y^b_{\mathrm{B}} G^b_{\mathrm{B}} + y^b_{\mathrm{A}} G^b_{\mathrm{A}} + y^b_{\mathrm{Va}} G^b_{\mathrm{Va}} \\ & - T[S_{\mathrm{mix}}(\{y^a_i\}) + S_{\mathrm{mix}}(\{y^b_i\})] \} \end{aligned} \tag{37}$$

and the fractional site occupancies are defined by

$$y^a_i = \frac{n^a_i}{N}, \qquad y^b_i = \frac{n^b_i}{N}. \tag{38}$$

2.4.1 Calculation of parameters for the defect-centric B2 NiAl model We first describe here the approach to calculating the relevant defect energies, which are required for all the methodologies for calculating compositions and chemical potentials. It resembles the method described above for the pure element. However, unlike the pure elements, the individual values of G^a_{A} and G^b_{B} are indeterminate by a direct calculation. Fortunately, as it turns out, we never need to know their separate values but only their sum, which is the Gibbs energy per formula unit of the hypothetical defect-free AB compound and which *can* be calculated. This Gibbs energy is a reference quantity and its value should not effect the calculated equilibrium defect concentrations for a specified stoichiometry. Corresponding to each defect, we define a *raw* Gibbs energy of formation, which we shall denote by the lower case g. This is calculated simply as the difference when the Gibbs energy of a perfect crystal or supercell of N sites per sublattices (staying with our simple example of equal sublattices) is subtracted from a crystal of the same number of sites containing the defect. These raw defect energies are related to the formation energies previously found by writing

down explicitly the energies of the N-site crystals with and without a defect. Thus:

$$g^a_{\mathrm{Va}} = [G^a_{\mathrm{Va}} + (N-1)G^a_{\mathrm{A}} + NG^b_{\mathrm{B}}] - N[G^a_{\mathrm{A}} + G^b_{\mathrm{B}}] = G^a_{\mathrm{Va}} - G^a_{\mathrm{A}}, \tag{39}$$

$$g^b_{\mathrm{Va}} = [G^b_{\mathrm{Va}} + (N-1)G^b_{\mathrm{B}} + NG^a_{\mathrm{A}}] - N[G^a_{\mathrm{A}} + G^b_{\mathrm{B}}] = G^b_{\mathrm{Va}} - G^b_{\mathrm{B}}, \tag{40}$$

$$g^a_{\mathrm{B}} = [G^a_{\mathrm{B}} + (N-1)G^a_{\mathrm{A}} + NG^b_{\mathrm{B}}] - N[G^a_{\mathrm{A}} + G^b_{\mathrm{B}}] = G^a_{\mathrm{B}} - G^a_{\mathrm{A}}, \tag{41}$$

$$g^b_{\mathrm{A}} = [G^b_{\mathrm{A}} + (N-1)G^b_{\mathrm{B}} + NG^a_{\mathrm{A}}] - N[G^a_{\mathrm{A}} + G^b_{\mathrm{B}}] = G^b_{\mathrm{A}} - G^b_{\mathrm{B}}. \tag{42}$$

The important points to note about these relationships are that the quantities on the left hand side are the ones that are calculated and form the input to a model, with which the actual point defect concentrations and the Gibbs energy will be derived by a numerical, or in the present case analytical, minimization. The individual quantities G^a_{A} and G^b_{B} will never be required, but their sum ${}^{\circ}G_{\mathrm{A:B}}$, which is the Gibbs energy per mole of a hypothetical defect free crystal, is essential. This is easily seen if we express Eq. (37) in terms of the raw defect Gibbs energies and ${}^{\circ}G_{\mathrm{A:B}}$:

$$G = N\left\{{}^{\circ}G_{\mathrm{A:B}} + y^a_{\mathrm{B}}g^a_{\mathrm{B}} + y^a_{\mathrm{Va}}g^a_{\mathrm{Va}} + y^b_{\mathrm{A}}g^b_{\mathrm{A}} + y^b_{\mathrm{Va}}g^b_{\mathrm{Va}} - T[S_{\mathrm{mix}}(\{y^a_i\}) + S_{\mathrm{mix}}(\{y^b_i\})]\right\}. \tag{43}$$

All the calculated or reference Gibbs energies should be at the same pressure, which for practical purposes can usually be taken as zero.

2.4.2 The defect-centric B2 NiAl model: Canonical formalism Now let us summarize the simple mathematical problem in hand. Given the composition is fixed, we have seven unknown quantities, the concentrations per sublattice site of each of the species and the number of sites N on a sublattice. These will be determined by setting to zero the corresponding seven derivatives of G while respecting the four constraints that must be satisfied, namely Eqs. (33–36), which it is convenient to write in terms of the site occupancies thus:

$$1 = \sum_i y^a_i, \tag{44}$$

$$1 = \sum_i y^b_i, \tag{45}$$

$$n_{\mathrm{A}} = (y^a_{\mathrm{A}} + y^b_{\mathrm{A}})N, \tag{46}$$

$$n_{\mathrm{B}} = (y^a_{\mathrm{B}} + y^b_{\mathrm{B}})N. \tag{47}$$

The alloy composition is measured by the mole fraction x of component A, e.g. $\mathrm{Ni}_x\mathrm{Al}_{1-x}$. Thus it will be convenient to combine the constraint Eqs. (46) and (47) to eliminate N and the actual numbers of atoms in favour of x:

$$\frac{x}{1-x} = \frac{y^a_{\mathrm{A}} + y^b_{\mathrm{A}}}{y^a_{\mathrm{B}} + y^b_{\mathrm{B}}}. \tag{48}$$

Each constraint introduces an unknown Lagrange multiplier into the minimization procedure, which we designate λ_a, λ_b, λ_{A} and λ_{B}, respectively. We therefore end up with ten unknowns, but with ten equations to solve for them. Alternatively, the constraints and variables could be reduced in number by simple substitution, but the systematic procedure described here has the virtue of being straightforward to generalize or automate. The seven minimization conditions are as follows:

$$\frac{\partial G}{\partial y^a_{\mathrm{A}}} - \lambda_a - N\lambda_{\mathrm{A}} = 0, \tag{49}$$

$$\frac{\partial G}{\partial y^b_{\mathrm{B}}} - \lambda_b - N\lambda_{\mathrm{B}} = 0, \tag{50}$$

$$\frac{\partial G}{\partial y^a_{\mathrm{B}}} - \lambda_a - N\lambda_{\mathrm{B}} = 0, \tag{51}$$

$$\frac{\partial G}{\partial y^b_{\mathrm{A}}} - \lambda_b - N\lambda_{\mathrm{A}} = 0, \tag{52}$$

$$\frac{\partial G}{\partial y^a_{\mathrm{Va}}} - \lambda_a = 0, \tag{53}$$

$$\frac{\partial G}{\partial y^b_{\mathrm{Va}}} - \lambda_b = 0, \tag{54}$$

$$G - \lambda_{\mathrm{A}}n_{\mathrm{A}} - \lambda_{\mathrm{B}}n_{\mathrm{B}} = 0. \tag{55}$$

Equation (55) has been written in its present form by substituting the constraint Eqs. (46) and (47). Comparing it to the standard Euler relation for G in terms of chemical potentials:

$$G = \mu_{\mathrm{A}}n_{\mathrm{A}} + \mu_{\mathrm{B}}n_{\mathrm{B}}, \tag{56}$$

we see that the Lagrange multipliers λ_{A} and λ_{B} are none other than the chemical potentials of the components.

After evaluating the derivatives of G and eliminating the Lagrange multipliers, we obtain the following simultaneous equations for the concentrations in the form of 'laws of mass action':

$$y^a_{\mathrm{Va}}y^b_{\mathrm{Va}} = k_v(T), \tag{57a}$$

$$\frac{y^b_{\mathrm{A}}y^a_{\mathrm{B}}}{y^a_{\mathrm{A}}y^b_{\mathrm{B}}} = k_s(T), \tag{57b}$$

$$\frac{y^a_{\mathrm{B}}(y^b_{\mathrm{Va}})^2}{y^b_{\mathrm{B}}} = k_t(T). \tag{57c}$$

where the equilibrium constants are given by

$$k_v(T) = \exp\left(-\frac{G^a_{\mathrm{Va}} + G^b_{\mathrm{Va}}}{k_{\mathrm{B}}T}\right) = \exp\left(-\frac{g^a_{\mathrm{Va}} + g^b_{\mathrm{Va}} + {}^{\circ}G_{\mathrm{A:B}}}{k_{\mathrm{B}}T}\right), \quad (58a)$$

$$k_s(T) = \exp\left(-\frac{G^a_{\mathrm{B}} + G^b_{\mathrm{A}} - G^a_{\mathrm{A}} - G^b_{\mathrm{B}}}{k_{\mathrm{B}}T}\right) = \exp\left(-\frac{g^a_{\mathrm{B}} + g^b_{\mathrm{A}}}{k_{\mathrm{B}}T}\right), \quad (58b)$$

$$k_t(T) = \exp\left(-\frac{2G^b_{\mathrm{Va}} + G^a_{\mathrm{B}} - G^b_{\mathrm{B}}}{k_{\mathrm{B}}T}\right) = \exp\left(-\frac{2g^b_{\mathrm{Va}} + g^a_{\mathrm{B}} + {}^{\circ}G_{\mathrm{A:B}}}{k_{\mathrm{B}}T}\right). \quad (58c)$$

These equilibrium constants, which are amenable to calculation, have been given the subscripts v, s and t because they correspond to three particular reactions that could occur to create multiple defects, although these are not associated as clusters, namely:

v – formation of a vacancy on each sublattice
s – formation of an antisite on each sublattice
t – formation of two vacancies on b and an antisite on a, called a 'triple defect'.

The special feature of these combinations is that they conserve the atoms, that is they can be realized physically without any addition or subtraction of atoms. Thus, they represent three degrees of freedom corresponding to *thermal* defects, which can be excited by temperature at any given stoichiometry of the crystal. Any other possible combination of defects that one can imagine creating subject to this constraint is expressible as a linear combination of these three. Knowledge of the raw defect energies is also sufficient to tell us immediately what the constitutional defect should be for the two different cases of non-stoichiometry, $x > 0.5$ (Ni-rich) and $x < 0.5$ (Al-rich).

Finally, we can eliminate five of the six concentrations from the set of Eqs. (44), (45), (48) and (57) to obtain an equation for one of them, e.g. for y^a_{Va}:

$$k_t^2(y^a_{\mathrm{Va}})^4 + (k_v - k_t)k_t k_v(y^a_{\mathrm{Va}})^3 + (k_v - k_t)k_s k_v^3 y^a_{\mathrm{Va}} - k_s k_v^4 + (1 - k_s)k_t k_v^2[-y k_v y^a_{\mathrm{Va}} + (2y - 1)(y^a_{\mathrm{Va}})^2 - y(y^a_{\mathrm{Va}})^3] = 0. \quad (59)$$

Having solved this quartic equation numerically for y^a_{Va}, the other concentrations can be obtained by back substitution. A natural extension to these ideal solution approaches would be to include a regular solution term to describe the interactions, but in practice interactions are handled by the CEF.

2.4.3 The grand canonical formalism In Section 2.4.2, the equilibrium properties were calculated by keeping the number of atoms of each species fixed and minimizing the Gibbs energy. Within the same conceptual framework we can keep the number of lattice sites and the chemical potentials of the components, μ_{A} and μ_{B}, fixed. Then we must minimize the *grand potential*, defined as

$$\mathcal{G} = G - PV - \mu_{\mathrm{A}} n_{\mathrm{A}} - \mu_{\mathrm{B}} n_{\mathrm{B}}, \quad (60)$$

for $P = 0$ with respect to n_{A}, n_{B} and the defect concentrations. In equilibrium, this grand potential must vanish, to satisfy the familiar Euler Eq. (56).

Since in this approach N is defined to be constant, the task is to minimize $\mathcal{G}$ with respect to the six site concentrations y_i at fixed μ_{A} and μ_{B}, subject of course to the two constraints in Eqs. (44) and (45). These constraints are simply included by eliminating the numbers of A and B atoms, leaving only the four defect concentrations as independent variables. Hence by setting the derivatives to zero, it is easily shown that the defect concentrations in this ideal solution model are given by

$$\frac{y^a_{\mathrm{Va}}}{y^a_{\mathrm{A}}} = \exp\left(-\frac{g^a_{\mathrm{Va}} + \mu_{\mathrm{A}}}{RT}\right), \quad (61)$$

$$\frac{y^b_{\mathrm{Va}}}{y^b_{\mathrm{B}}} = \exp\left(-\frac{g^b_{\mathrm{Va}} + \mu_{\mathrm{B}}}{RT}\right), \quad (62)$$

$$\frac{y^a_{\mathrm{B}}}{y^a_{\mathrm{A}}} = \exp\left(-\frac{g^a_{\mathrm{B}} + \mu_{\mathrm{A}} - \mu_{\mathrm{B}}}{RT}\right), \quad (63)$$

$$\frac{y^b_{\mathrm{A}}}{y^b_{\mathrm{B}}} = \exp\left(-\frac{g^b_{\mathrm{A}} + \mu_{\mathrm{B}} - \mu_{\mathrm{A}}}{RT}\right). \quad (64)$$

It is often convenient to work directly with these equations, displaying chemical potentials explicitly as external control variables, rather than deriving them from the canonical formalism, from which they are implicitly dependent on the composition; the results must be the same. To convince oneself that this is the case one might use Eq. (56) and the above four equations for defect concentrations to eliminate the chemical potentials and recover the previous formulae Eqs. (57) and (58). In either case the solution for G as a function of composition will entail a very small numerical effort. The practical application of the grand canonical approach, described in the following section, is even simpler, since for low defect concentrations G in Eq. (56) can be accurately approximated by ${}^{\circ}G_{\mathrm{A:B}}$. Also, in principle, defect interactions might be included in a regular or sub-regular solution formalism. The main computational effort is the calculation of the point defect energies themselves, which in the case of NiAl have been obtained by several groups using standard DFT methods. In view of its convenience and wide use, we discuss the grand canonical approach in more detail in the following sub-section.

2.4.4 *Ab initio* defect chemistry and the grand canonical formalism The grand canonical approach was

originally developed in the *ab initio* defect chemistry community [14], where it is widely used. In this section, we give a simple derivation of the formulae for defect concentrations, which leads, somewhat more intuitively, to the same equations as above in the usual limit of small defect concentrations. The difference in the present derivation is that the host chemical potentials are assumed at the outset to be unchanged by the presence of the defects.

The key quantity within this approach is a modified defect formation energy defined in terms of a raw defect energy by the addition of a term referring to chemical potentials, thus:

$$\mathcal{G}_\delta = g_\delta - \sum_k \Delta n_k \mu_k \tag{65}$$

with δ labelling one of the four types of defect, i.e. $g_\delta \equiv \{g^a_{\mathrm{Va}}, g^a_{\mathrm{B}}, g^b_{\mathrm{Va}}, g^b_{\mathrm{A}}\}$. The index k runs over the chemical components ($k \equiv \{\mathrm{A}, \mathrm{B}\}$), Δn_k is the number of atoms, which have to be added to create the defect δ, it is negative if an atom is removed from the crystal and positive if one is added. The chemical potential μ_k refers to atom k in its external chemical reservoir, with which the system is assumed to be in thermodynamic equilibrium, i.e. the reservoir from which the atoms to create the defect are brought to or taken from. The raw defect energies are defined in Eq. (39), and evaluated as the *ab initio* computed difference between the Gibbs energy of a bulk cell containing one localized defect and the same cell but without the defect. If electronic, vibrational or magnetic contributions are neglected $G = E_{\mathrm{tot}}$, i.e. the total energy as obtained from the electronic structure calculations, normally referring to zero pressure.

For the ordered AB alloy considered here, exemplified by B2 NiAl, these definitions become specifically:

$$\begin{aligned}
\mathcal{G}^{\mathrm{a}}_{\mathrm{Va}} &= g^a_{\mathrm{Va}} + \mu_{\mathrm{A}},\\
\mathcal{G}^{\mathrm{b}}_{\mathrm{Va}} &= g^b_{\mathrm{Va}} + \mu_{\mathrm{B}},\\
\mathcal{G}^{\mathrm{a}}_{\mathrm{B}} &= g^a_{\mathrm{B}} + \mu_{\mathrm{A}} - \mu_{\mathrm{B}},\\
\mathcal{G}^{\mathrm{b}}_{\mathrm{A}} &= g^b_{\mathrm{A}} + \mu_{\mathrm{B}} - \mu_{\mathrm{A}}.
\end{aligned} \tag{66}$$

The above equations are physically intuitive. Creating a defect can be separated into two steps: (i) removing/adding atoms from the ideal bulk, which is described by the first term in Eq. (65) and (ii) placing/removing these atoms in their respective chemical reservoir μ_k, which is described by the second term in Eq. (65). In contrast to a unary system, where the chemical potential is a constant, in binaries and multicomponent systems they can be varied within certain bounds. A generic upper limit for any chemical potential is that it must be smaller than or equal to the value it takes in its bulk unary phase, i.e. $\mu_{\mathrm{A}} \leq \mu_{\mathrm{A}}(\mathrm{bulk})$ with $\mu_{\mathrm{A}}(\mathrm{bulk})$ the chemical potential of the thermodynamically most favourable unary phase of element A. Otherwise, the system would be unstable against decomposing into this phase. Further limits may be imposed if other stoichiometries of this alloy form thermodynamically stable phases. For example to prevent that the AB system becomes unstable against forming an $\mathrm{A}_m\mathrm{B}_n$ phase, we have to impose the following inequality:

$$m\mu_{\mathrm{A}} + n\mu_{\mathrm{B}} \leq \mu_{\mathrm{A}_m\mathrm{B}_n}. \tag{67}$$

In order to ensure thermodynamic equilibrium with the AB bulk phase, as long as the defects are in sufficiently low concentrations we can approximate:

$$\mu_{\mathrm{A}} + \mu_{\mathrm{B}} = {}^{\circ}G_{\mathrm{A:B}}. \tag{68}$$

Note this is not quite an exact equality because the Gibbs energy of the compound with defects is slightly modified by their presence. Thus in $\mathrm{A}_{1+\delta}\mathrm{B}_{1-\delta}$ we would have strictly

$$(1+\delta)\mu_{\mathrm{A}} + (1-\delta)\mu_{\mathrm{B}} = \text{Gibbs energy of } \mathrm{A}_{1+\delta}\mathrm{B}_{1-\delta}. \tag{69}$$

Since μ_{A} and μ_{B} cannot be varied independently, in the following we choose μ_{A} as independent variable, μ_{B} is then simply:

$$\mu_{\mathrm{B}} = {}^{\circ}G_{\mathrm{A:B}} - \mu_{\mathrm{A}}. \tag{70}$$

Using Eq. (65) and referring to Table 1 for the number of atoms that need to be added/removed to create the defect we get the effective formation energy for all defects considered here. To be more specific, let us consider a vacancy and an antisite defect. For the vacancy on the *a* sublattice site (which may be denoted Va_A or Va_a) the corresponding formation energy reads:

$$\mathcal{G}^{\mathrm{a}}_{\mathrm{Va}} = g^a_{\mathrm{Va}} + \mu_{\mathrm{A}}. \tag{71}$$

For substituting an A atom on a b sublattice site (which may be denoted A_{B} or A_b) the corresponding energy is:

$$\mathcal{G}^{\mathrm{b}}_{\mathrm{A}} = g^b_{\mathrm{A}} - 2\mu_A + {}^{\circ}G_{\mathrm{A:B}}. \tag{72}$$

Since the four raw defect formation energies g_δ and ${}^{\circ}G_{\mathrm{A:B}}$ are directly accessible from total energy calculations all defect energies are expressed by a linear equation in a single free parameter μ_{A}.

