

15

Formation Energies of Point Defects at Finite Temperatures

Blazej Grabowski, Tilmann Hickel, and Jörg Neugebauer

15.1

Introduction

A crucial quantity for the *ab initio* study of point defects is the defect formation free energy $F^f(V, T)$ as a function of volume V and temperature T . The dominant contribution to F^f is due to the zero temperature formation energy $E^f(V) = F^f(V, T = 0\text{K})$, which can be calculated at a relatively low computational cost. The calculation of higher order contributions such as quasiharmonic excitations (= non-interacting harmonic vibrations + effect of thermal expansion; Section 15.2.2.3) and anharmonic excitations (= interacting vibrations; Section 15.2.2.4) significantly increases the required computational resources. Since these effects are also expected to yield only a comparatively small contribution to F^f they are typically neglected. Indeed, it is expected that their influence on defect properties in semiconducting materials is far smaller than the inaccuracies resulting from the band gap problem. Thus, the majority of defect studies in semiconductors (see Ref. [1] for a recent review) are based on E^f with only a few exceptions [2, 3].

The situation is likely to change in the near future. Recent progress in the development of new exchange-correlation functionals [4, 5] and methods going beyond density functional theory (DFT) [6, 7] allows for a highly accurate prediction of band structures. At present, such calculations are computationally too expensive to be routinely applicable to total energy calculations of defects. However, the steady progress in methodological development and hardware components will soon close this gap. Then, the determination of the above mentioned higher order contributions will become critical.

For metals which do not suffer from the band gap problem, the situation is different. The highly efficient screening in metallic materials removes a large part of the self-interaction error which is mainly responsible for the band gap problem in semiconductors. As a consequence, significantly more accurate defect formation energies are obtained even with common local or semi-local exchange-correlation functionals such as the local density approximation (LDA) or the generalized gradient approximation (GGA). Thus, in order to reach the next accuracy level in defect

Table 15.1 Representative *ab initio* studies of point defect calculations in unary metals for the specific case of vacancies. The abbreviations are: 3d/4d/5d: respective transition elements; xc: exchange-correlation functional; LDA: local density approximation; GGA: generalized gradient approximation; PWps: planewaves with pseudopotentials; FP-LMTO: full potential linearized muffin tin orbitals; PW-PAW: planewaves with projector augmented waves; V: rescaled volume approach; P: constant pressure approach; volOpt: volume optimized approach (Section 15.2.1.1); F^f : defect formation free energy; E^f : ($T = 0$ K) contribution to F^f ; el/qh/ah: electronic/quasiharmonic/anharmonic contribution to F^f ; 1s: first shell (around the defect) contribution to the dynamical matrix; emp: empirical potential approach.

| year | ref. | elements | methodology | | | contributions to F^f | | | |
|------|------|-------------|-------------|-------------|----------|------------------------|----|----------|-----------|
| | | | xc | potential | strain | E^f | el | qh | ah |
| 1989 | [8] | Al | LDA | PWps | V | x | | | |
| 1991 | [9] | Li | LDA | PWps | V | x | | | |
| 1993 | [10] | Al,Cu,Ag,Rh | LDA | FP-LMTO | V | x | | | |
| 1995 | [11] | 3d,4d,5d | LDA | FP-LMTO | V | x | | | |
| 1997 | [12] | Al | LDA | PWps | P | x | | | |
| 1998 | [13] | W | LDA | PWps | P | x | x | | |
| 1999 | [14] | Ta | LDA | PWps | P | x | x | | |
| 2000 | [15] | Al | LDA/GGA | PWps | P | x | | x^{1s} | x^{emp} |
| 2003 | [16] | Al | LDA/GGA | PWps | P | x | | x^{1s} | x^{emp} |
| 2009 | [17] | Fe | GGA | PW-PAW | V | x | | x | |
| 2009 | [18] | Al | LDA/GGA | PWps/PW-PAW | volOpt/P | x | x | x | x |

calculations, the inclusion of finite temperature contributions to F^f is already now of importance. Indeed, a review of the related literature clearly reveals such efforts (Table 15.1): starting in the late 1980s with the seminal work by Gillan [8], DFT-based studies of point defects were limited to the $T = 0$ K contribution E^f . This situation persisted roughly until the beginning of the new century, when studies [13, 14] of the electronic contribution to F^f – of crucial importance for some metallic materials [19] – appeared. In 2000 and 2003, Carling *et al.* [15, 16] provided a first *ab initio* based assessment of the quasiharmonic contribution to the vacancy of aluminum. To make such a study feasible at that time, the authors had to restrict the dynamics of the system to the first shell around the vacancy, *i.e.*, to the atomic shell which experiences the largest effect as compared to the perfect bulk. An *ab initio* based evaluation of the anharmonic contribution was computationally prohibitive at that time, which made it necessary to resort to empirical potentials. Major methodological improvements and the boost in computer power provide now the opportunity to study all relevant free energy contributions of defect formation in a rigorous *ab initio* manner (*cf.* Table 15.1).

In the present paper, we review the methodology required to compute defect concentrations from *ab initio* including the electronic, quasiharmonic, and anharmonic contributions to the formation free energy (Section 15.2.2). For their correct evaluation and interpretation it is important to correctly treat the strain induced by the periodic array of defects in a supercell approach, since an improper treatment

may lead to errors of the same order of magnitude. We therefore review first the possible strategies and available correction schemes (Section 15.2.1). We then present results demonstrating the quality and performance of the methods (Section 15.3). The focus will be on point defects in aluminum since this material system can be produced with high chemical purity and crystalline quality, thus providing accurate experimental data as needed for a critical comparison. On the theoretical side, a good performance of available exchange-correlation functionals can be expected due to the free electron character of Al. All these aspects render aluminum to be a particularly attractive system for evaluating the performance of *ab initio* simulations of point defects. Indeed, as shown in Table 15.1 most theoretical studies focused on this system.

15.2

Methodology

15.2.1

Analysis of Approaches to Correct for the Spurious Elastic Interaction in a Supercell Approach

In the literature, two major approaches have been proposed and employed to correct for the artificial strain fields in a supercell approach arising from a collective interplay of all periodic images: (i) the rescaled volume and (ii) the constant pressure approach. Within the rescaled volume approach [8], the volume of the supercell containing the defect is rescaled such as to account for the volume of the missing atom. In contrast, within the constant pressure approach [12], the volume of the defect supercell is adjusted such as to correspond to the same pressure as is acting on the perfect bulk supercell (commonly zero pressure is assumed). In the limit of asymptotically large supercell sizes both approaches will converge to the same result. For realistic finite sized *ab initio* supercells the two approaches give different results. It is commonly accepted that the constant pressure method is superior to (more accurate than) the rescaled volume one. This is due to the fact that the latter imposes additional constraints on the system while the constant pressure approach allows the system to relax along all degrees of freedom including the shape of the supercell. A disadvantage of the constant pressure approach is that additional relaxations are needed, which significantly increase computational effort. This fact is illustrated in Table 15.1: early calculations of defect properties, when computer power was severely limited, were solely based on the rescaled volume approach. Only at the end of the 1990s one was able to achieve the next level of accuracy and employ the constant pressure approach. For the accurate calculation of finite temperature contributions to the defect formation energy employing the more accurate constant pressure approach is mandatory. Recently a more general approach, the volume optimized scheme, was proposed [18]. It takes higher order terms in the concentration dependence of F^f into account thus going beyond the constant pressure treatment. A consequence is that the formation free energy becomes concentration dependent.