Table 1 Number of A and B atoms to be added (positive sign) or to be removed to create the two possible vacancy defects, namely vacancy on an *a* or *b* site, or the two analogous antisite defects.

defect	Δn_{A}	Δn_{B}
Va_a	-1	0
Va_b	0	-1
A_b	$+1$	-1
B_a	-1	$+1$

Using these formation energies, the equilibrium concentrations of the defects can be directly computed:

$$y_\delta^*(\mu_A) = \exp\left(-\frac{\mathcal{G}_\delta(\mu_A)}{k_B T}\right). \tag{73}$$

As we would expect, this result is almost identical to Eq. (61) derived above, with the small difference that by not making the approximation of Eq. (68), the concentrations in Eq. (61) include a denominator that is not quite unity, but tends to 1 in the limit of low defect concentrations. In practice, this will not be important within the range of validity of the defect-centric model. From the defect concentrations, it is a straightforward matter to obtain the equilibrium stoichiometry as a function of the chemical potentials.

In comparing the grand canonical approach described here with the canonical or the CEF described above, we make the following remarks. First, as in the CEF most equations given here do not have to be solved by hand, their solution is implemented in the computer. In practice, all one needs to apply this formalism are the raw defect formation energies g_α, which can be directly computed from standard DFT total energy codes (Eq. 66). The calculation of a raw defect energy also entails a calculation of the Gibbs energy per mole of the ideal phase ${}^\circ G_{A:B}$. Thus, this approach is conceptually straightforward, and avoids the need to explicitly formulate and solve conditions with respect to particle conservation as in the canonical approach. Thus, for small concentrations of point defects in binary and ternary systems the grand canonical approach is the method of choice. For multicomponent systems, with larger defect concentrations, and to connect to the CALPHAD community and their methods, however, the CEF has its own distinct advantages, which are outlined below.

2.4.5 Compatibility of defect-centric and a CEF models There are clearly problems if we simply extend the parameter space of the previous, defect-centric B2 model to defect concentrations higher than 1–5% (the exact range of validity has not been investigated), even if we proceed by including polynomials in the defect concentrations to represent their interactions. An obvious problem is that if we had 100% antisites on each sublattice we should recover the energy of the perfect compound, ${}^\circ G_{A:B}$, which is certainly not the case in the defect-centric model. A CEF model, with more parameters at our disposal, should be able to address this difficulty, and as well as providing a means of extrapolation to a wider range of concentration, it should provide compatibility with the parameters used to model other phases of the Al–Ni system.

Following the usual notation, the CEF model that represents the same system as the defect-centric model described above must be written: $(A,B,Va)_1(A,B,Va)_1$. This model has a 3×3 matrix of end-member energies and concentrations, and for the case in hand A and B stand for Ni and Al, respectively. By analogy with Eq. (15) a CEF expression for the Gibbs energy per mole is formulated as

$$\begin{aligned} G_m = {} & y_A^a y_A^b {}^\circ G_{A:A} + y_A^a y_B^b {}^\circ G_{A:B} + y_A^a y_{Va}^b {}^\circ G_{A:Va} \\ & + y_B^a y_A^b {}^\circ G_{B:A} + y_B^a y_B^b {}^\circ G_{B:B} + y_B^a y_{Va}^b {}^\circ G_{B:Va} \\ & + y_{Va}^a y_A^b {}^\circ G_{Va:A} + y_{Va}^a y_B^b {}^\circ G_{Va:B} + y_{Va}^a y_{Va}^b {}^\circ G_{Va:Va} \\ & + RT\left(y_A^a \ln y_A^a + y_B^a \ln y_B^a + y_{Va}^a \ln y_{Va}^a \right. \\ & \left. + \; y_A^b \ln y_A^b + y_B^b \ln y_B^b + y_{Va}^b \ln y_{Va}^b\right) + {}^E G_m \end{aligned} \tag{74}$$

with

$$1 = y_A^a + y_B^a + y_{Va}^a \tag{75}$$

and

$$1 = y_A^b + y_B^b + y_{Va}^b. \tag{76}$$

${}^\circ G_{A:A}$ and ${}^\circ G_{B:B}$ represent the Gibbs energy for two moles of atoms of pure A and B, respectively, in their metastable bcc (A2) structure and ${}^\circ G_{A:B}={}^\circ G_{B:A}$ is the Gibbs energy of one mole of the stable B2 phase. The end-members ${}^\circ G_{A:Va} = {}^\circ G_{Va:A}$ must be related to the energy of forming a vacancy on one sublattice in the B2 structure when the other sublattice is filled with A and ${}^\circ G_{B:Va} = {}^\circ G_{Va:B}$ the same when one sublattice is filled with B. Finally the ${}^\circ G_{Va:Va}$ end-member is used as the 'penalty energy' to prevent forming a phase consisting of just vacancies, as derived in Eq. (30). In many cases, the metastable end-members are even mechanically unstable but for the modelling a value must be assigned that ensures it will never appear as a stable phase and, possibly together with an interaction parameter, gives a reasonable behaviour for small defect fractions.

The nine end-member energies might be expressed formally in terms of the six energies entering the defect-centric formula Eq. (37) as follows:

$${}^\circ G_{A:A} = G_A^a + G_A^b = g_A^b + {}^\circ G_{A:B}, \tag{77a}$$

$${}^\circ G_{A:B} = G_A^a + G_B^b = {}^\circ G_{A:B}, \tag{77b}$$

$${}^\circ G_{A:Va} = G_A^a + G_{Va}^b = g_{Va}^b + {}^\circ G_{A:B}, \tag{77c}$$

$${}^\circ G_{B:A} = G_B^a + G_A^b = g_B^a + g_A^b + {}^\circ G_{A:B}, \tag{77d}$$

$${}^\circ G_{B:B} = G_B^a + G_B^b = g_B^a + {}^\circ G_{A:B}, \tag{77e}$$

$${}^\circ G_{B:Va} = G_B^a + G_{Va}^b = g_B^a + g_{Va}^b + {}^\circ G_{A:B}, \tag{77f}$$

$${}^\circ G_{Va:A} = G_{Va}^a + G_A^b = g_{Va}^a + g_A^b + {}^\circ G_{A:B}, \tag{77g}$$

$${}^\circ G_{Va:B} = G_{Va}^a + G_B^b = g_{Va}^a + {}^\circ G_{A:B}, \tag{77h}$$

$${}^\circ G_{Va:Va} = G_{Va}^a + G_{Va}^b = g_{Va}^a + g_{Va}^b + {}^\circ G_{A:B}. \tag{77i}$$

We include on the right hand side the expressions in terms of the quantities that can be directly calculated, namely the raw defect energies and the perfect lattice energy. It is indeed encouraging that if we substitute Eq. (77) into Eq. (74) we recover the defect-centric formula Eq. (37), apart from an excess term, which we neglect in Eq. (37). However, this would *not* be the correct way to determine the parameters of the CEF model for this system! The above comparison of the

defect-centric formalism and CEF in this example illustrates an important aspect of the modelling of point defects. The symmetries of the end-member energies mentioned above are not respected by the physical energies to which they appear formally equivalent. For example the CEF end-members in Eqs. (77b) and (77d) must have the same energy, but their counterparts in the defect-centric model have different values; the first is the energy of the perfect lattice, while the second is the energy of the perfect lattice plus the formation energy of one pair of antisites. If we create more and more pairs of antisites, until all the atoms have been switched between the *a* and the *b* sublattices, then we know we must recover the original energy of the perfect lattice. Thus in order for this symmetry to be preserved in energies of the CEF end-members, some of the most important physical information about the point defect formation energies is modelled within the excess term ${}^{E}G_{\mathrm{m}}$, as we have already seen for the case of vacancy formation energies in a unary phase.

To describe the details of the experimental and theoretical information about the B2 phase in this system, and the equilibria with the other phases, several additional excess parameters have been assessed by Dupin [15]. We present here some of the results.

The resulting variation of the constituent fractions at 1500 K are shown in Fig. 4. The vertical dashed lines represent the stable composition range of the B2 phase at this temperature. The fraction of vacancies on the sublattice occupied by Ni increases to more than 20 at% when moving from the equiatomic composition towards pure Al. This is experimentally verified. But just at the solubility limit this high fraction of vacancies becomes unstable and the vacancy fraction decreases drastically as they are replaced by Ni. In Fig. 4b, the same curves are shown using a logarithmic scale. In Fig. 5a, the chemical potentials of Al and Ni are shown for the B2 phase, at 1500 K. Note that ouside the stable range of B2 these curves represent a metastable state and the endpoints of the curves are not zero because the reference state is fcc (A1). The chemical potential of vacancies is always zero. One can note that there is a metastable miscibility gap in the B2 phase on the Al-rich side, when the fraction of vacancies decrease rapidly. Finally in Fig. 5b, the Gibbs energy for the B2 phase is shown together with the metatstable bcc (A2) phase. One can note that the ordering has a significant influence. There is no explicit ordering parameter in CEF, the chemical ordering is calculated from the constituent fractions, which are found by the Gibbs energy minimization together with the chemical potentials. The rapid change in the chemical potentials in Fig. 5a at equiatomic composition corresponds to the sharp curvature of the Gibbs energy.

From the examples described here it should be clear that CEF models are inherently versatile; a CEF model can readily be constructed to describe any kind of point defect in a multicomponent alloy. Once you have constructed a model for each competing phase of interest, with sublattice parameters that respect their crystal symmetries, thermodynamic software is available to automate the process of Gibbs energy minimization subject to constraints, and to report the resulting phase boundaries and chemical potentials. The difficulty of course is in establishing values for the many parameters that are allowed.

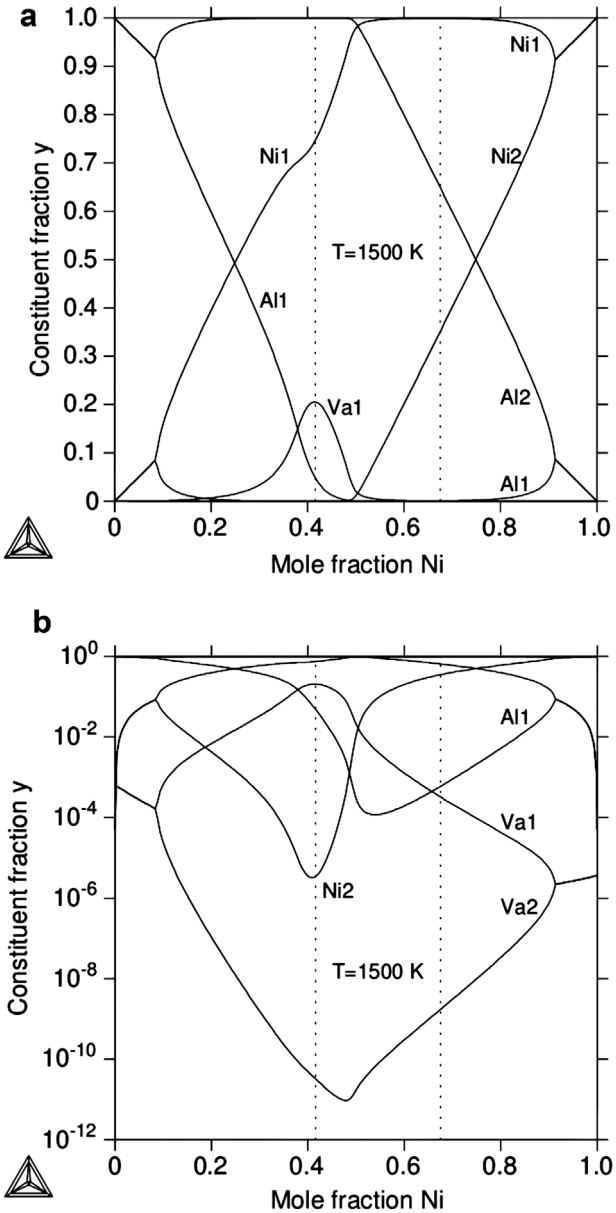

Figure 4 Calculations in AlNi B2 at 1500 K across the whole composition range. In (a), the constituent fractions in the different sublattices, Al1 means fraction of Al on first sublattice, etc. In (b), the same fractions but with a logarithmic scale.

This is where equations such as those introduced in the defect-centric model can be helpful to find the relations between the CEF parameters and various physical properties that can be measured or calculated independently. The end-member energies and combinations of *L* parameters (excess parameters) of the CEF may be fitted in an assessment procedure to experimental data and/or data from DFT calculations.

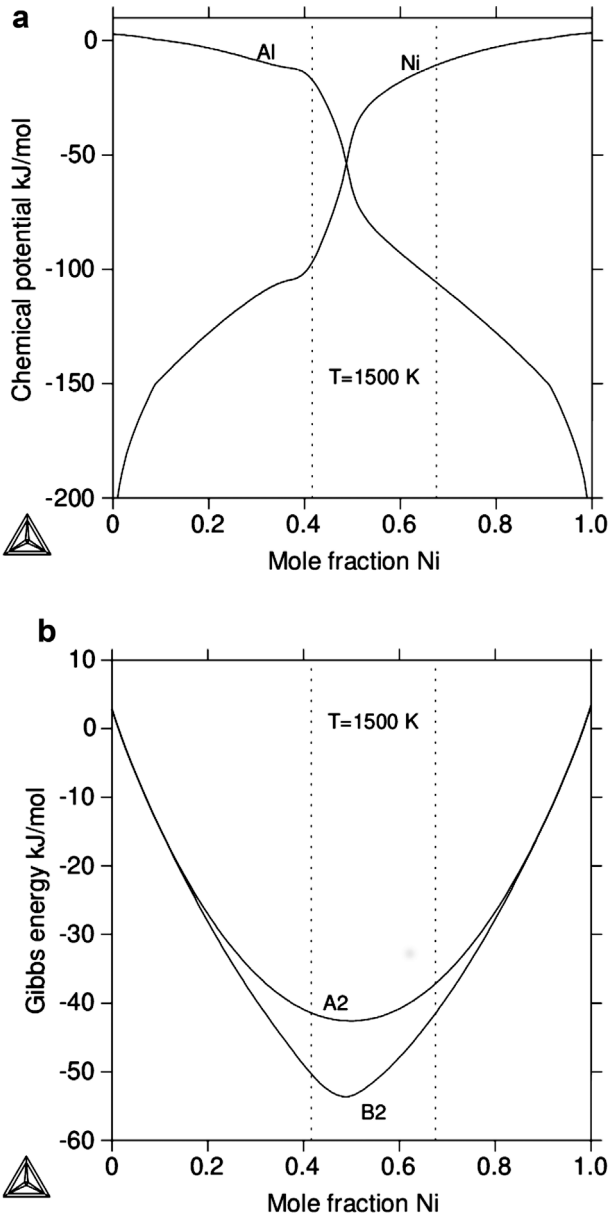

Figure 5 Calculations in AlNi B2 at 1500 K across the whole composition range. In (a) the chemical potentials of Al and Ni and in (b) the Gibbs energy of the B2 and metastable disordered A2 (with same fractions in both sublattices).

In order to have a good fit to such data, it may be necessary to increase the number of excess parameters employed.

The model for the B2 phase described here can also describe the disordered bcc (A2) phase. In order to simplify merging assessments of the bcc phase, with and without ordering, one may separate the parameters in an 'ordered' and a 'disordered' partition as described in Ref. [3].

In conclusion, while the defect-centric approach, using the five computed Gibbs energies, can successfully account for the defect composition and chemical potentials of Ni and Al in B2 Ni_xAl_{1-x} at compositions with dilute defect concentrations (x close to 0.5 on either side of stoichiometry) the greater flexibility of the CEF is necessary in order to account for the boundaries to adjacent phases.

2.5 When is a defect not a defect? If one is only interested in defects when they are in low concentration, in ordered compounds, then there is no need for more than the standard tools of defect chemistry, canonical or grand canonical approaches combined with DFT or more accurate total energy calculations. However, what appears as a defect in one system may be a major component in another, with which it is in equilibrium. This situation is commonplace in the analysis and calculation of phase diagrams. A few examples will serve to show how the CEF may then be helpful in ensuring compatibility in the treatment of Gibbs energies; it is essential that the same energies are attributed to the reference states or end-members in different phases whose Gibbs energies are to be compared.

A simple example of the CEF in action requires the unary phase diagram of iron as a major input. The model is of an important phase in the Fe–C system, in which C is dissolved in bcc iron, the phase known as ferrite. Carbon dissolves by occupying the octahedral interstitial sites, of which there are three for each Fe site. Ignoring thermal vacancies we have a CEF model:

$$(\mathrm{Fe})_1(\mathrm{C,Va})_3.$$

Notice that the concentration of Fe on its sublattice is fixed at 1, so the only variables are the concentrations y_C and y_{Va}, and we can omit the sublattice index. The simplest Gibbs energy expression is:

$$\begin{aligned} G_m = {} & y_{Va}{}^{\circ}G_{Fe:Va} + y_C{}^{\circ}G_{Fe:C} \\ & + 3RT(y_C \ln y_C + y_{Va} \ln y_{Va}) \\ & + y_C y_{Va} L_{Fe:C,Va} + {}^{magn}G_m, \end{aligned} \tag{78}$$

where

$$y_{Va} = 1 - y_C. \tag{79}$$

The quantity ${}^{\circ}G_{Fe:Va}$ is the Gibbs energy of defect-free Fe in a bcc structure with all interstitial sites vacant and the parameter ${}^{\circ}G_{Fe:C}$ represents the Gibbs energy of a hypothetical state in which all interstitial sites are filled with carbon. The last term should be added to deal with magnetism, which is beyond the scope of the present paper and is discussed in [3] and in the paper by Körmann et al. [5]. What is of most interest here is the role of the interaction term, $y_C y_{Va} L_{Fe:C,Va}$. Physically this 'regular solution' term on the sublattice of octahedral sites is introduced in order to account for the interaction between C atoms. Without it we would have a simple ideal solution of C right up to 100% occupancy of the interstitial sites ($y_C = 1$). The bilinear interaction term ensures that this term vanishes in the limits of pure end-members, Fe or FeC_3, which are defect-free. The canonical formalism allows no variations within the fixed composition, but in the grand canonical formalism, or with the standard calculus employed in the CEF as described previously, we can easily minimize

the grand potential

$$\mathcal{G}_m = G_m - y_C \mu_C,$$

which allows the system to be in contact with a reservoir of carbon at chemical potential μ_C, and thus we can obtain the equilibrium concentration of carbon. Since there is only one degree of freedom, represented by y_C, it is very simple to obtain the result for the carbon concentration:

$$G_C = \frac{1}{3}\left(G_{Fe:C} - G_{Fe:Va}\right), \tag{80}$$

$$^{\circ}G_{Fe:C} - {}^{\circ}G_{Fe:Va} + 3RT\left(\ln\frac{y_C^*}{1-y_C^*}\right) + (1-2y_C^*)L_{Fe:C,Va} - \mu_C = 0. \tag{81}$$

Thus, if we assume the carbon concentration is very small and neglect terms in y_C compared to unity, we can express the equilibrium carbon concentration as

$$y_C^* = \exp\left(-\frac{{}^{\circ}G_{Fe:C} - {}^{\circ}G_{Fe:Va} + L_{Fe:C,Va} - \mu_C}{3RT}\right). \tag{82}$$

From the defect-centric point of view, we would identify the first three energies in the numerator as equivalent to the raw carbon interstitial formation energy, as calculated for example by DFT as described previously:

$$g_C^i \equiv {}^{\circ}G_{Fe:C} - {}^{\circ}G_{Fe:Va} + L_{Fe:C,Va}. \tag{83}$$

Therefore, since the first two energies are in principle known or can also be calculated, the interaction parameter $L_{Fe:C,Va}$ would have to be adjusted to fit g_C^i. This is the same situation we found for a dilute vacancy concentration with ${}^{\circ}G_{Va}$ and $L_{A,Va}$ in Eq. (28).

In the first assessments of the C–Fe system with this model [16], only the ${}^{\circ}G_{Fe:C}$ was used to fit the experimental data and that worked well when calculating up the 25 mol% of carbon. However, the assessed value of ${}^{\circ}G_{Fe:C}$ had the undesired effect of stabilizing the ferrite phase relative to graphite at high temperatures and high carbon compositions, i.e. ${}^{\circ}G_{Fe:C}$ was too negative at high temperature. To prevent this a positive value was added to ${}^{\circ}G_{Fe:C}$ and $L_{Fe:C,Va}$ was set to the negative of the same value. In this way, it was ensured that the same chemical potentials were generated within the stable dilute carbon range at temperatures where ferrite is experimentally stable. Nowadays, the possibility of calculating ${}^{\circ}G_{Fe:C}$ by DFT enables one to assess a realistic value for $L_{Fe:C,Va}$ for ferrite. Care must always be taken in this situation to align the reference energies in DFT calculations with conventions used in thermodynamics as discussed in the following.

2.5.1 The use of DFT calculations and multicomponent consistency A few remarks are made in this section on the matter of consistency between thermodynamic calculations and calculated input for such studies, especially concerning defects.

The reader who has used DFT calculations in thermodynamics will be well aware that care must be taken to align the reference energies in DFT calculations with conventions used in thermodynamics. For example the chemical potential, or Gibbs energy per mole, of an element in its pure equilibrium phase at a certain temperature can certainly be calculated by DFT. This number may be shifted to zero for reference purposes in a thermodynamic calculation, but the DFT energies of all compounds containing that element will then have to be shifted by the same amount. This applies, for example to carbon in the previous section; if the ferrite is in equilibrium with pure graphite, the thermodynamicist might take $\mu_C = 0$ for convenience, and other reference states of pure elements similarly, and the calculated data will have to be shifted accordingly. However, the problem is a very general one.

When developing multicomponent thermodynamic databases, the modelling of defects in intermetallic phases is carried out in many systems, such as Laves phases, σ, B2, etc. It is important to consider how these defects are combined. Considering, for example austenite in the C–Fe system and cubic carbide in C–Ti , what is a defect in the first system may be the major constituent in the second. This is one of the main reasons for using models that treat defects on the same footing as major constituents.

In order to merge assessments of systems where an intermetallic phase ϕ is stable and modelled with defects, the Gibbs energies of a pure element in this intermetallic structure, denoted ${}^{\circ}G_{A:A}^{\phi}$, must be the same in all systems. This requires some international agreement and there is a recent publication for some common ordered intermetallic by Sluiter [17].

3 Extension to charged defects

3.1 *Ab initio* defect chemistry The formalism introduced in Section 2.4.4 can be straightforwardly extended to cover also the case of charged defects. Two extensions are needed. First, the defect formation energy in Eq. (65) depends not only on the chemical potentials but, since for charged defects electrons have to be added/removed, also on the chemical potential of the electrons, i.e. the Fermi level:

$$G_\delta = g_\delta - \sum_i \Delta n_i \mu_i + q(E_{VB} + E_{Fermi}). \tag{84}$$

Here, q is the charge state of the defect. A positive integer number indicates that q electrons have been removed from the defect and transferred to the electron reservoir. E_{Fermi} is the Fermi energy (which can be regarded as the chemical potential of the electron reservoir) and E_{VB} is the reference electron energy against which the Fermi level is measured. Typically, as reference energy the top of the valence band (VB) edge is used. Both, E_{VB} and G_δ (which is now q-dependent), can again be computed directly by DFT. For practical DFT defect calculations, great care has to be devoted to spurious effects

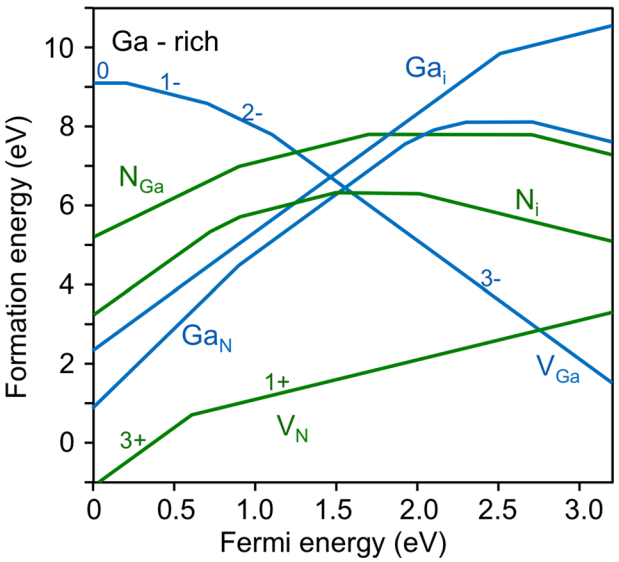

Figure 6 Formation energy of all relevant native point defects in GaN as a function of the Fermi level as obtained from DFT calculations together with Eq. (84). Ga-rich conditions are assumed. Figure taken from Ref. [22].

that arise from the finite size of supercells [18–20]. The particularly large magnitude of such effects for charged defects is due to the long-range character of the Coulomb interaction.

Having expressed in Eq. (84), the defect formation energy in terms of quantities amenable to *ab initio* calculation, together with chemical potentials and a Fermi level that are variable, we can apply the same strategy as outlined in Section 2.4.4. To be more specific let us consider point defects in GaN, a defect system that has been extensively studied in the past [21]. Since GaN is a wide-bandgap semiconductor the range of possible Fermi energies is large, making charged defects particularly important. Possible intrinsic point defects are antisites (Ga_N, N_{Ga}), vacancies (Va_{Ga}, Va_N) and interstitials (Ga_{int}, N_{int}). The defect formation energy depends on two chemical potentials (μ_{Ga}, μ_N) that are coupled via the condition of being in thermodynamic equilibrium with bulk GaN:

$$\mu_{Ga} + \mu_N = {}^{\circ}G_{Ga:N}. \tag{85}$$

Here, because the defect concentration is very low, ${}^{\circ}G_{Ga:N}$ is an adequate approximation to the chemical potential of GaN, for which the above equality would be exact. For wide bandgap semiconductors and insulators a common and very intuitive approach to visualize and discuss the energetics of the various defects is to plot their energy for fixed chemical potential(s) as a function of the Fermi level. For GaN such a plot is shown in Fig. 6, in which the chemical potential of Ga is fixed to its maximum value, representing the Ga-rich condition.

Equation (73) implies that the dominant defects, i.e. the defects with the highest concentration, are the ones which have the lowest defect formation energy. Figure 6 thus reveals immediately the dominant defect: under p-type conditions this is the N-vacancy, under n-type conditions it is the Ga-vacancy. All other defects are higher in energy and occur thus in much lower concentrations. Another important piece of information that becomes directly accessible from this figure is the electrical activity of the defect (i.e. whether it acts as an acceptor, a donor or an amphoteric defect center), and the position of the electronic *charge transfer level* (i.e. at which position of the Fermi level the defect changes its charge state).

The character of a defect is given by the slope of its line on Fig. 6. The slope is simply the charge state q of the defect: a positive slope thus means that the defect has donated q electrons (i.e. it is a donor) whereas a negative slope indicates an acceptor. Regions where the slope is zero indicate that the defect is for this range of Fermi levels charge neutral. As can be seen in Fig. 6, in its neutral charge state the defect has attained its highest formation energy, i.e. in this state its concentration will be lowest. The lowest formation energies arise at the band edges: Since under p-type conditions defects with a positive slope are favoured, which act as donors, these defects partly compensate the origin of the p-type doping and thus are detrimental to device performance. The same argument applies to the other extreme (n-type doping) where acceptors become dominant.

The second piece of information, the position of the charge transfer levels, is given by the kinks in the defect energies in Fig. 6. The kinks arise due to a change in the slope (which is proportional to the charge q), i.e. the position on the x-axis where the kink occurs is the Fermi level position at which the defect changes its charge. This quantity is an important fingerprint for identifying defects and crucial for understanding and optimizing the performance of semiconductor devices.

In a real semiconductor, the Fermi level is not a variable but has to be chosen such that the system is charge neutral. This condition can be enforced by requiring the excess charge of the system to vanish:

$$\begin{aligned} 0 &= q_{\text{excess}}(E_{\text{Fermi}}) \\ &= e(E_{\text{Fermi}}) - h(E_{\text{Fermi}}) - \sum_{\delta} q_\delta y_\delta(E_{\text{Fermi}}). \end{aligned} \tag{86}$$

Here, e is the number of electrons per mole of GaN in the conduction band, h the number of holes per mole of GaN in the VB, and the index δ runs over all defects and charge states. y_δ is the concentration of defect δ that is obtained by substituting the defect formation energy computed via Eq. (84) in Eq. (73). For a given set of defects the charge neutrality condition together with fixing the chemical potentials uniquely determines the Fermi level and thus the defect concentrations.

The grand canonical approach outlined above applies not only to the case of intrinsic point defects discussed here, but can be easily extended to the technologically more relevant

scenario of extrinsic defects such as intentionally or unintentionally incorporated dopants and their interaction with the native defects. The only modifications needed are to extend the index α to include the additional defects and defect complexes and include the additional chemical reservoirs needed to create the defect. For example to consider p-type doping by Mg in GaN additional defects to include would be Mg_{Ga}, Mg_N and Mg_{int}. In addition, complexes between the Mg acceptor and the dominant point defects may occur. A prominent and thermodynamically stable complex is, e.g. Mg-Va_N. An additional chemical potential μ_{Mg} is introduced and additional constraints on the chemical potentials to prevent the formation of parasitic phases such as Mg_3N_2 have to be applied, namely in this case:

$$3\mu_{Mg} + 2\mu_N \leq \mu_{Mg_3N_2} \equiv {}^{\circ}G_{Mg:N} \ . \tag{87}$$

3.2 Modelling defects in GaN by combining DFT with CEF As described in Section 3.1, GaN can change its composition mainly by creating Ga and N vacancies since all other defects such as interstitials or antisites are higher in energy (see Fig. 6). As shown in Fig. 6, the dominant defect over a large range of Fermi energies is the singly positively charged nitrogen vacancy. In the CEF model, rather than dealing directly with charged vacancies, we suppose that the charge state of the defect is realized by a change in the charge state of the neighbouring elements. Thus, if only this defect is included, the corresponding CEF model is written as:

$$(Ga^{+3}, Ga^{+2})_1(N^{-3}, Va)_1 \ . \tag{88}$$

The Gibbs energy with ${}^{E}G_m = 0$ is:

$$\begin{aligned} G_m = {} & y_{Ga^{+3}} y_{N^{-3}} \, {}^{\circ}G_{Ga^{+3}:N^{-3}} + y_{Ga^{+3}} y_{Va} \, {}^{\circ}G_{Ga^{+3}:Va} \\ & + y_{Ga^{+2}} y_{N^{-3}} \, {}^{\circ}G_{Ga^{+2}:N^{-3}} + y_{Ga^{+2}} y_{Va} \, {}^{\circ}G_{Ga^{+2}:Va} \\ & + RT(y_{Ga^{+3}} \ln(y_{Ga^{+3}}) + y_{Ga^{+2}} \ln(y_{Ga^{+2}}) \\ & + y_{N^{-3}} \ln(y_{N^{-3}}) + y_{Va} \ln(y_{Va})) \ . \end{aligned} \tag{89}$$

The constitutional square, similar to that in Fig. 1, is shown in Fig. 7. Note that all end-members except ${}^{\circ}G_{Ga^{+3}:N^{-3}}$ have a net charge and that only compositions along the neutral line have a physical meaning. This line represents the constraint:

$$y_{Ga^{+2}} = 3y_{Va}. \tag{90}$$

The chemical potential for N in this model can be derived from the end-members in the same way as for C by Eq. (80) in Section 2.5, taking also the electronic charge neutrality into account:

$$G_N = 3 \, {}^{\circ}G_{Ga^{+3}:N^{-3}} - 2 \, {}^{\circ}G_{Ga^{+2}:N^{-3}} - {}^{\circ}G_{Ga^{+2}:Va}. \tag{91}$$

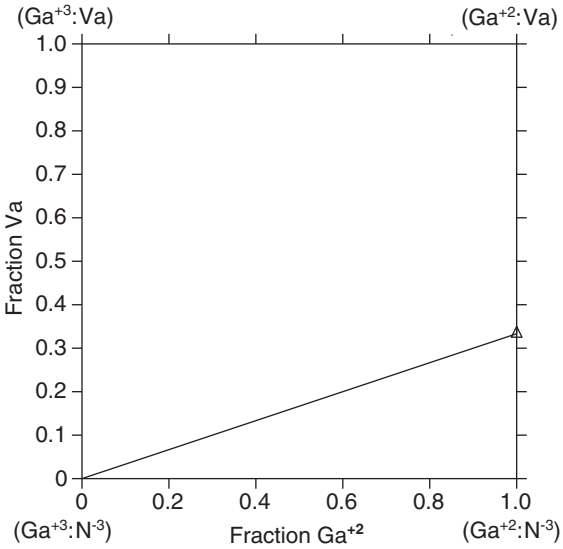

Figure 7 The constitutional square of GaN in the CEF. The line represents electrically neutral combinations of defects.

Eliminating some common terms gives

$$\begin{aligned} G_N = {} & 3\left(\frac{\partial G_m}{\partial y_{Ga^{+3}}} + \frac{\partial G_m}{\partial y_{N^{-3}}}\right) - 2\left(\frac{\partial G_m}{\partial y_{Ga^{+2}}} + \frac{\partial G_m}{\partial y_{N^{-3}}}\right) \\ & - \left(\frac{\partial G_m}{\partial y_{Ga^{+2}}} + \frac{\partial G_m}{\partial y_{Va}}\right). \end{aligned} \tag{92}$$

It is convenient for writing the subsequent equations to abbreviate the labels of the constituents in the CEF model that are given in full by (88); let us replace their labels as follows: $Ga^{+3} \to 3$, $Ga^{+2} \to 2$, $N^{-3} \to N$ and $Va \to V$. The CEF model can then be represented as

$$(3, 2)(N, V). \tag{93}$$

Using Eq. (89) with this notation we find:

$$\begin{aligned} G_N = {} & 3(y_N \, {}^{\circ}G_{3:N} + y_V \, {}^{\circ}G_{3:V} + RT \ln(y_3) \\ & + y_3 \, {}^{\circ}G_{3:N} + y_2 \, {}^{\circ}G_{2:N} + RT \ln(y_N)) \\ & - 2(y_N \, {}^{\circ}G_{2:N} + y_V \, {}^{\circ}G_{2:V} + RT \ln(y_2) \\ & + y_3 \, {}^{\circ}G_{3:N} + y_2 \, {}^{\circ}G_{2:N} + RT \ln(y_N)) \\ & - (y_N \, {}^{\circ}G_{2:N} + y_V \, {}^{\circ}G_{2:V} + RT \ln(y_2) \\ & + y_3 \, {}^{\circ}G_{3:V} + y_2 \, {}^{\circ}G_{2:V} + RT \ln(y_V)) \\ = {} & 3y_N({}^{\circ}G_{3:N} - {}^{\circ}G_{2:N}) + 3y_V({}^{\circ}G_{3:V} - {}^{\circ}G_{2:V}) \\ & + y_3({}^{\circ}G_{3:N} - {}^{\circ}G_{3:V}) + y_2({}^{\circ}G_{2:N} - {}^{\circ}G_{2:V}) \\ & + RT \ln\left(\frac{y_3^3 y_N}{y_2^3 y_V}\right). \end{aligned} \tag{94}$$

Individual end-members that carry a net charge cannot have any physical meaning, and they can only enter a

physical equation in neutral combinations. Thus it is only the sum $2\,{}^{\circ}G_{2:\mathrm{N}} + {}^{\circ}G_{2:\mathrm{V}}$ that has a physical meaning, as it is the end point of the neutral line. The corollary of this observation is that it is only necessary to obtain data for the neutral combination of end-members, and how the total energy of such a combination is divided between them is purely a matter of convenience. To understand this point it may be helpful to consider the energy terms that enter the physical process of creating a vacancy on the N sublattice by removing a nitrogen atom. In terms of the given end-members, this can be done by converting three neutral molecules of energy ${}^{\circ}G_{3:\mathrm{N}}$ into one molecule of ${}^{\circ}G_{2:\mathrm{V}}$ with charge $= +2$ and and two molecules of ${}^{\circ}G_{2:\mathrm{N}}$, each with charge $= -1$. The constraint of charge neutrality, which requires $y_{\mathrm{V}} = y_2/3$, does not allow us any other combination.