Approximating to first order in the defect concentration, the constant pressure approach is obtained. Performing an additional approximation in the volume of defect formation, the rescaled volume approach can be derived. The relation and hierarchy between these approaches is discussed in the following.

15.2.1.1 The Volume Optimized Approach to Point Defect Properties

The central quantity, which contains all thermodynamic information about the system such as *e.g.*, the defect concentration, is the free energy surface $F(\Omega, T; N, n)$ of a macroscopic crystal. In general, it depends on the crystal volume Ω (we reserve the symbol V for the atomic volume introduced below), temperature T , the number of atoms N , and the number of defects n . For the following discussion, we consider a large fictitious supercell (Born–von Karman cell) representing this macroscopic crystal (Figure 15.1).

The term fictitious refers to the fact that an actual calculation of this supercell is not feasible (and as will be discussed not necessary). The basic assumption is that the presence of defects leads to two kinds of effects: (i) strong distortions of the atoms close to the defect away from their ideal perfect bulk positions and (ii) long ranged volumetric distortions affecting only the lattice constant. Around each defect, a cell/box is constructed which we call defect cell. The defect cell needs to be large enough to cover the first kind of effect but not necessarily the second kind since this will be accounted for by the volume optimization introduced below. The defect cell contains N^d atoms, has a volume Ω^d , and a free energy $F^d = F^d(\Omega^d, T; N^d)$. According to this construction, the crystal outside the defect cells can be described

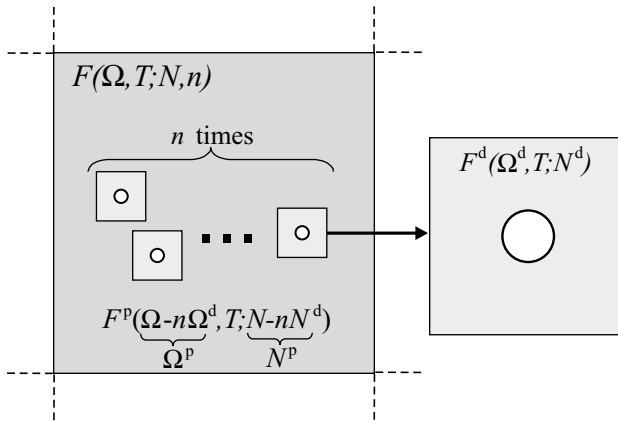


Figure 15.1 Schematic illustration of the concept to compute the free energy $F(\Omega, T; N, n)$ of a macroscopic crystal with defects. The larger box represents a supercell of volume Ω at temperature T containing N atoms and n defects. A light-gray shaded box with a white circle represents a cell of volume Ω^d ,

containing N^d atoms, exactly one defect, and having the free energy $F^d(\Omega^d, T; N^d)$. The dark-gray shaded region represents the perfect crystal without defects, with volume $\Omega^p = \Omega - n\Omega^d$, $N^p = N - nN^d$ atoms, and free energy $F^p(\Omega^p, T; N^p)$. The dashed lines indicate periodic boundary conditions.

by a perfect crystal with volume $\Omega^p = \Omega - n\Omega^d$, $N^p = N - nN^d$ atoms, and free energy $F^p = F^p(\Omega^p, T; N^p)$. The free energy of the fictitious supercell is then given by

$$F(\Omega, T; N, n, \Omega^d, N^d) = F^p(\Omega^p, T; N^p) + nF^d(\Omega^d, T; N^d) + F^{\text{conf}}(T, N, n), \quad (15.1)$$

where we have also explicitly indicated the dependence of F on the volume of the defect cell Ω^d and the number of atoms in the defect cell N^d . These dependencies and their treatment will be discussed in the following. Further in Eq. (15.1), F^{conf} is the configurational free energy of the defects in the dilute limit. It is approximated using Stirling's formula by [20] $F^{\text{conf}} \approx -k_B T [n - n \ln(n/N)]$, with the Boltzmann constant k_B . The volume optimized approach [18] is based on the key observation that in equilibrium the total free energy F is minimal with respect to changes in the volume of the defect cell,

$$\partial F / \partial \Omega^d \equiv 0, \quad (15.2)$$

so that the volume of the perfect crystal and of the defect cells will adjust self consistently, *i.e.*, until the optimum volume for both is achieved when minimizing F . The actual result of the minimization procedure will depend on the free energy volume curves of the perfect crystal and of the defect cell.

Equation (15.1) can be further transformed. The free energy of the perfect crystal F^p scales with the number of atoms due to its extensivity property,

$$\begin{aligned} F^p(\Omega^p, T; N^p) / N^p &= F^p(V^p, T; 1) \\ &=: F^p(V^p, T) = F^p\left(\frac{V - c\Omega^d}{1 - cN^d}, T\right), \end{aligned} \quad (15.3)$$

with

$$V^p = \frac{\Omega^p}{N^p} = \frac{V - c\Omega^d}{1 - cN^d}, \quad (15.4)$$

where we have defined the volume per atom $V = \Omega/N$ and the concentration of defects $c = n/N$. Using this property to rewrite Eq. (15.1) yields

$$\begin{aligned} F(\Omega, T; N, n, \Omega^d, N^d) / N &= (1 - cN^d) F^p(V^p, T) \\ &+ cF^d(\Omega^d, T; N^d) + F^{\text{conf}}(T, c) =: F(V, T; c, \Omega^d, N^d), \end{aligned} \quad (15.5)$$

where the configurational free energy depends now on the concentration as $F^{\text{conf}}(T, c) = -ck_B T (1 - \ln c)$. Equation (15.5) is independent of N and defines the free energy per atom $F(V, T; c, \Omega^d, N^d)$ of the full supercell consisting of the perfect crystal and the defect cells. Now, a formation free energy F^f can be defined, which turns out to be concentration dependent

$$F^f(V, T; c, \Omega^d, N^d) = F^d(\Omega^d, T; N^d) - N^d F^p(V^p, T), \quad (15.6)$$

and thus:

$$F(V, T; c, \Omega^d, N^d) = F^p(V^p, T) + cF^f(V, T; c, \Omega^d, N^d) + F^{\text{conf}}(T, c). \quad (15.7)$$

Applying next the equilibrium condition with respect to the defect concentration, $\partial F / \partial c \equiv 0$, to Eq. (15.7) yields an equation for the equilibrium defect concentration c^{eq}

$$-k_B T \ln c^{\text{eq}} = F^f(V, T; c^{\text{eq}}, \Omega^d, N^d) - \frac{v^f P^p}{c^{\text{eq}} N^d - 1}, \quad (15.8)$$

where the volume of defect formation $v^f = \Omega^d - N^d V$ and the pressure inside the perfect bulk $P^p = -\partial F^p / \partial V^p$ have been defined. It is straightforward to show that the latter equals the external pressure $P = -\partial F / \partial V$:

$$P = -\frac{\partial F}{\partial V} = -(1 - cN^d) \frac{\partial F^p}{\partial V^p} \frac{\partial V^p}{\partial V} = -\frac{\partial F^p}{\partial V^p} = P^p. \quad (15.9)$$