In order to eliminate the unphysical end-member ${}^{\circ}G_{3:\mathrm{V}}$ from the problem another constraint between the energies of the end-members can be written down, which expresses the fact that we are assigning the same energy to the system irrespective of whether a gallium ion carrying a charge +2 is next to a vacancy or far away from it. This is expressed by the relation

$$ {}^{\circ}G_{3:\mathrm{V}} + {}^{\circ}G_{2:\mathrm{N}} = {}^{\circ}G_{2:\mathrm{V}} + {}^{\circ}G_{3:\mathrm{N}}\,. \tag{95} $$

Thus by eliminating ${}^{\circ}G_{3,\mathrm{V}}$ and applying also the constraints of total site occupancy, $y_3 + y_2 = 1$ and $y_{\mathrm{N}} + y_{\mathrm{V}} = 1$, we get:

$$ G_{\mathrm{N}} = 3\,{}^{\circ}G_{3:\mathrm{N}} - 2\,{}^{\circ}G_{2:\mathrm{N}} - {}^{\circ}G_{2:\mathrm{V}} + RT\ln\left(\frac{y_3^3 y_{\mathrm{N}}}{y_2^3 y_{\mathrm{V}}}\right). \tag{96} $$

Notice that total charge neutrality is preserved in all these equations. In the dilute limit $y_3 \to 1$ and $y_{\mathrm{N}} \to 1$. The chemical potential of N in an N_2 gas is $G_{\mathrm{N}} = 0.5RT\ln(p_{\mathrm{N}_2})$, where p_{N_2} is the partial pressure of N_2 and hence we can write the relation between the partial pressure of N_2 and the fraction of N vacancies, y_{V} as

$$ \begin{aligned} \frac{1}{y_{\mathrm{V}}^4} &\sim \sqrt{p_{\mathrm{N}_2}}, \\ y_{\mathrm{V}} &\sim p_{\mathrm{N}_2}^{-1/8}. \end{aligned} \tag{97} $$

This relation can be derived more easily using the Kröger–Vink notation [23] but if there are several competing reactions, as in this case where the DFT calculations show that a dominant reaction is creating a neutral Ga vacancy plus 3 N vacancies with a charge of -2, it can become complicated to handle all of them using Kröger–Vink. For such cases it is more convenient to use the grand canonical defect approach as outlined in Section 2.4.4.

The CEF model taking into account all three thermodynamically possible defects (i.e. the singly and triply charged N vacancy and the triply charged Ga vacancy, c.f. Fig. 6) would be

$$ (\mathrm{Ga}^{+3}, \mathrm{Ga}^{+2}, \mathrm{Va})(\mathrm{N}^{-3}, \mathrm{Va}^{-2}, \mathrm{Va})\,. \tag{98} $$

If also minority defects have to be included, i.e. defects that require a higher energy to be formed than the energetically most favourable ones (e.g. the $\mathrm{Ga_N}$ antisite defect under p-type conditions, c.f. Fig. 6), then the CEF has to be extended and additional end-members have to be taken into account. Another system with multiple defects in ionic crystals has also been modelled using CEF by Sundman et al. [24].

3.3 Comparison of methods For binaries, the formalism outlined in Section 2.4.4 in connection with *ab initio* total energy techniques is the method of choice to study intrinsic and extrinsic defects in semiconductors and insulators [21] and has been successfully used to address a wide range of materials and device questions in semiconductor physics. For multicomponent systems, it is advantageous to employ the tools and concepts developed within CALPHAD. As discussed in Section 3.2, CEF is well suited to include a large number of possible defects and charge states when properly extended. However, since the end-members are only auxiliary quantities needed to constrain defect concentrations to physical quantities great care has to be devoted to ensure that the energies deduced from models based on CEF in the dilute limit are identical to those obtained by the *ab initio* calculations.

4 Calculation of point defect properties

4.1 Total energy calculations The possibility to determine point defect properties independent from experimental data provides an important tool to improve our fundamental understanding of the effect of point defects on materials properties. Furthermore, first-principles calculations can be very useful in extending databases for thermodynamic modelling as discussed in Sections 2 and 3. Currently, the most widely used approach to study point defects in metals *ab initio* is DFT. One of the first, groundbreaking studies of point defects using DFT is the work by Gillan [25], where its application to the vacancy formation energy in aluminium is discussed. Since then, DFT calculations have been performed for many other materials and DFT is nowadays the method of choice due to its high accuracy at feasible computational cost.

The accuracy of defect formation energies obtained using DFT depends, however, on the particular setup chosen for the calculation. There are two types of parameters in the computational setup of DFT calculations: controllable parameters, which can be converged w.r.t. a certain property, for which a reliable error estimate can be provided, and non-controllable parameters, which arise from the inherent approximations within DFT implementations. Controllable parameters include, e.g. the quality of the basis set as well as the sampling density of k-points. In DFT codes based on plane waves the quality of the basis set is determined entirely by the energy cutoff for the plane waves. Furthermore, the size of the supercell has to be carefully chosen to avoid artificial interactions between periodic images of the defect. This is of particular importance if there are long-range interactions, e.g. in charged defects or due to long-range elastic

effects. The convergence of these parameters can significantly influence the investigated defect properties and must be carefully assessed. Non-controllable parameters depend on the particular flavour of DFT and its implementation. The reliability of the often used pseudopotential approach can be tested against all-electron calculations [26]. The choice of the exchange-correlation (XC) functional is, however, not unique since it cannot be said *a priori* which XC functional performs best to describe a certain point defect. The most commonly used XC functionals include the local-density approximation (LDA) [27, 28] and the generalized-gradient approximation (GGA) [29–31]. In metals, the GGA tends to give more accurate results for bulk properties such as equilibrium lattice constants or bulk moduli [32]. But it has been shown that LDA provides generally more accurate values for vacancy formation energies [33–35]. This was attributed to the fact that the cohesion in metals is treated differently by the different XC functionals, which influences bulk properties like the equilibrium lattice constant and thus also the vacancy formation energy. It is noteworthy that the formation of a vacancy creates an internal surface, and it has been shown that LDA gives better results for surface energies due to a cancellation of errors [32]. This suggests a way around the problem: on the assumption that GGA and LDA do not correctly describe the internal surface of a vacancy, a surface correction was introduced, which improved results for vacancy formation energies in Al [33] as well as Pt, Pd and Mo [34]. A recent study discusses in detail the influence of XC functionals and available correction schemes for the internal surface on vacancy formation energies in fcc metals [36]. When discussing values for the vacancy formation energy from literature one has to be careful, though, since the convergence of the computational parameters can have a significant influence as shown in Section 4.5.

Although DFT has been very successful in calculating defect properties in metals, the uncertainty arising from the XC functional remains. Various improved XC functionals are being developed but since there is no rigorous systematic improvement it is difficult to decide which XC functional provides the most accurate results. The quantum Monte Carlo (QMC) method [37] can provide an alternative to DFT for the calculation of point defect properties [38], in particular the more sophisticated diffusion quantum Monte Carlo (DMC) approach. Recently, point defects in aluminium have been studied using the DMC approach which is the first application of the QMC method to point defects in metals [39]. Compared to DFT, the DMC calculations are computationally extremely demanding. Furthermore, the calculation of forces is very expensive and still under development [40]. The structures calculated with QMC are thus typically relaxed using DFT. The high computational cost prevents QMC from being widely used at present, but in the future with increasing computational resources and further methodological developments QMC calculations can become a reliable benchmark.

Constructing appropriate supercells to calculate point defect properties is crucial as discussed in more detail in Section 4.2. In particular, when studying the interaction between various point defects large supercells are required to avoid artificial interactions between periodic images. At this point DFT calculations start to become prohibitively expensive, requiring further developments in the field of linear scaling electronic structure methods [41]. Alternatively, less accurate, but much faster and still reliable interatomic potentials may be applied. Interatomic potentials can roughly be divided into two groups: potentials based on physical intuition and pure mathematical interpolation schemes. The quality and transferability to point defect calculations of these potentials can vary significantly and there is no systematic prediction of how well a potential will perform in determining defect properties. A successful approach in the construction of interatomic potentials for unary metals is the embedded atom method (EAM) [42] and good agreement was found for vacancy formation energies in various simple metals such as Cu or Al [43–45]. Multicomponent systems beyond binaries are, however, difficult to treat and the EAM potentials exhibit only a low transferability. Another example of empirical potentials are ReaxFF potentials. They have mainly been developed for molecular systems, but there have also been some applications to bulk systems and defect properties in reasonable agreement with the respective DFT data [46]. If the electronic structure needs to be explicitly considered tight-binding (TB) models [47] are a good compromise between DFT calculations and empirical potentials. TB models are directly derived from the electronic structure and besides their wide applicability to point defects in unaries [48, 49] it is also possible to treat the magnetic properties of point defects [50]. A coarser approximation to the electronic structure at lower computational cost is provided by bond-order potentials (BOP) [51]. BOPs have also been applied to different unary metals and show good results with respect to defect properties [52, 53]. Two approaches that aim for a mathematical interpolation of the high dimensional potential energy surface are Gaussian approximation potentials (GAP) [54] and neural networks (NN) [55, 56]. In contrast to interatomic potentials where the potential parameters are fitted to reproduce certain system properties, GAPs or NNs are fitted to a large set of structures that represent the different local environments of an atom. Examples include the migration of a vacancy in diamond using a GAP [54] and defect properties in Cu using a NN [57]. Since these potentials are constructed for interpolation, they give excellent agreement with the corresponding DFT data if defect structures are included in the training set. If not, the agreement can still be reasonable but has to be carefully evaluated.

A short comparison of the different methods to evaluate the energetics of point defect properties is given in Table 2.

4.2 Supercell setup for defect calculations

Defect formation energies are obtained by calculating the difference between the total energy of a supercell containing the defect and a corresponding defect-free supercell, as shown for vacancy formation energies in Eq. (7). This is

Table 2 Comparison of computational approaches to obtain the energetics of point defect properties. The range −− to ++ refers to the capability of a method to capture different atomic environment (transferability) respectively to properly treat interactions between an arbitrary number of elements (multicomponent).

	computational cost	transferability	multicomponent	magnetism	charge transfer
DFT	high	++	++	explicit	yes
QMC	very high	++	++	explicit	yes
TB	medium	+	−	explicit	yes
BOP	low	+	−	explicit	yes
EAM	very low	−	− −	implicit	no
ReaxFF	low	−	−	implicit	yes
GAP	low	−	+	implicit	yes
NN	low	−	+	implicit	yes

generally applicable to all types of defects and the calculation of a single supercell, including relaxation of atomic positions, is in itself a straightforward task. Special care must, however, be taken in constructing the appropriate supercell. The presence of the defect affects the surrounding lattice in two ways: electronic effects due to the modified bonds to the neighbouring atoms, and elastic effects due to the distortion of the surrounding lattice. The supercell setup of a DFT calculation thus needs to ensure that both effects are taken into account and that the periodic images do not interact with each other.

Since the defects carry an excess volume that is constrained by the periodicity of a supercell, different relaxation procedures of the supercell may lead to different results. In the constant volume approach, the lattice constant of the supercell is kept at the value for the perfect bulk, only the atomic positions are relaxed. Alternatively, one may also rescale the lattice constant to account for the excess defect volume ('rescaled volume approach'). The constant pressure approach amounts to fully relaxing both the lattice constant and the atomic positions. With increasing supercell size the respective energies converge to the energy of a truly isolated defect in the bulk [58]. This is illustrated in Fig. 8 for a vacancy in fcc Al [59]. The top graph shows the energy versus volume curves for a perfect bulk cell and for $(1 \times 1 \times 1)$, $(2 \times 2 \times 2)$, $(3 \times 3 \times 3)$, $(4 \times 4 \times 4)$ supercells with one vacancy and respectively 3, 31, 107 and 255 atoms. With increasing size of the supercell the equilibrium volume converges towards the equilibrium volume of the perfect bulk cell. The bottom graph shows the difference in vacancy formation energy at constant volume as well as zero pressure. As the size of the supercell increases the two approaches converge.

An additional complication arises in the context of multicomponent alloys. While the notion of a single defect in a pure bulk system is well defined, the properties of a defect in an alloy are dependent on the local environment, and as such the whole situation is much more complex. In alloys with several components, the sheer volume of the configuration space that needs to be sampled prohibits a direct approach. Several approaches have been developed to tackle these problems, including the coherent potential approximation (CPA) [60–62], the cluster expansion method [63] or special quasi-random structures (SQS). The latter is discussed in more detail in the following.

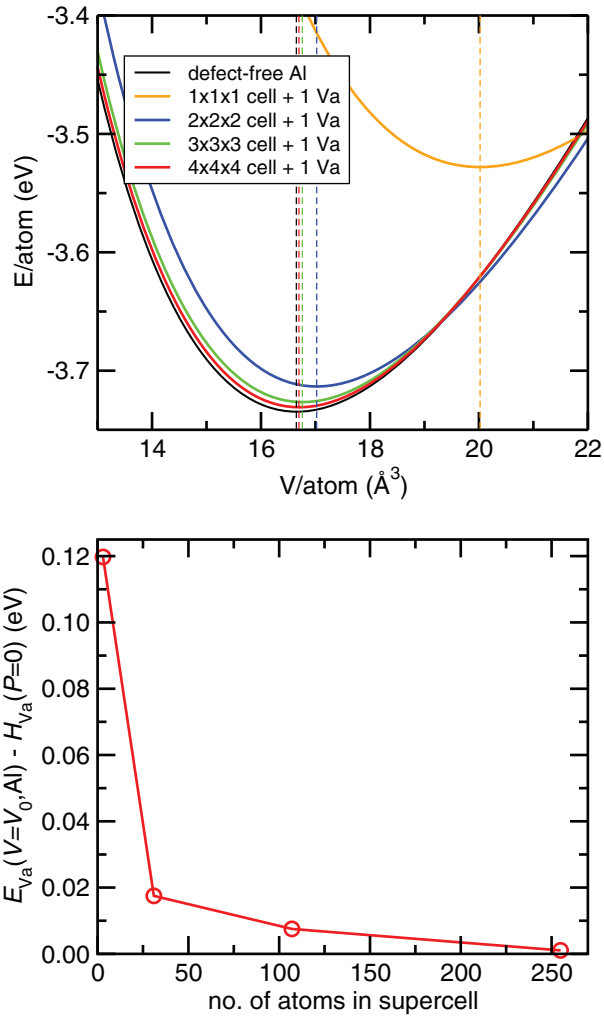

Figure 8 Top graph: energy versus volume curves for perfect fcc Al and supercells with one vacancy; bottom graph: difference in vacancy formation energy calculated at $T = 0$ K and constant volume, respectively zero pressure versus size of the supercell.

4.3 Disordered alloys: Example Cu–Ni The calculation of vacancy formation energies in disordered alloys has recently received increased attention [64–69]. For simplicity, we limit the discussion to configurationally random alloys. The extension to alloys with short range order will only be briefly discussed. In a disordered alloy with atomic species A and B, it is important to recognize that vacancies can be surrounded by a varying number of A and B atoms. In the nearest neighbour shell in an fcc alloy the 12 nearest neighbours of a vacancy can range from 12 A and 0 B atoms to 0 A and 12 B atoms. The composition of the nearest neighbour, and of more distant neighbour shells, affects the vacancy formation energy. It is then apparent that 'the vacancy formation energy' in a disordered alloy requires a careful definition because the vacancy formation energy must be a function of the atomic neighbourhood of the vacancy, the composition of the alloy and other factors.

In a random A–B alloy, once there is a vacancy, it cannot be known whether the removed atom was an A or a B atom. In order to preserve the composition of the alloy, only a 'random' atom could be removed, i.e. x_A times an A atom and x_B times a B atom. It follows that a weighted average over A and B removed supercells must be considered, and Eq. (7) becomes

$$G^f_{Va}(x_A, x_B) = x_A G_{1Va,A} + x_B G_{1Va,B} - \frac{N-1}{N} G_{perfect}, \quad (99)$$

where N is the number of atoms in the cell without a vacancy. Note that in this and the following equations, we ignore the difference between x_k and y_k for real components, since the vacancy concentration is assumed to be very low. Actual supercells contain only small numbers of atoms, and are therefore never truly 'random'. The best possible choice typically is achieved by constructing a SQS, which for all important correlation functions in the alloy reproduces the values for truly random structures [70]. In such a supercell, one can then remove one atom at a time, and define the vacancy formation energy as an appropriate average

$$G^f_{Va}(x_A, x_B) = \frac{1}{N}\sum_{i=1}^{N}\left[G_{1Va,i} - \frac{N-1}{N} G_{perfect}\right]. \quad (100)$$

The above equation is not physically relevant because it averages over local vacancy neighbourhoods. In an actual alloy, vacancies would occur where favourable local neighbourhoods exist, so that the effective vacancy formation energy should be biased towards the lowest energy neighbourhoods. In an alloy where high A vacancy coordination gives a lower vacancy formation energy, but with low A concentration, it is improbable to find a neighbourhood with A atoms only. Therefore, the bias towards the lowest energy configurations is limited by combinatorial factors. The combinatorial factors are easily evaluated for the configurationally random alloy, but can be determined also for a prescribed state of short range order, e.g. through constrained Monte Carlo simulations. If the effective interactions between vacancies and A or B atoms are limited to the nearest neighbours, the sum in Eq. (100) could be limited to those atomic positions, which have a particular neighbourhood α only,

$$G^f_{Va}(\alpha; x_A, x_B) = \frac{1}{N_\alpha}\sum_{i_\alpha=1}^{N_\alpha}\left[G_{1Va,i_\alpha} + \mu_i - G_{perfect}\right], \quad (101)$$

where μ_i is the chemical potential of atom i. It should be noted that this definition is akin to Refs. [71, 72], but differs from those in Ref. [68] where the chemical potential term is replaced with an elemental energy. The concentration of each type of vacancy $y_{Va}(\alpha)$ can then be computed from

$$y^*_{Va}(\alpha; x_A, x_B) = P_\alpha e^{-\beta G^f_{Va}(\alpha; x_A, x_B)}, \quad (102)$$

where $\beta = 1/k_B T$ and P_α is the probability that any site in the alloy has a neighbourhood of type α. In the random alloy

$$P_\alpha = m_\alpha x_A^{n_{A\alpha}} x_B^{n_{B\alpha}}, \quad (103)$$

where m is the degeneracy of a particular neighbourhood, and $n_{A\alpha}(n_{B\alpha})$ is the number of A (B) atoms in the neighbourhood of type α. The sum of P_α over all types of neighbourhoods yields unity. The total vacancy concentration is then obtained from

$$y^*_{Va}(x_A, x_B) = \sum_\alpha y^*_{Va}(\alpha; x_A, x_B). \quad (104)$$

It follows that the total vacancy concentration does not necessarily obey a simple Arrhenius equation, although in the case considered below, it is observed to do so. In practice, it is rather cumbersome to generate SQS supercells that contain all types of neighbourhoods. Just considering the nearest neighbour shell in fcc solid solutions alone gives 144 distinct configurations [73] in a binary alloy. Therefore, it is usually more efficient to compute neighbourhood-dependent vacancy formation energies through cluster expansions [65, 67].

As an illustration we discuss here fcc Cu–Ni alloys. The alloys are assumed to be configurationally random, which is approximated using SQSs with 16 atoms per cell with compositions Cu_4Ni_{12} , $Cu_{12}Ni_4$ (both structures listed as SQS-1 in Ref. [74]), and Cu_8Ni_8 (two variants listed as SQS-2 and SQS-3 in Ref. [74]). It should be remarked that generally such small cells do not give very accurate vacancy formation energies and the currently obtained values are a bit too small [36], but here the objective is not high accuracy but proof of principle. The pure Cu and Ni phases are considered using the same type of supercell as for the Cu_4Ni_{12} and $Cu_{12}Ni_4$ compositions [75]. Here, only $T = 0$ K is considered with zero point vibrational corrections being neglected. Enthalpy of formation of the SQS, computed as described in

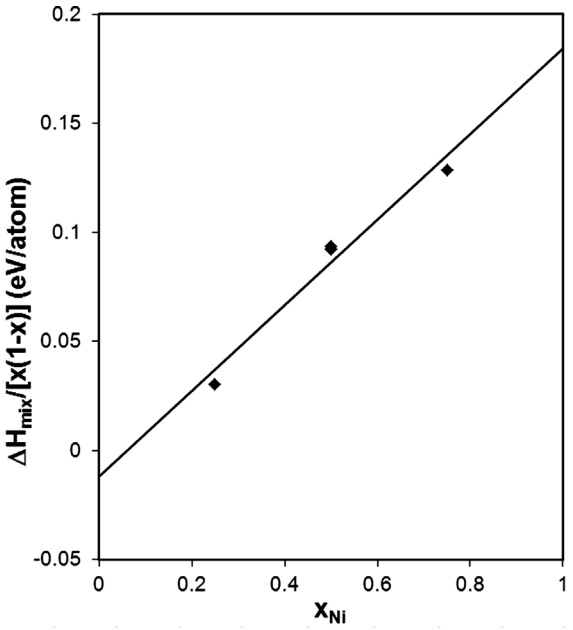

Figure 9 $\Delta H_{\mathrm{mix}}/[x_{\mathrm{Ni}}(1-x_{\mathrm{Ni}})]$ as function of the atomic concentration of Ni as extracted from *ab initio* computed formation energies of SQSs (diamond symbols).

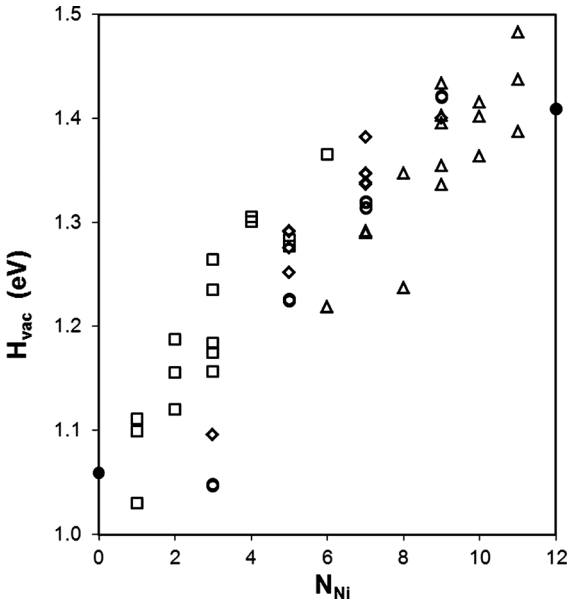

Figure 10 Vacancy formation enthalpy as function of the number of Ni atoms in the nearest neighbour shell (N_{Ni}) as computed with Eq. (101) at various compositions; solid circles: in pure Cu and Ni; squares: in SQS-1 at x_{Ni}=0.25; diamonds: in SQS-2 at x_{Ni}=0.5; open circles: in SQS-3 at x_{Ni}=0.5; triangles: in SQS-1 at x_{Ni}=0.75.

Ref. [74], is used as a proxy for the mixing enthalpy ΔH_{mix}. Figure 9 illustrates that the compositional dependence of the mixing enthalpy is well approximated by a subregular solution model. The chemical potential of Cu and Ni is then easily extracted from the mixing enthalpy as function of the alloy composition.

In the SQSs, single vacancies were introduced by removing one atom at a time, followed by a structural relaxation. The enthalpies so obtained could be used in Eq. (101) to extract vacancy formation enthalpies for various vacancy neighbourhoods, see Fig. 10. It is evident that the larger the number of Ni atoms in the nearest neighbour shell, the larger the vacancy formation energy. This is consistent with the larger vacancy formation energy in pure Ni in comparison with pure Cu. The composition of the alloy also plays a role: at a given number of Ni atoms in the nearest neighbour shell the larger Ni concentrations are associated with vacancy formation enthalpies at the bottom of the range, while at smaller Ni concentrations vacancy formation enthalpies are at the top of the range. This is evident for N_{Ni} ranging from 3 to 9 in Fig. 10.

A consequence of the neighbourhood dependence of the vacancy formation enthalpy is that as a function of temperature the relative contributions from the various neighbourhoods will change. Neglecting contributions from vibrations and other thermal excitations, a qualitative analysis of configurational thermal effects can be made. In the equiatomic alloy, one might expect at low temperature to find almost exclusively vacancies surrounded by Cu atoms only with a vacancy formation enthalpy of about 1 eV. As the temperature is raised, an increasing number of vacancies with more Ni atoms in the nearest neighbour shell will appear. At very high temperatures ultimately all vacancies can appear, but for combinatorial reasons, the ones with $N_{\mathrm{Ni}} \approx 6$ will be most abundant so that the 'apparent' vacancy formation enthalpy increases to about 1.3 eV. The effective vacancy formation enthalpy is thus temperature dependent, as envisaged half a century ago on the basis of a simple Braggs–Williams model [76]. Detailed calculations show that the total vacancy concentration, given by Eq. (104), behaves as if the vacancy formation energy is linearly increasing with temperature. Therefore, we consider it as an effective Gibbs energy of vacancy formation. The effective Gibbs energy of vacancy formation is associated with a *negative* effective vacancy formation entropy, at equiatomic composition amounting to about $-0.25\,k_{\mathrm{B}}$. Non-configurational excitations tend to give significantly positive contributions to the effective vacancy formation entropy, see, e.g. Table 2 in Ref. [77]. Therefore, the actual total vacancy formation entropy in alloys probably remains positive, but it is of interest to note that quite generally the configurational contribution in an alloy is negative.

4.4 Finite temperature properties DFT total energies provide defect properties at $T = 0\,\mathrm{K}$, but to be able to compare with experimental data finite temperature effects have to be taken into account. Defect properties at finite temperatures are nowadays mostly calculated in the context of empirical potentials. With increasing temperature, the phase space that atoms can probe is drastically enlarged and with it the computational demands. This is reflected by the history of vacancy calculations in metals performed using DFT: calculations on vacancies were limited to $T = 0\,\mathrm{K}$ throughout the 1990s [78–81]. With advances in hardware

technology, electronic and quasiharmonic (non-interacting phonons) contributions started to be included [82–84]. Yet, as of 2005 it was generally accepted that the full set of high-temperature excitations cannot be captured within first-principles calculations, because the computational costs would be prohibitive [85]. The extension of defect calculations to temperatures close to the melting point where pronounced anharmonic excitations become important has been demonstrated only very recently on a fully *ab initio* DFT level [86].

The methodology to derive accurate free energies as a function of volume and temperature for bulk systems *ab initio* is outlined in detail in Ref. [87] of this series. Here, we focus on the particular aspects that have to be taken into account to obtain the defect free energy. For defect calculations we have to keep in mind that the defect and the perfect bulk energies have to be treated on an equal footing, i.e. the inclusion of the same excitation mechanisms is necessary, in order to maximize the cancellation of errors.

4.4.1 Electronic excitations Calculating electronic contributions for defects is generally more demanding compared to perfect bulk systems since extensivity is not applicable. As a consequence, the entire supercell has to be calculated instead of just the primitive cell that would completely represent a bulk system. Nevertheless, the calculation of the electronic contribution to the free energy can nowadays be performed with moderate computational resources even for defects. The electronic contribution has a substantial effect on derivative quantities of bulk systems [88]. For the vacancy formation energy in aluminium the effect turned out to be small [86]. Unfortunately, however, this cannot be generalized, since other elements show significantly higher electronic contributions to the free energy than aluminium, which might therefore also affect defect formation. Known examples include Rh, Pd, Ir, Pt, Ca and Ti [77, 88].

4.4.2 Quasi-harmonic approximation Lattice vibrations contribute by far the largest entropy in well ordered and pure systems. Phonon calculations are usually performed in the *quasi-harmonic* approximation, which is based on treating the lattice vibrations as harmonic but volume-dependent.

An important difference between bulk and defect phonon calculations is the required accuracy: converging the free energy of bulk systems at the melting point is nowadays easily achieved with a precision of 0.1 meV per atom. For a defect supercell with 100 atoms calculated with the same precision the errors in formation energies would already be 10 meV. It immediately becomes clear that the demands on precision increase enormously when defect formation energies are considered. This also results in tighter convergence criteria (denser k-point mesh/larger energy cutoff) compared to simple bulk calculations. The problem is particularly acute in phonon calculations due to the increased sensitivity of free energies at elevated temperatures.

There are two well-established methods for calculating non-interacting phonons. The first is the 'direct' or 'direct force constant' method, where the dynamical matrix is built directly from force constants obtained in a corresponding energy calculation by displacing single atoms. The second is a perturbative approach in reciprocal space using linear response theory [89]. Both approaches have been found to yield phonon spectra of comparable precision and computational effort in bulk systems. The usual advantage of the linear response method for perfect crystals is that all calculations are made on only a primitive unit cell. However, for defect calculations this advantage disappears, since, as previously discussed, a large supercell has to be considered for both perfect and defective systems, and thus the direct method is generally used.

In practical calculations of the phonon contributions, the defect cells require special attention: due to the loss of translational symmetry introduced by the presence of the defect a Fourier interpolation cannot be applied to generate dense wave vector meshes in the Brillouin zone. A correction scheme to overcome this difficulty is presented in [86], which is based on the difference between Fourier interpolated and non-interpolated quasi-harmonic free energies of the bulk systems.

4.4.3 Anharmonic vibrational contributions In order to account for phonon–phonon interactions, methods going beyond the quasi-harmonic approximation have to be applied. A direct calculation of the free energy is possible in principle by molecular dynamics (MD) simulations, by integration of internal energy with respect to temperature. However, this would entail a prohibitively high computational cost [77, 90]. Fortunately, as discussed in Ref. [87] of this series, a method of thermodynamic integration can be used to calculate anharmonic contributions to the free energy.

To converge the anharmonic contribution to the *defect formation energy* to a similar accuracy as the contribution to *bulk free energies* is an enormous challenge. For all chosen temperature and volume points of the free energy surface, MD runs have to be performed. Let us suppose that we desire an accuracy of, say, 1 meV to sample the anharmonic contribution. As in the quasi-harmonic case we choose a defect cell containing 100 atoms. Our MD runs are performed for the bulk and defect system on a quasi-harmonic reference of similar quality. This also means that the standard deviation σ of the sampled energy differences is very similar in the bulk and defect case, $\sigma_{\mathrm{bulk}} \approx \sigma_{\mathrm{defect}}$. In a bulk calculation with 100 atoms extensivity holds and the standard error SE can be calculated as a per atom quantity. In the case of a defect formation energy, we are interested in energy differences of the two supercells (defect minus bulk) and the standard error will therefore be 100 times larger compared to the bulk free energy calculation. Using the definition of the standard error

$$SE := \frac{\sigma}{\sqrt{n}} \tag{105}$$

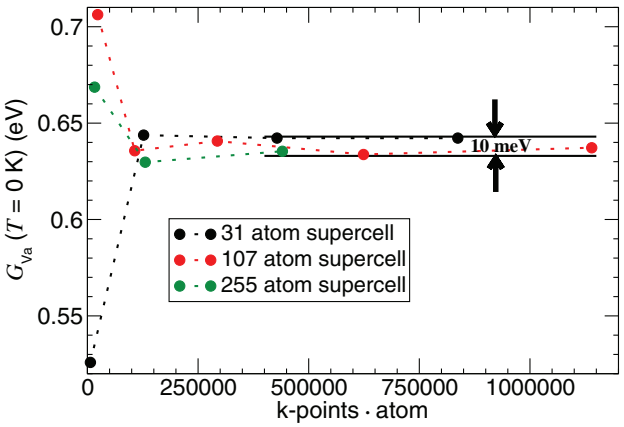

Figure 11 Vacancy formation energy in Al at $T = 0$ K (using the PBE XC functional). It can be seen that k-points × atom converges only at approximately 100 000.

with n being the number of time steps in the corresponding MD run, we see a dramatic effect on $n_{\text{defect}}/n_{\text{bulk}}$: to obtain the same accuracy in the defect formation energy as in the bulk free energy calculation we need a sampling time 10^4 times longer. Since converging the bulk free energy to 1 meV is already a difficult task, it becomes clear that to converge defect formation energies to a similar quality is a huge challenge. Nevertheless, for the *ab initio* anharmonic contribution to the vacancy formation energy of Al and Cu it has recently been shown that a standard error of 0.1 meV per atom can be achieved even in a supercell containing 108 atoms [91]. It was found that the anharmonic contribution to the Gibbs energy of formation is significant, and it changes the formation entropy by an order of magnitude.

4.4.4 Magnetic contributions The magnetic contributions to the free energy are discussed in detail by Körmann et al. (Ref. [5], this issue).

4.5 Precision and accuracy As discussed in Sections 4.1 and 4.2, the values obtained for defect formation energies can depend strongly on the actual computational setup. Here, by way of illustration we briefly discuss the dependence of the vacancy formation energy in aluminium on the k-point sampling. As noted in Sections 4.1 this is in principle simply a technical obstacle, with the possibility of systematic convergence by sufficiently long and diligent calcualtions. In practice, it turns out that the demands on the computer are simply impossibly large, and in some cases converged results are not currently feasible. In this section, we also review some published values for the vacancy formation energy in Al that have been calculated using DFT.

The influence of the k-point density on the vacancy formation energy in Al (at $T = 0$ K) is illustrated in Fig. 11. It can be seen that k-point×atom values below ≈ 100 000 can change the vacancy formation energy significantly: energy differences of more than 0.1 eV are found compared to the converged value. These findings are important in assessing the literature, since it has to be noted that this demanding k-point convergence was impossible to achieve in the pioneering years of defect calculations using DFT.

Calculations performed with different convergence criteria can easily show a spread of more than 0.2 eV in vacancy formation energies and when comparing literature values it is thus necessary to carefully analyze the reported details of the computation. It can further be seen that already for a 31 atom supercell the vacancy formation energy converges to within 10 meV provided a sufficient k-point sampling is used.

The importance of being aware of the convergence with respect to the controllable parameters as well as the influence of the non-controllable parameters is illustrated by the values listed in Table 3. Even though all values are obtained using DFT they spread over a range of 0.3 eV. Correction schemes to the error introduced by the XC functional due to the creation of an internal surface yield a slight improvement, but still the values range from 0.62 to 0.77 eV. As discussed above, such a spread in values can also be due to different levels of accuracy in the convergence of controllable parameters.

4.6 Combining insight from *ab initio* calculations and experiments The current model for G^f_{Va} in the vast majority of literature on point defects assumes $G^f_{\text{Va}}(T) = H^f_{\text{Va}} - TS^f_{\text{Va}}$ with temperature *independent* enthalpy and entropy of formation H^f_{Va} and S^f_{Va}. Inserting $G^f_{\text{Va}}(T)$ into the general expression for defect concentrations, $c = \exp(-G^f_{\text{Va}}(T)/k_{\text{B}}T)$, yields $y_{\text{Va}} = \exp(S^f_{\text{Va}}/k_{\text{B}})\exp(-H^f_{\text{Va}}/k_{\text{B}}T)$, which exhibits an Arrhenius behaviour with a slope equal to the vacancy formation enthalpy. Many well-documented cases are known where a distinct deviation from a linear formation energy is found [97–100]. One of the most thoroughly measured examples is Al [97] with considerably different H^f_{Va} values measured with differential dilatometry (cf. Section 5.1) and positron annihilation (cf. Section 5.2) in different temperature ranges. There are two non-exclusive possibilities to explain the situation: either the assumption of a linear Arrhenius model is wrong for monovacancies and correspondingly H^f_{Va} and S^f_{Va} are temperature dependent; or other defects contribute and therefore the experimentally measured concentrations are effective concentrations arising from several defect species.

In practice, the fraction of interstitials, tri-vacancies and other point defect species is expected to be negligible; the only candidate defect under consideration besides the monovacancy is the divacancy. A situation where the non-Arrhenius behaviour is explained by divacancy contributions is illustrated in Fig. 12. Although the constituent mono- and divacancy formation energies taken individually would predict perfect Arrhenius behaviour, their total Arrhenius plot is effectively curved and as such fits the experimental data.