Further, it follows from Eq. (15.2) that

$$\begin{aligned} \frac{\partial F}{\partial \Omega^d} &= (1 - cN^d) \frac{\partial F^p}{\partial V^p} \frac{\partial V^p}{\partial \Omega^d} + c \frac{\partial F^d}{\partial \Omega^d} \\ &= c \left(\frac{\partial F^d}{\partial \Omega^d} - \frac{\partial F^p}{\partial V^p} \right) = c(-P^d + P^p) \equiv 0, \end{aligned} \quad (15.10)$$

with $P^d = -\partial F^d / \partial \Omega^d$ the pressure inside the defect cell. Hence, the equilibrium defect cell volume $\Omega^{d,\text{eq}}$ is obtained when the pressure inside the defect cells equals the pressure inside the perfect cell which, according to Eq. (15.9), equals the external pressure, i.e., $P^d = P^p = P$.

Using $\Omega^{d,\text{eq}}$ and Eqs. (15.8, 15.9), Eq. (15.7) can be transformed to the final expression for the free energy of a crystal at (atomic) volume V and temperature T and with an equilibrium concentration of thermally excited defects:

$$\begin{aligned} F(V, T) &= F(V, T; c^{\text{eq}}, \Omega^{d,\text{eq}}, N^d) \\ &= F^p \left(\frac{V - c^{\text{eq}} \Omega^{d,\text{eq}}}{1 - c^{\text{eq}} N^d}, T \right) + \frac{c^{\text{eq}} v^f P}{c^{\text{eq}} N^d - 1} - c^{\text{eq}} k_B T. \end{aligned} \quad (15.11)$$

The parameter N^d is determined by the specific supercell used for the defect calculation and has to be checked for convergence. Note that Eq. (15.8) [and thus Eq. (15.11)] cannot be solved for c^{eq} in closed form due to the dependence of F^f on c . The actual equilibrium defect concentration must be thus solved self consistently.

15.2.1.2 Derivation of the Constant Pressure and Rescaled Volume Approach

Employing well defined approximations, the constant pressure and rescaled volume approaches can be easily derived from the volume optimized approach. Let us first show the relation to the constant pressure approach. We therefore Taylor expand F in Eq. (15.5) as a function of c around $c = 0$:

$$F(V, T) = F^p(V, T) + c[F^d(\Omega^d, T; N^d) - N^d F^p(V, T) + P v^f] + F^{\text{conf}}(T, c) + O(c^2), \quad (15.12)$$

where Eq. (15.9) has been used. Retaining only the terms linear in c , the expression for the free energy within the constant pressure approach is obtained [21]:

$$F(V, T) \approx F^p(V, T) + F^{\text{conf}}(T, c) + c \underbrace{[F^d(\Omega^d, T; N^d) - N^d F^p(V, T) + P v^f]}_{=F_p^f, \text{formation free energy at const. pressure.}}. \quad (15.13)$$

The term in the square brackets defines the defect formation free energy at constant pressure, since Ω^d needs to be chosen such as to satisfy the pressure equality, Eq. (15.10). The term correctly includes the enthalpic $P v^f$ contribution. This contribution has been intensively discussed over many years in literature: in their textbook, Varotsos and Alexopoulos [21] stress that it has been frequently ignored in point defect studies. For a correct description at nonzero pressure it needs however to be included. Based on the above derivation, we can straightforwardly analyze the necessity of this contribution. It naturally arises from the Taylor expansion in Eq. (15.12) and needs to be taken into account since it is part of the first order term. Physically, the $P v^f$ term reflects the fact that the work needed to form a defect depends on the pressure and likewise on the volume changes it induces. Equation (15.13) can be simplified in a standard way [21] by using the defect equilibrium condition $\partial F / \partial c \equiv 0$ to yield $F(V, T) \approx F^p(V, T) - k_B T c^{\text{eq}}(V, T)$, with the equilibrium defect concentration $c^{\text{eq}}(V, T) = \exp[-F_p^f(V, T)/(k_B T)]$.

The rescaled volume approach can be easily derived from Eq. (15.12). We therefore note that $F^d(\Omega^d, T; N^d)$ is the zeroth order term of a Taylor expansion of $F^d(\Omega^d + v^f, T; N^d)$ in the volume of defect formation v^f around $v^f = 0$:

$$\begin{aligned} F^d(N^d V, T; N^d) &= F^d(\Omega^d + v^f, T; N^d) \\ &= F^d(\Omega^d, T; N^d) + P v^f + O([v^f]^2). \end{aligned} \quad (15.14)$$

In the above equation the first equality follows from the definition of the volume of defect formation. Further, $-\partial F^d / \partial \Omega^d = P$ due to Eqs. (15.9) and (15.10) has been used. Approximating to first order in v^f , rearranging with respect to $F^d(\Omega^d, T; N^d)$, and plugging into Eq. (15.13) yields:

$$F(V, T) \approx F^p(V, T) + F^{\text{conf}}(T, c) + c \underbrace{[F^d(N^d V, T; N^d) - N^d F^p(V, T)]}_{=F_V^f, \text{formation free energy at rescaled volume.}}. \quad (15.15)$$

The quantity in square brackets is the formation free energy of the rescaled volume approach [8]. To make this even more apparent, the extensivity property of F^p can be used to write F_V^f as

$$F_V^f(V, T; N^d) = F^d(N^d V, T; N^d) - \frac{N^d}{N^d \pm 1} F^p([N^d \pm 1] V, T; N^d \pm 1), \quad (15.16)$$

with the plus (minus) sign referring to vacancies (self interstitials). As for the constant pressure case, Eq. (15.15) can be simplified to $F(V, T) \approx F^p(V, T) - k_B T c^{\text{eq}}(V, T)$, with the equilibrium defect concentration given now by $c^{\text{eq}}(V, T) = \exp[-F_V^f(V, T)/(k_B T)]$.

The preceding derivations show that the two standard approaches arise as natural approximations of the volume optimized method. In particular, a hierarchy of approximations can be identified: first, the constant pressure approach arises by terminating the Taylor series in the defect concentration in Eq. (15.12) after the first order term. The rescaled volume approach needs a further approximation by terminating the Taylor series in the volume of defect formation in Eq. (15.14) likewise after the first order term. For practical purposes we note the following: The approximation in the defect concentration, Eq. (15.13), is well motivated, since the basic assumption of non interacting defects, *i.e.*, the dilute limit, is valid only for low defect concentrations. We therefore recommend to employ the constant pressure approach, also because of numerical instabilities in the volume optimized method when approaching $c^{\text{eq}} N^{\text{d}} \approx 1$ due to the denominator in Eq. (15.11). In contrast, the approximations needed to derive the rescaled volume approach are not appropriate for realistic supercell sizes and will result in sizeable errors even for low defect concentrations (see *e.g.*, Figure 15.7).

15.2.2

Electronic, Quasiharmonic, and Anharmonic Contributions to the Formation Free Energy

Let us now focus on the methodology needed to compute the electronic, quasiharmonic, and anharmonic contribution to the free energy surface. For a defect calculation, we need both the free energy of the defect cell $F^{\text{d}}(V, T)$ and the free energy of the perfect bulk $F^{\text{p}}(V, T)$ as a function of volume and temperature as outlined in the previous section. In the following we will derive the necessary steps with particular emphasis on numerical efficiency. The latter issue is crucial to allow a full *ab initio* determination of all contributions, specifically including the anharmonic one. Except for the remark at the end of Section 15.2.2.3 all considerations refer to both free energy surfaces (F^{d} and F^{p}) and we therefore use generically the symbol F .

15.2.2.1 Free Energy Born–Oppenheimer Approximation

Starting point is an expression for the free energy surface in which the ionic and electronic degrees of freedom are decoupled quantum mechanically (Born–Oppenheimer approximation), but which still contains the effect of thermodynamic electronic excitations on the ionic vibrations. For that purpose, the “standard” Born–Oppenheimer approximation [22] needs to be extended to the so called free energy Born–Oppenheimer approximation which was introduced by Cao and Berne [23] in 1993.

Within the standard Born–Oppenheimer approximation, the free energy F of a system consisting of electrons and ions is written as:

$$F = -k_B T \ln Z \quad \text{where} \quad Z = \sum_{v,\mu} e^{-\beta E_{v,\mu}^{\text{nuc}}}, \quad (15.17)$$

with

$$E_{v,\mu}^{\text{nuc}} = \langle \Lambda_{v,\mu} | (\hat{T}^{\text{nuc}} + \hat{1} E_v^{\text{el}}) | \Lambda_{v,\mu} \rangle, \quad (15.18)$$

and $\beta = (k_B T)^{-1}$. In Eq. (15.17), we have defined the partition function Z of the system in which the sums run over an electronic quantum number v and an ionic quantum number μ . The energy levels $E_{v,\mu}^{\text{nuc}}$ and eigenfunctions $\Lambda_{v,\mu}$ are solutions to the nuclei Schrödinger equation with the Hamiltonian $(\hat{T}^{\text{nuc}} + \hat{1} E_v^{\text{el}})$, in which \hat{T}^{nuc} is the ionic kinetic energy operator, $\hat{1}$ is the identity operator, and E_v^{el} are potential energy surfaces generated by the electronic system, *i.e.*, the solutions to the electronic Schrödinger equation (*cf.* Figure 15.2). (As commonly done, we include the nucleus–nucleus interaction into E_v^{el} .) We can transform the partition function as

$$\begin{aligned} Z &= \sum_{v,\mu} e^{-\beta \langle \Lambda_{v,\mu} | (\hat{T}^{\text{nuc}} + \hat{1} E_v^{\text{el}}) | \Lambda_{v,\mu} \rangle} \\ &= \sum_{v,\mu} \langle \Lambda_{v,\mu} | e^{-\beta (\hat{T}^{\text{nuc}} + \hat{1} E_v^{\text{el}})} | \Lambda_{v,\mu} \rangle, \end{aligned} \quad (15.19)$$

since the $\Lambda_{v,\mu}$ are eigenfunctions of $(\hat{T}^{\text{nuc}} + \hat{1} E_v^{\text{el}})$. It would be now desirable to factorize the exponential to separate the E_v^{el} . This factorization needs however to be performed

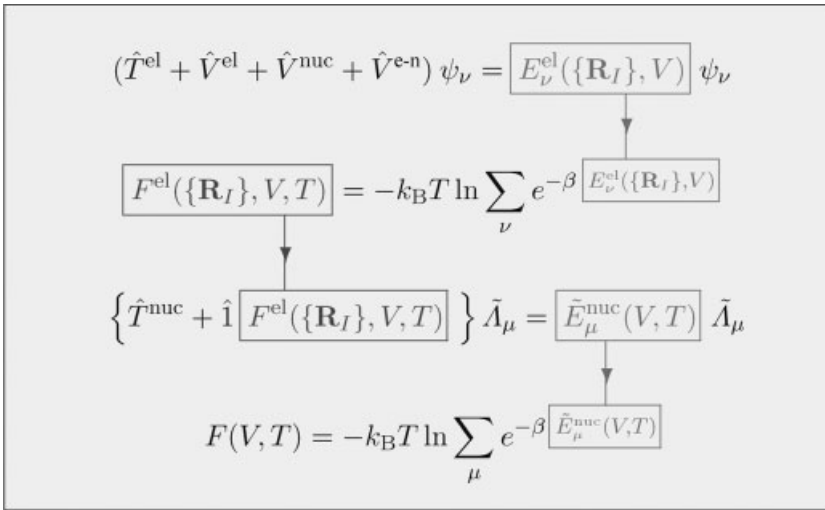


Figure 15.2 (online colour at: www.pss-b.com) Key equations to compute the free energy Born–Oppenheimer surface. Here, ψ_ν are electronic wave functions and $(\hat{T}^{\text{el}} + \hat{V}^{\text{el}} + \hat{V}^{\text{nuc}} + \hat{V}^{\text{e-n}})$ is the electronic Hamiltonian with the electronic kinetic energy

operator, electron–electron repulsion operator, nucleus–nucleus repulsion operator, and electron–nucleus attraction operator, respectively. The remaining quantities are defined in Section 15.2.2.1.

with caution since \hat{T}^{nuc} and $\hat{1} E_v^{\text{el}}$ are non-commuting operators (both depend on the ionic coordinates). Therefore, we have to apply the so-called Zassenhaus formula [24]:

$$e^{-\beta(\hat{T}^{\text{nuc}} + \hat{1} E_v^{\text{el}})} = e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} E_v^{\text{el}}} e^{\beta^2/2[\hat{T}^{\text{nuc}}, \hat{1} E_v^{\text{el}}]} e^{-\beta^3/6(2[\hat{1} E_v^{\text{el}}, [\hat{T}^{\text{nuc}}, \hat{1} E_v^{\text{el}}]] + [\hat{T}^{\text{nuc}}, [\hat{T}^{\text{nuc}}, \hat{1} E_v^{\text{el}}]])} \dots \quad (15.20)$$