Unfortunately, any experimental verification of this explanation appears to be impossible, since a separation of the total vacancy concentration into mono- and divacancies has so far not been feasible. One peculiarity of fitting experimental data to a mono-/divacancy model is that the divacancy entropy is often larger by a significant factor than

Table 3 Representative literature values of vacancy formation energies in Al obtained from *ab initio* calculations. The abbreviations are: XC, exchange correlation functional; LDA, local density approximation; GGA, generalized gradient approximation (PW91 from Refs. [29, 30], PBE from Ref. [31]), AM05 from Refs. [92, 93]; PW, planewaves; ps, pseudopotential approach; PAW, projector augmented wave method; FP-LMTO, full potential linearized muffin tin orbitals; V = rescaled volume approach, P = constant pressure approach, volOpt = volume optimized approach (see [77]); $G^f_{\mathrm{Va}}(T = 0\,\mathrm{K})$ = vacancy formation energy, 'surf. corr.' refers to correction schemes due to the internal surface created by the vacancy.

year	ref.	XC	DFT flavour	strain	$G^f_{\mathrm{Va}}(T = 0\,\mathrm{K})$	$G^f_{\mathrm{Va}}(T = 0\,\mathrm{K})$ + surf. corr.
1989	[25]	LDA	PWps	V	0.56	–
1993	[79]	LDA	FP-LMTO	V	0.57	–
1995	[94]	LDA	PWps	–	0.66	–
1997	[81]	LDA	PWps	P	0.66	–
2000	[33]	LDA	PWps	P	0.70	0.76
		GGA-PW91			0.54	0.69
2003	[95]	LDA	PWps	P	0.70	–
		GGA-PW91			0.55	–
2008	[96]	DA	PWps	P	0.64	0.70
		GGA-PBE			0.62	0.77
2009	[86]	LDA	PWps	volOpt	0.65	–
		GGA-PBE			0.58	–
2012	[36]	LDA	PW-PAW	P	0.57	0.63
		GGA-PBE			0.51	0.66
		GGA-PW91			0.43	0.62
		AM05			0.74	–

the corresponding monovacancy entropy, as can be seen in Table 4. Nevertheless, this model is frequently used to fit experimental data: for Al about 40% of vacancy-type defects at T_{m} have to be assumed to be divacancies [101] (18% in the recent study [97]). For Cu, a similar situation with 30% was anticipated [100] with the corresponding values for fitted Gibbs energies of formation being indicated in Fig. 12 by the green lines.

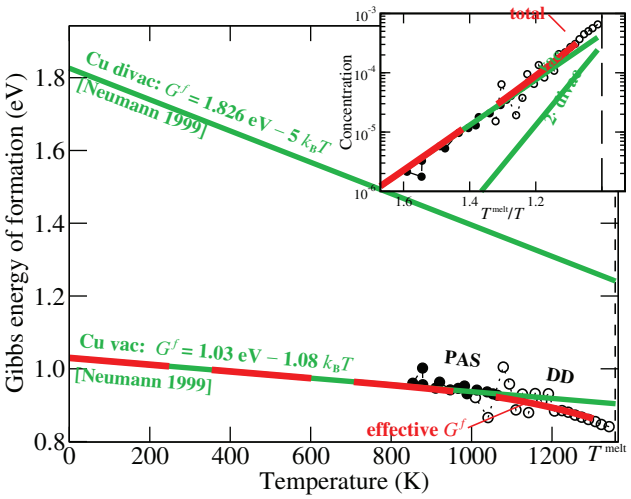

Figure 12 Mono-/divacancy situation for the temperature dependence of the Gibbs energy of vacant sites in Cu taken from [100]. Green solid lines show the Gibbs energy of formation, G^f_{Va}, and the concentration (inset) for mono- and divacancies with temperature-independent enthalpy and entropy of formation. The red dashed line is the effective G^f_{Va} obtained from the total concentrations.

Since it recently became possible to calculate anharmonic contributions of defects on a fully *ab initio* basis [86, 77], highly accurate DFT based calculations of mono- and divacancies in Cu and Al, including all relevant finite temperature excitations, have been performed [91]. A knowledge of the explicit temperature dependence allows us to exclude divacancies as a possible source of the non-Arrhenius behaviour. As shown for the monovacancy in Fig 13, even for these simple fcc metals anharmonicity is large and introduces an unexpectedly strong non-Arrhenius behaviour. With these results we are thus able to reconcile experimental and DFT data by bridging the enormous gap between $T = 0\,\mathrm{K}$ calculations and high temperature experimental data. The interpretation of the experimental measurements using DFT calculations clearly favours the temperature dependence of the entropy over a divacancy contribution to explain the non-Arrhenius behaviour. A more detailed discussion of experimental techniques and problems is given in the following section.

Table 4 Mono- and divacancy entropies of formation for Al and Cu. The factor = $S^{\mathrm{di}}_{\mathrm{Va}}/S^{\mathrm{mono}}_{\mathrm{Va}}$ illustrates the remarkably high divacancy entropies, which have to be assumed in a mono-/divacancy picture to fit the corresponding experimentally found curvature of G_{Va} as illustrated in Fig. 12.

element	$S^{\mathrm{mono}}_{\mathrm{Va}}$	$S^{\mathrm{di}}_{\mathrm{Va}}$	factor	refs.
Al	0.76	3.2	4.2	[98]
Cu	1.08	5	4.6	[100]
Cu	2.5	6.7	2.7	[102]
Cu	1.4	4.9	3.5	[103]

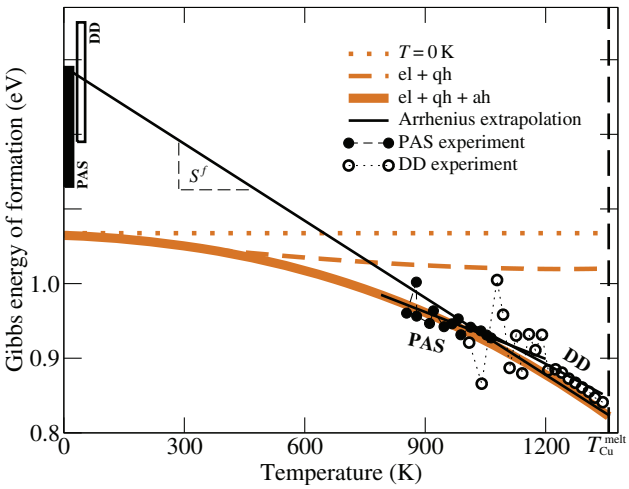

Figure 13 Experimental (black symbols) and DFT (orange lines) Gibbs energy of formation of mono vacancies in Cu. Experiments (PAS, positron annihilation spectroscopy [104]; DD, differential dilatometry [97]) are limited to a region close to the melting point, T_m. Extrapolations of the PAS/DD data to 0 K introduce scatter in the reported values (filled/empty black bar marks the corresponding interval). Common *ab initio* approximations such as the $T = 0$ K (dotted line) and the electronic plus quasiharmonic (el+qh; dashed line) approach are shown. The full curve (el + qh + ah) includes also anharmonicity. Figure adopted from Ref. [91].

5 Advances in experimental techniques Point defects are well known to play a central role in determining the kinetic behaviour of materials through their dominant influence on diffusive transport. Moreover, it is also well-established that there is an equilibrium concentration of intrinsic point defects such as vacancies (and vacancy complexes) and self-interstitials. At the same time, since the intrinsic point defects can be in equilibrium at high temperature, the measured value of thermodynamic properties such as heat capacity, expansion coefficient and compressibility already include the contribution from point defects. The challenge to modeling comes when the properties measured at high temperature are extended beyond the measurement range. In this case, in order to properly develop physically realistic models it is essential to develop quantitative data on the separate contribution of point defects to the thermodynamic properties. In turn this requires the direct determination of point defect concentrations as a function of temperature.

Many experimental efforts have been directed to measure the point defect concentrations. Since the point defect concentrations are small (i.e. typically 0.01% or less even at the melting temperature) the measurements are focused on high temperatures to maximize the concentration. Most often a given physical property such as the electrical resistivity is used for measurement, see, e.g. [105]. For example samples are quenched from different high temperatures that will yield the corresponding values for the excess resistivity compared to a baseline temperature. These methods can yield a determination of the defect formation enthalpy, but without separate knowledge of the property change per defect, the defect concentration is not available. In addition, during the quenching process some defect migration and clustering is possible and can complicate the analysis. In order to circumvent these difficulties, it has been well known that equilibrium measurements at high temperature are necessary to provide the most reliable data on the equilibrium point defects. Equilibrium measurements are less sensitive to the microstructure and prior history of samples than non-equilibrium measurements. In this case the change in a physical property is caused only by defects whose concentration reversibly changes with temperature.

There are two additional important issues. The computational modeling requires input of reliable data concerning the absolute defect concentration, if one addresses their effect on thermodynamic properties. Thus, the defect concentrations have to be carefully measured, especially as a function of temperature. On the other hand, the direct measurements of the formation or migration enthalpies without a direct access to the defect concentrations – via some indirect techniques – might also be helpful in thermodynamic modeling.

One of most complete overviews of experimental techniques and results with respect to the measurements of point defect concentrations in metals was published by Kraftmakher [106, 107]. We are addressing the reader to these publications for specific details.

The information about point defects, e.g. vacancies and interstitials in pure metals, and their formation and migration can experimentally be determined basically using any physical property that is proportional to the defect concentration. In fact, a large number of such techniques were used in the past, such as resistivity measurements [105], neutron scattering [108], calorimetric measurements [109], or Mößbauer spectroscopy [110], to mention just a few. All these are indirect techniques and the contribution of point defects to the measured property has to be precisely known.

According to Seeger [111], only two direct methods are sensitive to the vacancies at thermal equilibrium, namely differential dilatometry and positron annihilation.

5.1 Differential dilatometry In the classic studies, first by Feder and Nowick [112] and later by Simmons and Balluffi [113], the equilibrium measurement approach was employed together with a comparison between the sample length change and the lattice parameter change to establish that the dominant point defects in metals were of a vacancy type. The basic concept behind the absolute measurement is that the expansion of a macroscopic volume of a sample should increase more rapidly than the unit cell volume as a function of temperature in the case of vacancy formation. In this case the key component of the experimental design was centered upon making two separate measurements on the same sample at the same time, Fig. 14. From measurements of the length and lattice parameter change for a cubic metal as shown in Fig 14, the point defect concentration is given

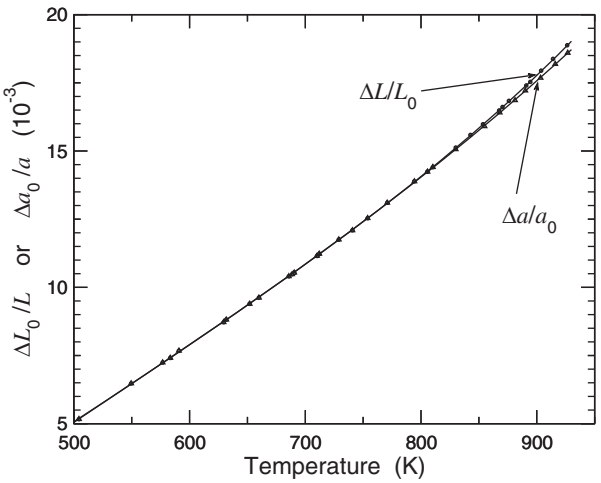

Figure 14 Relative changes of the sample length, $\Delta L/L_0$, and of the lattice parameter, $\Delta a/a_0$, of Al versus temperature after Simmons and Balluffi [113]. $\Delta L = L - L_0$ and $\Delta a = a - a_0$ where L_0 and a_0 are the corresponding values at 297 K.

by

$$y_{\mathrm{Va}}(T) - y_{\mathrm{i}}(T) = 3\left(\frac{\Delta L(T)}{L_0} - \frac{\Delta a(T)}{a_0}\right). \qquad (106)$$

Here L is the macroscopic length, a is the lattice parameter, T is the absolute temperature, L_0 and a_0 are the values at a low reference temperature. y_{Va} and y_{i} are the vacancy and interstitial concentrations. In metals the self-interstitial concentration is typically negligible due to their high formation energy. It is useful to note that the result of Eq. (106) does not depend on the state of aggregation of the point defects.

Since these pioneering works, the differential dilatometry method was significantly improved in sensitivity and accuracy. The recent data of Hehenkamp's group from Göttingen [114–118] represent the most accurate data measured so far. Recent progress in the dilatometry techniques and especially the application of *relaxation* dilatometry was documented in Ref. [119].

Some relevant data of absolute measurements of vacancy concentration in several pure metals are given in Table 5. It is seen that, for pure metals, the formation enthalpy of vacancies can be considered as a well-determined property, whereas the formation entropy (and thus the absolute vacancy concentration) has considerably larger uncertainty. This is related to the fact that measuring the given quantity over an extended temperature interval allows one to reduce the uncertainty of determination of the corresponding formation enthalpy (which might be an effective one for the given temperature range), while the formation entropy is a result of data extrapolation to $1/T \to 0$ that is prone to large uncertainties.

This uncertainty might provide conflicting explanations in terms of the vacancy-type defects, as discussed in the preceding Section. For example divacancies were discussed, see, e.g. [98], to be present in pure fcc metals near the melting temperature in significant amounts. Alternatively, recent measurements of the Göttingen group [116–118] substantiate that the divacancy concentrations remain low and no distinct contribution could be established in pure metals even at the melting point. These conclusions are strongly supported by the recent DFT calculations of the anharmonic contributions to the defect densities in Al [86], discussed above in Section 4.6.

Table 5 The absolute vacancy concentrations at melting point, y^0_{Va}, and determined enthalpies, H^f_{Va}, and entropies, S^f_{Va}, of vacancy formation of several fcc metals.

metal	y^0_{Va} (10^{-4})	H^f_{Va} (eV)	S^f_{Va} (k_{B})	Ref.
Al	8.8	0.76	2.4	[113]
	9.6	0.81	3.1	[120]
	9.2	0.75	2.3	[116]
Ag	1.7	1.09	1.5	[113]
	5.2	1.05	2.3	[115]
Au	7.2	0.94	1.0	[121]
	9.5	0.95	1.3	[117]
Cu	2.1	1.18	1.6	[122]
	7.6	1.19	3.0	[114]
Ni	7.8	1.56	3.3	[118]

5.2 Positron annihilation This technique provides accurate data on equilibrium vacancy concentrations in an interval of concentrations from 10^{-7} to 10^{-4} [123] that is complementary to the differential dilatometry data. The advantage of the positron annihilation method is that it is sensitive to vacancy-type defects and contributions of vacancy clusters can be separated by the lifetime of positrons. The situation might become complicated if other defects are present, such as dislocations and grain boundaries. Then the correct treatment of the positron annihilation data depends on the validity of the trapping model [124].

An Arrhenius plot of high-temperature positron trapping rates deduced from the positron lifetime spectroscopy data [123] is examplified in Fig. 15 for several fcc metals. The derived formation enthalpies are listed in Table 6.

5.3 Charged defects Point defects in pure semiconductors or dielectrics are additionally characterized by their charge state, and their concentration is extremely sensitive to the impurity content. Differential dilatometry measurements for pure Si for example revealed that the difference between temperature-induced lattice expansion and length changes was about zero. Those data were considered as a strong hint to the high concentration of self-interstitials in Si.

The charge state of the point defects is typically measured by spreading resistance [125] or scanning probe capacitance [126] techniques. Recently, the measurements of self- and impurity diffusion in isotopically modulated structures [127] were proposed as a powerful source of data on the intrinsic/extrinsic properties of charged defects. As an example, Fig. 16 represents the results of modeling of diffusion experiments on self- and dopant diffusion in

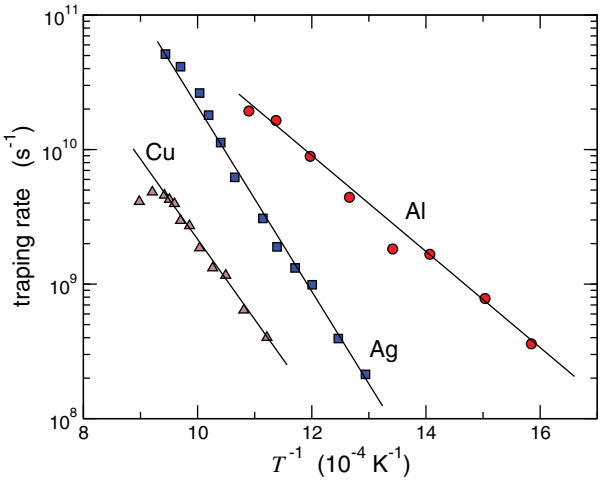

Figure 15 High-temperature positron trapping rates derived from positron lifetime spectroscopy versus inverse temperature for several fcc metals, after Schaefer [123].

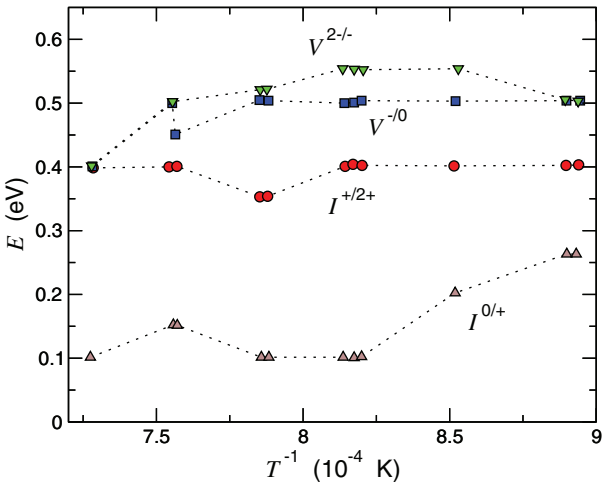

Figure 16 Energy-level positions of vacancies V and self-interstitials I within the band gap of Si for temperatures between 850 and 1100 °C after [128]. The symbols represent the results from modeling the simultaneous self- and dopant diffusion in isotopically modulated silicon [128]. The CEF notation is used here.

isotopically modulated Si grown as stacking of successive layers Si^{nat}/Si^{28} [128]. The charge state of point defects could be carefully determined as a function of temperature both for intrinsic and extrinsic conditions, see also Section 3.