Here, the dots denote exponentials corresponding to higher orders in β and with increasingly nested commutators. An explicit formula for the higher order terms is given in Ref. [24]. In Ref. [23], it is shown that exponentials corresponding to orders β^2 and higher will be small if $m_e \ll M$ (m_e = electron mass, M = nucleus mass), i.e., under the same condition as assumed in the standard Born–Oppenheimer approximation. It is therefore justified to approximate

$$e^{-\beta(\hat{T}^{\text{nuc}} + \hat{1} E_v^{\text{el}})} \approx e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} E_v^{\text{el}}}, \quad (15.21)$$

in any case in which the Born–Oppenheimer approximation is justified. Using Eq. (15.21), the fact that $(e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} E_v^{\text{el}}})$ corresponds to a block diagonal matrix and the invariance property of the trace (which allows to choose the same basis for different μ , e.g., $v = v'$; with a fixed v' ;) yields:

$$\begin{aligned} Z &= \sum_{v, \mu} \langle \Lambda_{v, \mu} | e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} E_v^{\text{el}}} | \Lambda_{v, \mu} \rangle \\ &= \sum_{v, \mu} \langle \Lambda_{v', \mu} | e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} E_v^{\text{el}}} | \Lambda_{v', \mu} \rangle \\ &= \sum_{\mu} \langle \Lambda_{v', \mu} | e^{-\beta\hat{T}^{\text{nuc}}} \left(\sum_{\mu} e^{-\beta\hat{1} E_v^{\text{el}}} \right) | \Lambda_{v', \mu} \rangle \\ &= \sum_{\mu} \langle \Lambda_{v', \mu} | e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} F^{\text{el}}} | \Lambda_{v', \mu} \rangle, \end{aligned} \quad (15.22)$$

with the electronic free energy defined by:

$$F^{\text{el}} := -k_B T \ln \sum_v \exp[-E_v^{\text{el}}/(k_B T)]. \quad (15.23)$$

In order to recombine the exponentials again, we need to apply the Baker–Campbell–Hausdorff formula which reads [24]:

$$e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} F^{\text{el}}} = e^{-\beta(\hat{T}^{\text{nuc}} + \hat{1} F^{\text{el}}) - \frac{\beta^2}{2}[\hat{T}^{\text{nuc}}, \hat{1} F^{\text{el}}] - \frac{\beta^3}{6}[[\hat{T}^{\text{nuc}}, \hat{1} F^{\text{el}}], \hat{1} F^{\text{el}} - \hat{T}^{\text{nuc}}]} \dots \quad (15.24)$$

The terms in the exponential of order β^2 and higher correspond again to terms which are small if $m_e \ll M$, thus justifying the following approximation:

$$e^{-\beta\hat{T}^{\text{nuc}}} e^{-\beta\hat{1} F^{\text{el}}} \approx e^{-\beta(\hat{T}^{\text{nuc}} + \hat{1} F^{\text{el}})}. \quad (15.25)$$

In fact, as shown in Ref. [23] the terms neglected in Eq. (15.25) lead to contributions which are of the same order as the contributions from the neglected terms in

Eq. (15.21) having however the opposite sign. Therefore, the approximations performed in Eqs. (15.21) and (15.25) partially compensate each other. Inserting Eq. (15.25) into Eq. (15.22) yields for the partition function and thus the free energy:

$$F = -k_B T \ln Z, \quad (15.26)$$

with

$$Z = \sum_{\mu} \langle \Lambda_{\mu} | e^{-\beta(\hat{T}^{\text{nuc}} + \hat{1} F^{\text{el}})} | \Lambda_{\mu} \rangle. \quad (15.27)$$

We can rewrite this to a more convenient notation as

$$F = -k_B T \ln \sum_{\mu} e^{-\beta \tilde{E}_{\mu}^{\text{nuc}}}, \quad (15.28)$$

with

$$\left(\hat{T}^{\text{nuc}} + \hat{1} F^{\text{el}}(\{\mathbf{R}_I\}, V, T) \right) \tilde{\Lambda}_{\mu} = \tilde{E}_{\mu}^{\text{nuc}} \tilde{\Lambda}_{\mu}, \quad (15.29)$$

where we have defined an effective nuclei Schrödinger equation with eigenfunctions $\tilde{\Lambda}_{\mu}$ and eigenvalues $\tilde{E}_{\mu}^{\text{nuc}}$. In Eq. (15.29), we have also explicitly written the dependence of F^{el} on the set of nuclei coordinates $\{\mathbf{R}_I\}$ and the crystal volume V which is a consequence of the fact that each of the electronic potential energy surfaces E_v^{el} depends on $\{\mathbf{R}_I\}$ and V . The key equations summarizing the preceding derivation are collected in Figure 15.2.

The central step which allows for a separation into an electronic, quasiharmonic, and anharmonic part is a Taylor expansion of F^{el} in Eq. (15.28) in the $\{\mathbf{R}_I\}$ around the $T = 0$ K equilibrium positions $\{\mathbf{R}_I^0\}$:

$$F^{\text{el}} = F_0^{\text{el}} + \frac{1}{2} \sum_{k,l} u_k u_l \left[\frac{\partial^2 F^{\text{el}}}{\partial R_k \partial R_l} \right]_{\{\mathbf{R}_I^0\}} + O(u^3). \quad (15.30)$$

Here, the zeroth order term is abbreviated as $F_0^{\text{el}}(V, T) := F^{\text{el}}(\{\mathbf{R}_I^0\}, V, T)$, k and l run over all nuclei of the system and additionally over the three spatial dimensions for each nucleus, and $u_k = R_k - R_k^0$ is the displacement out of equilibrium. The expansion does not contain a first order term, since such a term relates to the atomic forces that are absent in an equilibrium structure. Each of the other terms in Eq. (15.30) corresponds to a different excitation mechanism as emphasized in Figure 15.3. The exact derivations are given in the following sections.

15.2.2.2 Electronic Excitations

It is convenient to decompose the electronic part $F_0^{\text{el}}(V, T)$ in Eq. (15.30) into a temperature independent part $E_{g,0}^{\text{el}}(V)$, which corresponds to the ground state of the potential energy surfaces E_v^{el} at $\{\mathbf{R}_I^0\}$, and a remainder $\tilde{F}_0^{\text{el}}(V, T)$ carrying the temperature dependence of the electronic system, *i.e.*, the electronic excitations: $F_0^{\text{el}}(V, T) = E_{g,0}^{\text{el}}(V) + \tilde{F}_0^{\text{el}}(V, T)$. The reason for this separation is that $\tilde{F}_0^{\text{el}}(V, T)$ can be accurately described with low order polynomials, while $E_{g,0}^{\text{el}}(V)$ can be parame-

$$F^{\text{el}}(\{\mathbf{R}_I\}, V, T) = \underbrace{F_0^{\text{el}}(V, T)}_{\text{electronic contr.} \rightarrow \text{Sec. 2.2.2}} + \underbrace{\frac{1}{2} \sum_{k,l} u_k u_l \left[\frac{\partial^2 F^{\text{el}}(\{\mathbf{R}_I\}, V, T)}{\partial R_k \partial R_l} \right]_{\{\mathbf{R}_I\}}}_{\text{quasiharmonic contribution} \rightarrow \text{Sec. 2.2.3}} + \underbrace{\text{orders } u^3 \text{ and higher}}_{\text{anharmonic contribution} \rightarrow \text{Secs. 2.2.4 and 2.2.5}}$$

Figure 15.3 (online colour at: www.pss-b.com) Taylor expansion of the electronic free energy F^{el} as in Eq. (15.20) (see text for definitions). The color coding emphasizes the connection with the equations in Figure 15.2. The under braces indicate the type of contribution arising from each term and the section covering the issue.

trized using standard equations-of-state. To calculate the latter a standard DFT approach is sufficient. For the calculation of \tilde{F}_0^{el} , one needs instead to employ the finite temperature extension of DFT as originally developed by Mermin [25]. This approach is implemented in typical DFT codes and it amounts to using a Fermi–Dirac occupation distribution for the Kohn–Sham electronic energy levels,

$$\tilde{F}_0^{\text{el}}(V, T) \approx k_B T / 2 \sum_i [f_i \ln f_i + (1 - f_i) \ln (1 - f_i)], \quad (15.31)$$

with self consistently determined Kohn–Sham occupation numbers $f_i = f_i(V, T)$. Note that Eq. (15.31) is only approximately valid since there is a small contribution from the kinetic energy term, which is however fully accounted for in an actual finite temperature DFT calculation.