5.4 Calorimetric data While it is now well established that the predominant point defect in fcc metals at temperatures near the melting point is the vacancy, it is also evident that the Arrhenius plot of y_{Va} versus $1/T$ shows a curvature. As discussed in Sections 5.1 and 4.6, two possible explanations for this behaviour have been advanced based upon a divacancy contribution at high temperature [98] or due to anharmonicity effects without divacancies [33, 95, 138, 139, 86, 36]. In the absence of definitive evidence on the possible role of divacancies this remains an open issue.

Table 6 Comparison of the value of vacancy formation enthalpy, H^f_{Va}, determined by different methods for several fcc metals – differential dilatometry (DD); estimations from the nonlinear increase in the heat capacity (HC); positron annihilation (PA); and resistivity measurements (R).

metal	H^f_{Va} (eV)	method	Ref.
Al	0.75	DD	[116]
	0.79	HC	[129]
	0.77	R	[130]
	0.68	PA	[131]
Cu	1.19	DD	[114]
	1.05	HC	[132]
	1.3	R	[133]
	1.0	R	[134]
	1.13	PA	[123]
Ni	1.56	DD	[118]
	1.4	HC	[135]
	1.6	R	[136]
	1.63	PA	[137]

Based upon several *ab initio* calculations it is evident that anharmonic effects are important at high temperatures. At the same time, measurements of the molar heat capacity, C_p clearly reveal a rising value upon approaching the melting point [140]. However, the *ab initio* results indicate that explicit anharmonicity, i.e. anharmonicity beyond the quasi-harmonic, gives rise to a negative contribution to C_p. Thus, it was concluded that only the vacancies which the calculations indicate have a significant anharmonic contribution to the entropy of formation can account for the positive increase in C_p as observed experimentally [86]. In this case the vacancy contribution to C_p can be represented as follows [106]:

$$\Delta H = N_A H^f_{Va} y_{Va}, \tag{107}$$

$$\Delta C_p^{Va} = \left(\frac{d\Delta H}{dT}\right)_p = \frac{N_A (H^f_{Va})^2 A}{k_B T^2} \exp\left(-\frac{H^f_{Va}}{k_B T}\right). \tag{108}$$

Where ΔH is the extra enthalpy, N_A is the Avogadro number, A is $\exp(S^f_{Va}/k_B)$ and H^f_{Va} and S^f_{Va} are (generally) temperature independent.

Accounting for a separate contribution of vacancies to ΔC_p, Eq. (108), we simply represent the vacancy contribution as additive since the vacancy concentration is small,

$$C_p = C_p^{lattice} + \Delta C_p^{Va}. \tag{109}$$

The point defect contributions to C_p values near corresponding melting points are exemplified in Fig. 17 for a number of metals. The rising trend of specific heat with increasing temperature in Fig. 17 yields large values that are of the order of the lattice specific heat at T_m. Since the

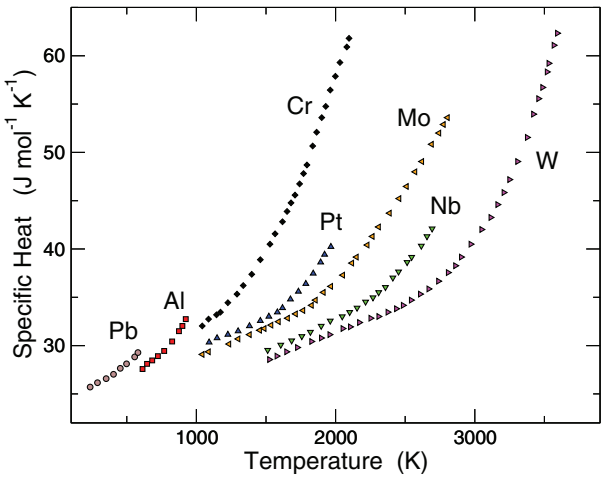

Figure 17 Specific heat (heat capacity) of metals as a function of temperature. The measurements were performed by adiabatic calorimetry (Pb and Al [129]), modulation calorimetry (W [141] and Pt [142]), drop method (Cr [143]); and by dynamic calorimetry (Mo [144] and Nb [145]).

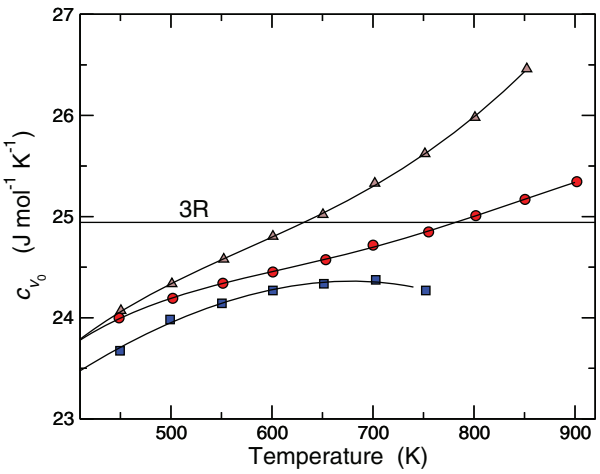

Figure 18 Reduced heat capacity of Al at 0 K volume (C_{V_0}) against temperature after Shukla et al. [140] derived based of the data of Brooks and Bingham [146] (triangles up), Ditmars at el. [147] (circles), and Leadbetter [148] (squares).

C_p data has been reported in several independent studies involving different experimental techniques and includes both low melting point and high melting point metals, it appears that the C_p behaviour is real. However, in this case the analysis of the data in Fig. 17 in terms of Eq. (108) with a temperature independent entropy term yields H^f_{Va} values that are 1.5–2.0 greater than the values determined by other more direct measurements such as dilatometry and PAS. This suggests that the upward trend of C_p indicated by the separate C_p measurements may not be due entirely to excess vacancies. A similar conclusion has been mentioned in the literature [144–146] as the required vacancy concentration approaches several percent, which seems to be unreasonable. Instead, the C_p trend may be due to a temperature dependent entropy as indicated by the DFT calculations in Section 4.6, but this issue warrants further study to resolve the origin of the high temperature C_p behaviour.

One of the ways to compare the calorimetric experimental data with theoretical estimates is to derive from the experimentally measured C_p values the heat capacity at constant volume reduced to the volume at $T = 0$ K [140]. An example of such recalculation of the experimental data after Shukla et al. [140] is given in Fig. 18, where the heat capacity at the 0 K volume, $C_{V_0} = T(dS_{V_0}/dT)$, is plotted and compared to the $3R$ value.

5.5 Summary The recent advances in experimental techniques and measurement methods based upon differential dilatometry and positron annihilation coupled with the need for accurate data on point defects calls for a renewed effort to extend the limited existing database beyond that for a few pure elements and an even smaller number of alloys. Indeed, as pointed out by Hehenkamp [97], the comparison of absolute measurement data with positron measurements allows for the assessment of vacancy–solute binding. As a result, the determination of point-defect behaviour can serve as an effective probe of alloy solution behaviour and provide a valuable test of CALPHAD models.

6 Conclusions and outlook In this paper, we have surveyed from several viewpoints the theory of point defects in unary systems, as well as alloys and compounds. We have included a discussion of the special problems associated with dielectric and semiconducting materials in which the defects may carry a charge. We have also discussed the treatment of random alloys, which present an ensemble of local environments in which vacancies may reside, such that the vacancy concentration is distributed more uniformly the higher the temperature, giving rise to effective vacancy formation energies that increase with temperature.

A major aim has been to bring together the perspective of CALPHAD with the statistical thermodynamics and atomistic simulation practised by many physicists, chemists and materials scientists. The point here is that the CALPHAD community takes a very broad view of materials and phases, and is concerned in particular with multicomponent phases and the prediction of their stability over the entire range of composition. On the other hand, the approaches of statistical thermodynamics, which we have characterized as *defect-centric*, generally take as their focus a perfect material in which vacancies, interstitials or impurity atoms, remain in low concentrations. Some care has been taken here to describe with a unified notation and nomenclature the methods based on the canonical and grand-canonical distributions, in order to make clear the common ground and the differences between them. Following our discussion of these defect-centric methods, we have maintained the same notation in our presentation of the CEF, widely used in the CALPHAD community, in order to make the methods of all communities transparent to all readers. The examples of NiAl and GaN serve to illustrate the concepts. Thus, we are

able to show how the same calculated defect energies can be used either within the machinery of CALPHAD or as input to canonical or grand-canonical methods for calculating stoichiometry variations and chemical potentials. This clarification of the role of input data across the disciplines has become a matter of urgency, since the field of phase diagram calculation is now well placed to benefit from the recent capabilities of *ab initio* calculation, particularly DFT, in order to implement the energetics of defects at high temperature. One obstacle to progress, which we hope to have eliminated with this paper, has been the lack of a common language for communication, even for a concept as apparently simple as defect formation energy.

A potential area of difficulty and misunderstanding for the *ab initio* community, concerns the requirement of CALPHAD for values of the free energies of end-member compounds. If these happen to be metastable, it is feasible to calculate their free energies in the manner discussed elsewhere in this volume, for example with DFT. However, should the end-member include a sublattice of 100% vacant sites, it is quite likely that it will be unstable, in which case no reasonable Gibbs energy can be calculated for it. This should not be a problem in practice. However, one needs to be aware that the extrapolation of a Gibbs energy towards this unstable end-member requires a more-or-less arbitrary function, whose role it is to penalize the phase with a sufficiently high Gibbs energy to ensure that it cannot enter the phase diagram when the proportion of the unstable end-member becomes high. As we have demonstrated, the provision of such a penalty function is also the role of the 'Gibbs energies' of the end-members representing pure vacancies. This use of the concept of unstable phases, including phases consisting partly or entirely of vacancies, is naturally unfamiliar to most of the practitioners of *ab initio* and atomistic simulation, and we hope that the present article will bridge this particular gap in understanding.

We have devoted the last two sections to reviewing the status of data on vacancy formation energies, both theoretical and experimental. Considerable progress has been made recently in the methodology of calculating vacancy formation energies at high temperature within the DFT framework, and we show results for Cu and Al which illustrate the significant effect that anharmonic lattice vibrations have in these systems on the calculated Gibbs energies of vacancy formation. This revitalizes the debate on the role of anharmonicity versus the role of divacancies in contributing to specific heats, a topic which is revisited in our review of the experimental situation. Although not conclusive, the evidence from DFT calculations now strongly favours the theory that anharmonicity causes the observed curvature in Arrhenius plots rather than divacancies. It remains to be seen how important the role of anharmonicity will prove to be in the defect concentrations exhibited by other elements and compounds.

Appendix A

Derivation of partial molar Gibbs energies We present here a derivation of the general expression for the partial Gibbs energies of constituents in a phase consisting of several sublattices, each with a fixed number of sites. This is a generalization of the original derivation of Sundman [152]. The standard definition of the partial Gibbs energy of a constituent labelled i is :

$$G_i = \left(\frac{\partial G}{\partial N_i}\right)_{T,P,\{N_{j\neq i}\}} . \tag{A.1}$$

These partial Gibbs energies become identical to chemical potentials when the phase concerned is in equilibrium, in the sense that the compositions of its several sublattices are equilibrated. However, it is important to note that if a constituent may be distributed over several sublattices there are degrees of freedom in the composition of the sublattices involved. Therefore the Gibbs energy of the phase has to be minimized with respect to these degrees of freedom before the partial Gibbs energy of a component can be identified with its chemical potential. However, Eq. (A.1) is inconvenient, because the quantity we normally work with is not the total Gibbs energy but the Gibbs energy per mole (or per formula unit), $G_{\rm m}$, and the variables we work with are not the total numbers of each constituent but the fractional occupancies of the sublattices. The purpose of this appendix is to show for a general phase, specified within the CEF, how we can define Eq. (A.1) in terms of derivatives of $G_{\rm m}$ with respect to the sublattice occupancies of each consituent, special cases of which were used in this article.

For generality, we treat a phase containing an arbitrary number of sublattices labelled s. Each sublattice is represented by a number a_s sites within a formula unit of the phase. Now within the CEF we choose the independent variables that specify the composition and quantity of this phase to be the numbers of formula units of the end-members, which we denote N_J. The suffix J ranges over the possible number of 'compounds' that can be constructed by filling each sublattice with a single constituent. To fix ideas, in the model $(A, B, C)_3(A, B, C)$, for example $a_1 = 3$, $a_2 = 1$ and there are 9 independent variables $N_J = \{N_1, N_2, \cdots N_9\}$.

To keep things general, let us refer to the set of end-members as E. A subset of the end-members having component i on sublattice s can be referred to as $E_i^{(s)}$. We also denote the constituent in sublattice s of an end-member J and $J^{(s)}$,

The number of formula units of each end-member has to sum to the total number of formula units N of the phase:

$$\sum_J N_J = N. \tag{A.2}$$

In terms of the N_J we can calculate the constituent fractions $y_i^{(s)}$:

$$y_i^{(s)} = \frac{\sum_{J\in E_i^{(s)}} N_J}{\sum_J N_J} . \tag{A.3}$$

These fractions must sum to unity on each sublattice,

$$\sum_i y_i^{(s)} = 1. \tag{A.4}$$

We shall need the partial derivatives of $y_i^{(s)}$ with respect to N_J, for which, as we see by inspection of Eqs. (A.2) and (A.3) there are two mutually exclusive cases, namely the positive result when the end-member J has constituent i in sublattice s:

$$\frac{\partial y_i^{(s)}}{\partial N_J} = \frac{1 - y_i^{(s)}}{N} \quad \text{if } J \in E_i^{(s)}, \tag{A.5}$$

and the negative result when J does not have constituent i in sublattice s:

$$\frac{\partial y_i^{(s)}}{\partial N_J} = -\frac{y_i^{(s)}}{N} \quad \text{if } J \notin E_i^{(s)}. \tag{A.6}$$

We can now proceed to take derivates of the total Gibbs energy with respect to the number of formula units of each independent end-member. We write the total Gibbs energy as

$$G = NG_{\mathrm{m}}, \tag{A.7}$$

where G_{m} depends only on the variables $\{y_i^{(s)}\}$, and differentiate with respect to N_I, with all derivatives at constant T and P, and find:

$$\begin{aligned} G_I &= G_{\mathrm{m}} + N \sum_s \sum_i \left(\frac{\partial G_{\mathrm{m}}}{\partial y_i^{(s)}} \right)_{y_j^{(t)}, \{i,s\} \neq \{j,t\}} \times \frac{\partial y_i^{(s)}}{\partial N_I} \\ &= G_{\mathrm{m}} + \sum_s \sum_i (\delta_{iI^{(s)}} - y_i^{(s)}) \left(\frac{\partial G_{\mathrm{m}}}{\partial y_i^{(s)}} \right)_{y_j^{(t)}, \{i,s\} \neq \{j,t\}} \end{aligned}$$

where $\delta_{iI^{(s)}}$ is unity when i is the constituent in sublattice s of the end-member I and zero otherwise.

Appendix B

Glossary Note that where there is no danger of confusion the same symbols have sometimes been used for quantities per mole and per formula unit, as mentioned in Section 2.2. The only difference to the thermodynamic formulae will be whether the gas constant R or Boltzmann's constant k_{B} appear.

BOP bond-order potential

CEF compound energy formalism

ΔC_p^{Va} vacancy contribution to the molar heat capacity C_p

DFT density functional theory

DMC diffusion quantum Monte Carlo

EAM embedded atom method

GAP Gaussian approximation potential

GGA generalized-gradient approximation

${}^{E}G_{\mathrm{m}}$ the excess Gibbs energy per formula unit, with reference to ideal solution behaviour

G the total Gibbs energy of a system of N formula units

G_{A} the Gibbs energy per atom of a hypothetical defect-free crystal of A; also used in the CEF to denote the *partial* Gibbs energy of a species A

${}^{\circ}G_{\mathrm{A:B}}$ the Gibbs energy of formation of a hypothetical defect-free compound of A and B, used as an 'end-member' in the CEF

$G_{\mathrm{A:C}}$ the partial Gibbs energy of the end-member A:C in the CEF

G_{m} the total Gibbs energy per formula unit

$G_{1\mathrm{Va}}$ the Gibbs energy of N atoms of crystal of A containing a single defect at a known lattice site

G_i^a energy attributed to a site of sublattice a when occupied by species i

G_{Va}^f the vacancy formation energy

$\mathcal{G}$ the grand potential of a system of N formula units (Legendre transform of G with respect to quantities of components)

$\mathcal{G}_\delta$ formation energy of defect δ expressed in terms of chemical potentials, as derived within the grand canonical formalism

g_{A}^b a 'raw defect energy' calculated for species A on sublattice b

H_{Va}^f vacancy formation enthalpy

k_s equilibrium constant for antisite pair formation in a binary ordered alloy

k_t equilibrium constant for triple defect formation in a binary ordered alloy

k_v equilibrium constant for vacancy pair formation in a binary ordered alloy

$L_{\mathrm{A,B...}}$ interaction parameters used in the CEF

MD molecular dynamics

N number of lattice sites

NN neural network

n_i number of species i (atoms or vacancies)

PAS Positron annihilation spectroscopy

QMC quantum Monte Carlo

S_{mix} entropy of mixing in an ideal solution

S_{Va}^f vacancy formation entropy

SE standard error

SQS special quasi-random structure

TB tight binding

XC exchange and correlation

y_{A}^b fraction of constituent A on sublattice b

$y_i^{(s)}$ similar to above: fraction of constituent i on sublattice s

y_{Va} vacancy fraction, measured as occupancy per lattice site

y^*_{Va} equilibrium vacancy fraction, measured as occupancy per lattice site

λ_i one of the Lagrange multipliers for minimization of the Gibbs energy, constraints are labelled $i = A, B, a$ and b

μ_i chemical potential of species i

Acknowledgements We are grateful to all participants of the Ringberg meeting for discussions. S.S. and J.R. acknowledge financial support through ThyssenKrupp AG, Bayer MaterialScience AG, Salzgitter Mannesmann Forschung GmbH, Robert Bosch GmbH, Benteler Stahl/Rohr GmbH, Bayer Technology Services GmbH, the state of North-Rhine Westphalia, the European Commission in the framework of the ERDF and the German Research Foundation (DFG) through project C2 of the collaborative research center SFB/TR 103. MWF would like to thank the EPSRC Program Grant [grant number: EP/K008749/1] Material Systems for Extreme Environments (XMat) for financial support. S.V.D. acknowledges financial support by the German Research Foundation (DFG) through research Grant DI 1419/3-2. J.H.P. is grateful for the support of the US Army Research Office (W911NF-12-1-0010). A.G. and J.N. gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (DFG) within the joint project PAK461. B.S. acknowledges the Humboldt Stiftung for a senior research award.