While the inclusion of F_0^{el} is important for metals at realistic temperatures, for semiconductors it is negligible except for narrow band gap semiconductors, where also partial occupations f_i may occur. In metals, \tilde{F}_0^{el} can become significant particularly if the density of states shows a peak close to the Fermi energy as found for instance for d-states in Pt, Pd, Rh, or Ir [19].

15.2.2.3 Quasiharmonic Atomic Excitations

Neglecting for the moment the higher order terms in Eq. (15.29), the quasiharmonic approximation results. The necessary steps to compute this contribution are

$$\begin{aligned} D_{k,l}(V, T) &:= \frac{1}{\sqrt{M_k M_l}} \left[\frac{\partial^2 F^{\text{el}}(\{\mathbf{R}_I\}, V, T)}{\partial R_k \partial R_l} \right]_{\{\mathbf{R}_I^0\}} \\ \rightarrow D(V, T) \mathbf{w}_i &= \omega_i^2(V, T) \mathbf{w}_i \\ \rightarrow E_{\{n_i\}}^{\text{qh}}(V, T) &= F_0^{\text{el}}(V, T) + \sum_i \hbar \omega_i(V, T) \left(n_i + \frac{1}{2} \right), \end{aligned} \quad (15.32)$$

with n_i the number of phonons in state i . In Eq. (15.32), the dynamical matrix D (with elements $D_{k,l}$) has been defined, which corresponds to the second derivative of F^{el} scaled by the masses M_k of the nuclei. The eigenvalue equation of D with eigenvectors \mathbf{w}_i and eigenvalues ω_i^2 defines the phonon frequencies ω_i . The energy $E_{\{n_i\}}^{\text{qh}}$ for a certain fixed phonon occupation configuration $\{n_i\}$ is given by a sum over the frequencies weighted by the corresponding occupation numbers. Note the important point that the phonon frequencies are not only volume dependent (quasiharmonic) but also explicitly temperature dependent through the temperature dependence of F^{el} . This temperature dependence does not correspond to an anharmonic atomic (*i.e.*, phonon–phonon) interaction. To make this point explicit, we will use the notation T^{el} for this temperature, while the temperature determining the thermodynamics of the nuclei will be denoted by T^{muc} . In fact, to speed up numeric convergence, T^{el} and T^{muc} can be varied independently of each other. Of course, at the end of the calculation both have to be ensured to be equal to the actual external temperature, *i.e.*, $T^{\text{el}} = T^{\text{muc}} = T$.

The final step of the quasiharmonic approximation is to approximate the eigenvalues $\tilde{E}_\mu^{\text{muc}}$ in Eq. (15.28) by the quasiharmonic energy $E_{\{n_i\}}^{\text{qh}}$ which yields (the tilde in

\tilde{F} indicating the approximation at this stage)

$$\tilde{F}(V, T) = [\tilde{F}(V, T^{\text{el}}, T^{\text{nuc}})]_{T^{\text{el}}=T^{\text{nuc}}=T}, \quad (15.33)$$

with

$$\tilde{F}(V, T^{\text{el}}, T^{\text{nuc}}) \approx -k_B T \ln \sum_{\{n_i\}} e^{-\beta \tilde{E}_{\{n_i\}}^{\text{qh}}} = \dots = F_0^{\text{el}}(V, T^{\text{el}}) + F^{\text{qh}}(V, T^{\text{el}}, T^{\text{nuc}}), \quad (15.34)$$

where the dots denote a series of straightforward transformations (see *e.g.*, Ref. [26]) and with

$$F^{\text{qh}}(V, T^{\text{el}}, T^{\text{nuc}}) = \frac{1}{N} \sum_i \left[\overbrace{\frac{\hbar}{2} \omega_i(V, T^{\text{el}})}^{T^{\text{nuc}}=0\text{K} \text{ zero point contr.}} + T^{\text{nuc}} \underbrace{k_B \ln[1 - \exp\{-\beta \hbar \omega_i(V, T^{\text{el}})\}]}_{\text{(negative) entropic contribution.}} \right]. \quad (15.35)$$

Here, N is an appropriately chosen scaling factor which is, *e.g.*, the number of sampled frequencies divided by three (number of spatial dimensions) if F^{qh} should refer to a “per atom” quantity. Note that F^{qh} fully contains quantum mechanical effects (*e.g.*, zero point vibrations) in contrast to a free energy obtained from a classical molecular dynamics run.

The quasiharmonic equations presented above are general and equally well applicable to defect and bulk cells. In practice, the calculation of the quasiharmonic free energy contribution of the defect cell requires special attention to ensure a consistent treatment with the corresponding perfect bulk cell. The reason for this is the break of translational symmetry introduced by the presence of the defect. As a consequence, the commonly applied [19] Fourier interpolation to generate dense wave vector meshes in the Brillouin zone cannot be employed to the defect cell. To nonetheless profit from the advantage of a Fourier interpolation for the perfect bulk (*e.g.*, a significantly improved description of the low temperature free energy) a correction scheme as *e.g.*, proposed in Ref. [18] should be used.

15.2.2.4 Anharmonic Atomic Excitations: Thermodynamic Integration

Let us now consider the higher order terms in Eq. (15.30). An intuitive and straightforward approach to include these terms would be an *ab initio* based classical molecular dynamics run. The fact that a direct computation of the free energy employing conventional molecular dynamics is not feasible (the free energy is an entropic quantity [27]) could be circumvented by calculating the inner energy and integrating it with respect to temperature. It turns out that such a “naive” phase space sampling leads to infeasibly long computational times [18]. Therefore, the development and application of highly efficient sampling strategies to perform the thermodynamic averages is crucial.

A fundamental concept is thermodynamic integration. The key idea is to start from a reference with an analytically known free energy or with a free energy that can be obtained numerically extremely fast. This reference is coupled adiabatically to the true *ab initio* potential energy surface and only the difference in free energies is sampled. Such an approach will be efficient, if one manages to construct a reference which closely approximates the true *ab initio* potential energy surface thus yielding a small free energy difference.