References

[1] H. B. Callen, Thermodynamics and an Introduction to Thermostatistics, second ed. (John Wiley and Sons, New York, 1985).
[2] R. DeHoff, Thermodynamics in Materials Science, second ed. (McGraw-Hill, Singapore, 2006).
[3] H. L. Lukas, S. G. Fries, and B. Sundman, Computational Thermodynamics (Cambridge University Press, Cambridge, 2007).
[4] M. Hillert, Physica B **103**, 31 (1981).
[5] F. Körmann, A. Breidi, S. L. Dudarev, N. Dupin, G. Ghosh, T. Hickel, P. Korzhavyi, J. Munoz, and I. Ohnuma, Phys. Status Solidi B **251**, 53–80 (2014), this issue.
[6] C. Wagner and W. Schottky, Z. Phys. Chem. B **11**, 163 (1930).
[7] M. Hagen and M. W. Finnis, Philos. Mag. A **77**, 447 (1998).
[8] P. A. Korzhavyi, A. V. Ruban, A. Y. Lozovoi, Y. K. Vekilov, and I. A. Abrikosov, Phys. Rev. B **61**, 6003 (2000).
[9] Y. Mishin and C. Herzig, Acta Mater. **48**, 589 (2000).
[10] Y. Mishin, M. J. Mehl, and D. A. Papaconstantopoulos, Phys. Rev. B **65**, 224114 (2002).
[11] G. P. Purja Pun and Y. Mishin, Philos. Mag. **89**, 3245 (2009).
[12] J. Mayer, C. Elsässer, and M. Fähnle, Phys. Status Solidi B **191**, 283 (1995).
[13] J. Mayer and M. Fähnle, Acta Mater. **45**, 2207 (1997).
[14] S. B. Zhang and J. E. Northrup, Phys. Rev. Lett. **67**(17), 2339 (1991).
[15] N. Dupin, Assessment of Al–Ni, unpublished.
[16] P. Gustafson, Scand. J. Metall. **14**, 259 (1985).
[17] M. H. F. Sluiter, Calphad **30**, 357 (2006).
[18] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).
[19] C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).
[20] N. D. M. Hine, K. Frensch, W. M. C. Foulkes, and M. W. Finnis, Phys. Rev. B **79**, 024112 (2009).
[21] C. G. Van de Walle and J. Neugebauer, J. Appl. Phys. **95**, 3851 (2004).
[22] J. Neugebauer and T. Hickel, WIREs Comput. Mol. Sci. **3**, 438 (2013).
[23] F. A. Kröger and H. J. Vink, Solid State Physics (Academic Press, New York, 1958).
[24] B. Sundman, C. Guéneau, and N. Dupin, Acta Mater. **59**, 6039 (2011).
[25] M. J. Gillan, J. Phys.: Condens. Matter **1**, 689 (1989).
[26] D. J. Singh, Planewaves, Pseudopotentials and the LAPW Method (Kluwer, Boston, 1994).
[27] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[28] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[29] J. P. Perdew, J. Chevary, S. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
[30] J. P. Perdew, J. Chevary, S. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **48**, 4978 (1993).
[31] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); Erratum – Phys. Rev. Lett. **78**, 1396 (1997).
[32] S. Kurth, J. P. Perdew, and P. Blaha, Int. J. Quantum Chem. **75**, 889 (1999).
[33] K. Carling, G. Wahnström, T. R. Mattsson, A. E. Mattsson, N. Sandberg, and G. Grimvall, Phys. Rev. Lett. **85**, 3862 (2000).
[34] T. R. Mattsson and A. E. Mattsson, Phys. Rev. B **66**, 214110 (2002).
[35] L. Delczeg, E. K. Delczeg-Czirjak, B. Johansson, and L. Vitos, Phys. Rev. B **80**, 205121 (2009).
[36] R. Nazarov, T. Hickel, and J. Neugebauer, Phys. Rev. B **85**, 144118 (2012).
[37] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[38] W. D. Parker, J. W. Wilkins, and R. G. Hennig, Phys. Status Solidi B **248**, 267 (2011).
[39] R. Q. Hood, P. R. C. Kent, and F. A. Reboredo, Phys. Rev. B **85**, 134109 (2012).
[40] A. Badinski, P. D. Haynes, J. R. Trail, and R. J. Needs, J. Phys.: Condens. Matter **22**, 074202 (2010).
[41] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[42] M. S. Daw and M. I. Baskes, Phys. Rev. B **29**, 6443 (1984).
[43] Y. Mishin, M. J. Mehl, D. A. Papaconstantopoulos, A. F. Voter, and J. D. Kress, Phys. Rev. B **63**, 224106 (2001).
[44] M. Y. D. Farkas, M. J. Mehl, and D. A. Papaconstantopoulos, Phys. Rev. B **59**, 3393 (1999).
[45] P. M. Derlet, R. Høier, R. Holmestad, K. Marthinsen, and N. Ryum, J. Phys.: Condens. Matter **11**, 3663 (1999).
[46] Q. Zhang, T. Çağin, A. van Duin, W. A. Goddard III, Y. Qi, and L. G. Hector, Jr., Phys. Rev. B **69**, 045423 (2004).
[47] M. W. Finnis, Interatomic Forces in Condensed Matter (Oxford University Press, Oxford, 2003).
[48] F. Cleri and V. Rosato, Phys. Rev. B **48**, 22 (1993).
[49] M. J. Mehl and D. A. Papaconstantopoulos, Phys. Rev. B **54**, 4519 (1996).
[50] G. Liu, D. Nguyen-Manh, B. G. Liu, and D. G. Pettifor, Phys. Rev. B **71**, 174115 (2005).
[51] T. Hammerschmidt, R. Drautz, and D. G. Pettifor, Int. J. Mater. Res. **100**, 1479 (2009).
[52] M. Mrovec, R. Gröger, A. G. Bailey, D. Nguyen-Manh, C. Elsässer, and V. Vitek, Phys. Rev. B **75**, 104119 (2007).

[53] M. Čák, T. Hammerschmidt, and R. Drautz, J. Phys.: Condens. Matter **25**, 265002 (2013).
[54] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
[55] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[56] J. Behler, Phys. Chem. Chem. Phys. **13**, 17930 (2011).
[57] N. Artrith and J. Behler, Phys. Rev. B **85**, 045439 (2012).
[58] Y. Mishin, M. Sorensen, and A. Voter, Philos. Mag. A **81**, 2591 (2001).
[59] Calculations have been performed using the projector augmented-wave method as implemented in the VASP code [149–151]. For the exchange–correlation functional the generalized gradient approximation (GGA, PBE) [31] was used. For the energy vs. volume curves the energy cutoff for the plane waves was set to 250 eV and the Brillouin zone integration was performed using the Monkhorst–Pack scheme with k-point meshes of [20 × 20 × 20], [10 × 10 × 10], [7 × 7 × 7], and [5 × 5 × 5], respectively. To determine the difference in vacancy formation energy at constant volume and zero pressure an energy cutoff of 350 eV was used and k-point meshes of [33 × 33 × 33], [17 × 17 × 17], [11 × 11 × 11], and [9 × 9 × 9], respectively.
[60] G. Stocks, W. Temmerman, and B. Gyorffy, Phys. Rev. Lett. **41**, 339 (1978).
[61] H. Akai, J. Phys. Soc. Jpn. **51**, 468 (1982).
[62] H. Winter and G. Stocks, Phys. Rev. B **27**, 882 (1983).
[63] J. Sanchez, F. Ducastelle, and D. Gratias, Physica A **128**, 334 (1984).
[64] F. Lechermann and M. Fähnle, Phys. Rev. B **63**, 012104 (2000).
[65] A. Van der Ven and G. Ceder, Phys. Rev. B **71**, 054102 (2005).
[66] A. Van der Ven and G. Ceder, Phys. Rev. Lett. **94**, 045901 (2005).
[67] A. Van der Ven, H. C. Yu, G. Ceder, and K. Thornton, Prog. Mater. Sci. **55**, 61 (2010).
[68] M. Muzyk, D. Nguyen-Manh, K. J. Kurzydłowski, N. L. Baluc, and S. L. Dudarev, Phys. Rev. B **84**, 104115 (2011).
[69] E. del Rio, J. M. Sampedro, H. Dogo, M. J. Caturla, M. Caro, A. Caro, and J. M. Perlado, J. Nucl. Mater. **408**, 18 (2011).
[70] A. Zunger, S. H. Wei, L. G. Ferreira, and J. Bernard, Phys. Rev. Lett. **65**, 353 (1990).
[71] B. L. Zhao, R. Najafabadi, and D. J. Srolovitz, Philos. Mag. A **70**, 519 (1994).
[72] L. Delczeg, B. Johansson, and L. Vitos, Phys. Rev. B **85**, 174101 (2012).
[73] P. C. Clapp, Phys. Rev. B **4**, 255 (1971).
[74] M. H. F. Sluiter and Y. Kawazoe, Europhys. Lett. **57**, 526 (2002).
[75] Generalized gradient approximation [31] projector augmented wave pseudopotentials as implemented in VASP [149–151] were used with collinear spin polarization. All structures are fully relaxed. Structures are deemed converged when the magnitude of the largest atomic force is 20 meV/Å or less, and the largest stress component is less than 1 kbar. With these convergence settings energy changes between the last ionic iterations are a few μeV/atom only. Reciprocal space integrations were carried out with 600 uniformly spaced Monkhorst–Pack points.
[76] F. W. Schapink, Philos. Mag. **12**, 1055 (1965).
[77] B. Grabowski, T. Hickel, and J. Neugebauer, Phys. Status Solidi B **248**, 1295 (2011).
[78] R. Pawellek, M. Fähnle, C. Elsässer, K. M. Ho, and C. T. Chan, J. Phys.: Condens. Matter **3**, 2451 (1991).
[79] H. M. Polatoglou, M. Methfessel, and M. Scheffler, Phys. Rev. B **48**, 1877 (1993).
[80] T. Korhonen, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **51**, 9526 (1995).
[81] D. E. Turner, Z. Z. Zhu, C. T. Chan, and K. M. Ho, Phys. Rev. B **55**, 13842 (1997).
[82] A. Satta, F. Willaime, and S. de Gironcoli, Phys. Rev. B **57**, 11184 (1998).
[83] A. Satta, F. Willaime, and S. de Gironcoli, Phys. Rev. B **60**, 7001 (1999).
[84] D. Andersson and S. Simak, Phys. Rev. B **70**, 115108 (2004).
[85] K. Nordlund and R. Averback, in: Handbook of Materials Modeling, edited by Sidney Yip (Springer, Berlin, 2005), p. 1864.
[86] G. Grabowski, L. Ismer, T. Hickel, and J. Neugebauer, Phys. Rev. B **79**, 134106 (2009).
[87] M. Palumbo, B. Burton, A. C. de Silva, B. Fultz, B. Grabowski, G. Grimvall, B. Hallstedt, O. Hellmann, B. Lindahl, A. Schneider, P. E. A. Turchi, and W. Xiong, Phys. Status Solidi B **251**, 14–32 (2014), this issue.
[88] B. Grabowski, T. Hickel, and J. Neugebauer, Phys. Rev. B **76**, 024309 (2007).
[89] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni, Phys. Rev. B **43**, 7231 (1991).
[90] B. Grabowski, P. Söderlind, T. Hickel, and J. Neugebauer, Phys. Rev. B **84**, 214107 (2011).
[91] A. Glensk, B. Grabowski, T. Hickel, and J. Neugebauer, Phys. Rev. X, accepted (2013).
[92] R. Armiento and A. E. Mattsson, Phys. Rev. B **72**, 085108 (2005).
[93] A. E. Mattsson, R. Armiento, J. Paier, G. Kresse, J. M. Wills, and T. R. Mattsson, J. Chem. Phys. **128**, 084714 (2008).
[94] N. Chetty, M. Weinert, T. S. Rahman, and J. W. Devenport, Phys. Rev. B **52**, 6313 (1995).
[95] K. Carling, G. Wahnström, T. Mattson, N. Sandberg, and G. Grimvall, Phys. Rev. B **67**, 054101 (2003).
[96] M. Mantina, Y. Wang, R. Arroyave, L. Q. Chen, Z. K. Liu, and C. Wolverton, Phys. Rev. Lett. **100**, 215901 (2008).
[97] T. Hehenkamp, J. Phys. Chem. Solids **55**, 907 (1994).
[98] A. Khellaf, A. Seeger, and R. Emrick, Mater. Trans. **43**, 186 (2002).
[99] G. Kögel, P. Sperr, and W. Triftshäuser, J. Nucl. Mater. **131**, 148 (1985).
[100] G. Neumann, V. Toelle, and C. Tuijn, Physica B **271**, 21 (1999).
[101] L. A. Girifalco (ed.), Statistical Physics of Materials (John Wiley & Sons, New York, 1973).
[102] J. E. Kluin, Philos. Mag. **65**, 1263 (1992).
[103] W. Schüle, Z. Metallkd. **85**, 78 (1994).
[104] J. E. Kluin and T. Hehenkamp, Phys. Rev. B **44**, 11597 (1991).
[105] J. Rivière and J. Grilhé, Acta Metall. **20**, 1275 (1972).
[106] Y. Kraftmakher, Phys. Rep. **299**, 79 (1998).
[107] Y. Kraftmakher, Lecture Notes on Equilibrium Point Defects and Thermophysical Properties of Metals (World Scientific, Singapore, 2000).
[108] O. G. Randl, G. Vogl, W. Petry, B. Hennion, B. Sepiol, and K. Nembach, J. Phys.: Condens. Matter **7**, 5983 (1995).

[109] H. Lafi, M. Dirand, L. Bouirden, J. Hertz, D. Weber, and P. Lesbats, Acta Metall. **34**, 425 (1986).
[110] B. Sepiol and G. Vogl, Phys. Rev. Lett. **71**, 731 (1993).
[111] A. Seeger, J. Phys. F, Metal Phys. **3**, 248 (1973).
[112] R. Feder and A. Nowick, Phys. Rev. **109**, 1959 (1958).
[113] R. Simmons and R. Balluffi, Phys. Rev. **117**, 52 (1960).
[114] T. Hehenkamp, W. Berger, J. E. Kluin, C. Lüdecke, and J. Wolff, Phys. Rev. B **45**, 1998 (1992).
[115] K. Mosig, J. Wolff, J. Kluin, and T. Hehenkamp, J. Phys.: Condens. Matter **4**, 1447 (1992).
[116] K. Mosig, PhD Thesis, University of Göttingen (1993).
[117] R. Kerl, PhD Thesis, University of Göttingen (1994).
[118] H. Scholz, PhD Thesis, University of Göttingen (2001).
[119] W. Sprengel, B. Oberdorfer, E. Steyskal, and R. Würschum, J. Mater. Sci. **47**, 7921 (2012).
[120] B. Guèrard, H. Peisl, and R. Zitzmann, Appl. Phys. **3**, 37 (1974).
[121] R. Simmons and R. Balluffi, Phys. Rev. **125**, 862 (1962).
[122] W. Trost, K. Differt, K. Maier, and A. Seeger, Point Defects in Metals in Thermal Equilibrium, Springer Proceedings in Physics, Vol. 10 (Springer, Berlin/Heidelberg, 1986).
[123] H. Schaefer, Phys. Status Solidi A **102**, 47 (1987).
[124] B. Oberdorfer and R. Würschum, Phys. Rev. B **79**, 184103 (2009).
[125] H. Kauppinen, C. Corbel, K. Skog, K. Saarinen, T. Laine, P. Hautojarvi, P. Desgardin, and E. Ntsoenzok, Phys. Rev. B **55**, 9598 (1997).
[126] J. Dagata, F. Perez-Murano, C. Martin, H. Kuramochi, and H. Yokoyama, J. Appl. Phys. **96**, 2386 (2004).
[127] H. Bracht, E. Haller, and R. Clark-Phelps, Phys. Rev. Lett. **81**, 393 (1998).
[128] H. Bracht, H. Silvestri, I. Sharp, and E. Haller, Phys. Rev. B **75**, 035211 (2007).
[129] W. Kramer and J. Nölting, Acta Metall. **20**, 1353 (1972).
[130] R. Simmons and R. Balluffi, Phys. Rev. **117**, 62 (1960).
[131] H. Schaefer, R. Gugelmeier, M. Schmolz, and A. Seeger, Mater. Sci. Forum **16/18**, 111 (1987).
[132] Y. Kraftmakher, Sov. Phys. Solid State **9**, 1458 (1967).
[133] T. Hehenkamp and L. Sander, Z. Metallkd. **70**, 202 (1979).
[134] R. Scholz and W. Schüle, Phys. Lett. A **64**, 340 (1977).
[135] S. Glazkov, High Temp. **25**, 51 (1987).
[136] W. Wycisk and M. Feller-Kniepmeier, Phys. Status Solidi A **37**, 183 (1976).
[137] M. Fluss, L. Smedskjaer, B. Chakraborty, and M. Chason, J. Phys. F, Metal Phys. **13**, 817 (1983).
[138] E. M. Forsblom, N. Sandberg, and G. Grimvall, Phys. Rev. B **69**, 165106 (2004).
[139] F. Varotsos, J. Appl. Phys. **105**, 083524 (2009).
[140] R. Shukla, C. Pint, and D. Ditmars, Int. J. Thermophys. **6**, 517 (1985).
[141] Y. Kraftmakher and P. Strelkov, Sov. Phys. Solid State **4**, 1662 (1962).
[142] Y. Kraftmakher and E. Lanina, Sov. Phys. Solid State **7**, 92 (1965).
[143] V. Kirillin, A. Sheindlin, V. Chekhovskoi, and I. Zhukova, High Temp. **5**, 1016 (1967).
[144] A. Cezairliyan, M. Morse, H. Berman, and C. Beckett, J. Res. Natl. Bur. Stand. A **74**, 65 (1970).
[145] F. Righini, R. Roberts, and A. Rosso, Int. J. Thermophys. **6**, 681 (1985).
[146] C. R. Brooks and R. E. Bingham, J. Phys. Chem. Solids **29**, 1553 (1968).
[147] D. Ditmars, C. Plint, and R. Shukla, Int. J. Thermophys. **6**, 499 (1985).
[148] A. Leadbetter, J. Phys. C **1**, 1489 (1968).
[149] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[150] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).
[151] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[152] B. Sundman and J. Ågren, J. Phys. Chem. Solids **42**, 297 (1981).