A good reference to perform thermodynamic integration is the quasiharmonic potential energy surface. Therefore, a system with free energy F_λ is introduced which couples the electronic + quasiharmonic system [having the free energy \tilde{F} as given in Eq. (15.34)] to the full system (having free energy F) by the adiabatic switching parameter λ . The boundary conditions are defined to be

$$[F_\lambda]_{\lambda=1} = F \quad \text{and} \quad [F_\lambda]_{\lambda=0} = \tilde{F}, \quad (15.36)$$

so that the anharmonic free energy is given by

$$F^{\text{ah}} = F - \tilde{F} = [F_\lambda]_{\lambda=1} - [F_\lambda]_{\lambda=0}. \quad (15.37)$$

In principle, methods based on the quantum-classical isomorphism exist allowing to perform thermodynamic integration including quantum mechanical effects [28]. Unfortunately, their actual application requires very large computational efforts and at present only investigations employing empirical potential energy surfaces can be afforded [28]. In practice, a quantum mechanical treatment of F^{ah} is not expected to be necessary since the quasiharmonic free energy contains the major part of quantum mechanical effects (as derived in the previous section). The reason for this is that at low temperatures, where quantum effects are important, the quasiharmonic approximation is excellent thus fully accounting for such effects, while at high temperatures where the quasiharmonic approximation fails and anharmonicity becomes significant quantum effects become small/negligible. The anharmonic contribution can therefore be calculated in the classical limit (superscript “clas;” note that correspondingly a classical quasiharmonic reference is used):

$$\begin{aligned} F^{\text{clas,ah}} &:= F^{\text{clas}} - (F_0^{\text{el}} + F^{\text{clas,qh}}) = [F_\lambda^{\text{clas}}]_{\lambda=1} - [F_\lambda^{\text{clas}}]_{\lambda=0} = \int_0^1 d\lambda \frac{\partial F_\lambda^{\text{clas}}}{\partial \lambda} \\ &\stackrel{\text{Eq. (15.39)}}{=} \int_0^1 d\lambda \left\langle \frac{\partial F_\lambda^{\text{el}}}{\partial \lambda} \right\rangle_{T,\lambda} \stackrel{\text{ergodicity hypothesis}}{=} \int_0^1 d\lambda \left\langle \frac{\partial F_\lambda^{\text{el}}}{\partial \lambda} \right\rangle_{t,\lambda}. \end{aligned} \quad (15.38)$$

Here, F_λ^{el} is the λ dependent electronic free energy surface determining the classical motion of the nuclei in the coupled system, $\langle \cdot \rangle_{T,\lambda}$ denotes the thermodynamic, and $\langle \cdot \rangle_{t,\lambda}$ the time average at a given λ . To obtain the first equality in the third line, we have used:

$$\begin{aligned} \frac{\partial F_{\lambda}^{\text{clas}}}{\partial \lambda} &= \frac{1}{Z_{\lambda}^{\text{clas}}} \int_{\Omega^{\text{ph}}} \frac{d\mathbf{R}_I}{e^{-\beta F_{\lambda}^{\text{el}}(\{\mathbf{R}_I\})}} \frac{\partial F_{\lambda}^{\text{el}}}{\partial \lambda} \\ &= \left\langle \frac{\partial F_{\lambda}^{\text{el}}}{\partial \lambda} \right\rangle_{T, \lambda}, \end{aligned} \quad (15.39)$$

with the standard definitions of a classical free energy F^{clas} , partition function Z^{clas} , and phase space volume Ω^{ph} . Besides the fixed boundary conditions, Eq. (15.36), any type of coupled system can be chosen. In practice, a simple linear coupling to the quasiharmonic reference (restoring the explicit notation for the volume and temperature dependencies),

$$\begin{aligned} F_{\lambda}^{\text{el}}(\{\mathbf{R}_I\}, V, T^{\text{el}}) &= \lambda F^{\text{el}}(\{\mathbf{R}_I\}, V, T^{\text{el}}) \\ &+ (1 - \lambda) \left[F_0^{\text{el}}(V, T^{\text{el}}) + \sum_{k,l} \frac{\sqrt{M_k M_l}}{2} u_k u_l D_{k,l}(V, T^{\text{el}}) \right], \end{aligned} \quad (15.40)$$

yields computationally efficient results. Finally, the anharmonic free energy reads:

$$\begin{aligned} F^{\text{clas,ah}}(V, T^{\text{el}}, T^{\text{nuc}}) &= \int_0^1 d\lambda \left\langle F^{\text{el}}(V, T^{\text{el}}) - F_0^{\text{el}}(V, T^{\text{el}}) \right. \\ &\quad \left. - \sum_{k,l} \frac{\sqrt{M_k M_l}}{2} u_k u_l D_{k,l}(V, T^{\text{el}}) \right\rangle_{t, \lambda}. \end{aligned} \quad (15.41)$$

Note that the dependence of $F^{\text{clas,ah}}$ on T^{nuc} is hidden in the time average $\langle \cdot \rangle_t$ which needs to be obtained e.g. from a molecular dynamics simulation. Combining the various contributions, the free energy of the system is given by:

$$\begin{aligned} F &= E_{g,0}^{\text{el}} + \tilde{F}_0^{\text{el}} + F^{\text{qh}} + F^{\text{clas,ah}} \\ &= [E_{g,0}^{\text{el}}(V) + \tilde{F}_0^{\text{el}}(V, T^{\text{el}}) + F^{\text{qh}}(V, T^{\text{el}}, T^{\text{nuc}}) \\ &\quad + F^{\text{clas,ah}}(V, T^{\text{el}}, T^{\text{nuc}})]_{T^{\text{el}}=T^{\text{nuc}}=T}. \end{aligned} \quad (15.42)$$

We stress that the dominant part of quantum mechanical effects contributing to F is accounted for by F^{qh} . For the case of (wide band gap) semiconductors the influence of T^{el} will be small (*cf.* Section 15.2.2.2) and the free energy can be approximated by:

$$F(V, T) \approx [E_{g,0}^{\text{el}}(V) + F^{\text{qh}}(V, 0\text{K}, T^{\text{nuc}}) + F^{\text{clas,ah}}(V, 0\text{K}, T^{\text{nuc}})]_{T^{\text{nuc}}=T}. \quad (15.43)$$

15.2.2.5 Anharmonic Atomic Excitations: Beyond the Thermodynamic Integration

While the thermodynamic integration approach boosts the efficiency by a few orders of magnitude as compared to conventional molecular dynamics simulations, the calculation of a full free energy surface $F(V, T)$ from *ab initio* is still a formidable task

even on today's high performance CPU architecture. Therefore, there are currently active efforts at exploring new methods to reduce computational time [18, 29, 30].

Wu and Wentzcovitch [29], *e.g.*, developed a semi empirical ansatz based on a single adjustable parameter to describe anharmonicity. To explain their approach, it is useful to note first that already the quasiharmonic frequencies ω_i^{qh} contain an implicit temperature dependence due to thermal expansion: $\omega_i^{\text{qh}} = \omega_i^{\text{qh}}(V^{\text{eq}}(T))$. Here, $V^{\text{eq}}(T)$ denotes the thermal expansion, *i.e.*, the equilibrium volume as a function of temperature T (for a fixed pressure). The thermal expansion is itself determined by the ω_i^{qh} so the problem must be solved self consistently.

The ansatz of Wu and Wentzcovitch [29] is to modify the implicit temperature dependence. Specifically, the following transformation to a renormalized frequency ω_i^{renorm} is proposed:

$$\omega_i^{\text{qh}}(V^{\text{eq}}(T)) \rightarrow \omega_i^{\text{renorm}}(T) := \omega_i^{\text{qh}}(V'(T)), \quad (15.44)$$

with a modified thermal expansion

$$V'(T) = V^{\text{eq}}(T) \left\{ 1 - C \left[\frac{V^{\text{eq}}(T) - V_0^{\text{eq}}}{V_0^{\text{eq}}} \right] \right\}, \quad (15.45)$$

where C is an adjustable parameter and $V_0^{\text{eq}} = V^{\text{eq}}(T = 0\text{K})$. The renormalized frequencies ω_i^{renorm} replace then the quasiharmonic frequencies in the quasiharmonic free energy expression Eq. (15.35). The resulting free energy surface includes therefore anharmonicity in an approximative manner and can be used to derive anharmonic thermodynamic quantities such as thermal expansion or defect concentrations. In the original work [29], C is adjusted by fitting to experimental data. In a more recent work [30], it was shown how the parameter can be obtained using *ab initio* based thermodynamic integration. The main advantage of the approach is its effectiveness, since to determine C and thus to generate a complete anharmonic free energy surface only a single thermodynamic integration run at fixed V and T suffices. The disadvantage is however that the ω_i^{renorm} (and thus the anharmonic free energy) are assumed to have a certain temperature dependence, specifically that they scale linearly with the quasiharmonic thermal expansion. This assumption has been shown to work well for diamond and MgO [30] up to high temperatures, it cannot be however *a priori* assumed in general and further evaluations are necessary.

An alternative approach that is able to account for the full anharmonic free energy surface without assuming any specific temperature dependence was developed recently in Ref. [18] and termed upsampled thermodynamic integration using Langevin dynamics (UP-TILD) method. The key ideas are as follows. In a first step, a usual thermodynamic integration run for discrete values of λ is performed:

$$\left\langle \frac{\partial F_{\lambda}^{\text{el}}}{\partial \lambda} \right\rangle_{t,\lambda}^{\text{low}} = \left\langle F^{\text{el}}(V, T^{\text{el}}) - F_0^{\text{el}}(V, T^{\text{el}}) - \sum_{k,l} \frac{\sqrt{M_k M_l}}{2} u_k u_l D_{k,l}(V, T^{\text{el}}) \right\rangle_{t,\lambda}^{\text{low}}. \quad (15.46)$$

A critical insight gained in Ref. [18] is that instead of employing fully converged DFT parameters (*e.g.*, with respect to basis set or electronic k sampling) computationally much less demanding DFT parameters can be used (indicated by the superscript “low”). In particular, they need to be chosen such that the corresponding phase space distribution (termed $\{\mathbf{R}_I\}_t^{\text{low}}$ in the following) closely resembles the phase space distribution $\{\mathbf{R}_I\}_t^{\text{high}}$ as would be obtained from fully/highly converged parameters. As a consequence of using the low converged parameters, the thermodynamic integration is computationally very efficient and speed up factors of ≈ 30 are achievable. Applying this approach makes it therefore possible to sample various λ , V , and T values even on modest computer resources. The resulting free energy surface, *i.e.*, $\langle \partial F_\lambda^{\text{el}} / \partial \lambda \rangle_{t,\lambda}^{\text{low}}$, needs however to be corrected in a second step. The actual correction step is straightforward, provided that the above stated condition $\{\mathbf{R}_I\}_t^{\text{low}}$ close to $\{\mathbf{R}_I\}_t^{\text{high}}$ holds. In such a case, a typically small set of N^{UP} uncorrelated structures $\{\mathbf{R}_I\}_t^{\text{low}}$ (indexed with t_u) is extracted from $\{\mathbf{R}_I\}_t^{\text{low}}$ and the upsampling average $\langle \Delta F_\lambda^{\text{el}} \rangle_\lambda^{\text{UP}}$ is calculated as:

$$\begin{aligned} \langle \Delta F_{\text{el}} \rangle_\lambda^{\text{UP}} &= \frac{1}{N^{\text{UP}}} \sum_u^{N^{\text{UP}}} F^{\text{el,low}}(\{\mathbf{R}_I\}_{t_u}^{\text{low}}) - F^{\text{el,low}}(\{\mathbf{R}_I^0\}) \\ &\quad - \left[F^{\text{el,high}}(\{\mathbf{R}_I\}_{t_u}^{\text{low}}) - F^{\text{el,high}}(\{\mathbf{R}_I^0\}) \right]. \end{aligned} \quad (15.47)$$

Here, $F^{\text{el,low}}$ ($F^{\text{el,high}}$) refers to the electronic free energy calculated using the low (high) converged set of DFT parameters. The λ dependence of $\langle \Delta F_\lambda^{\text{el}} \rangle_\lambda^{\text{UP}}$ is hidden in the trajectory $\{\mathbf{R}_I\}_t^{\text{low}}$, which is additionally dependent on the volume and temperature. In the last step, the quantity of interest, the converged $\langle \partial F_\lambda^{\text{el}} / \partial \lambda \rangle_{t,\lambda}^{\text{high}}$ is obtained

$$\langle \partial F_\lambda^{\text{el}} / \partial \lambda \rangle_{t,\lambda}^{\text{high}} = \langle \partial F_\lambda^{\text{el}} / \partial \lambda \rangle_{t,\lambda}^{\text{high}} - \langle \Delta F_\lambda^{\text{el}} \rangle_\lambda^{\text{UP}},$$

and thus the anharmonic free energy:

$$F^{\text{clas,ah}} = \int_0^1 d\lambda \langle \partial F_\lambda^{\text{el}} / \partial \lambda \rangle_{t,\lambda}^{\text{high}}. \quad (15.48)$$

The efficiency of the UP-TILD method is a direct consequence of the fact that the upsampling average $\langle \Delta F_\lambda^{\text{el}} \rangle_\lambda^{\text{UP}}$, which involves the computation of the CPU time consuming well converged DFT free energies $F^{\text{el,high}}$, converges extremely fast with respect to the number of uncorrelated configurations. In practice, less than 100 configurations are found to be sufficient to achieve an accuracy of better than 1 meV/atom in the free energy [18]. This small number of configurations has to be compared with the number of molecular dynamics steps needed to converge a thermodynamic integration run which is in the range of several thousands, *i.e.*, two orders of magnitude larger. Another advantage is that in many cases the λ dependence is invariant with respect to the upsampling procedure allowing to calculate the shift in Eq. (15.48) for a single λ value only. Using this invariance reduces computational costs further as illustrated in Figure 15.4.

The various methods needed to compute the electronic, quasiharmonic, and anharmonic contributions to the free energy surface are summarized in Figure 15.5.

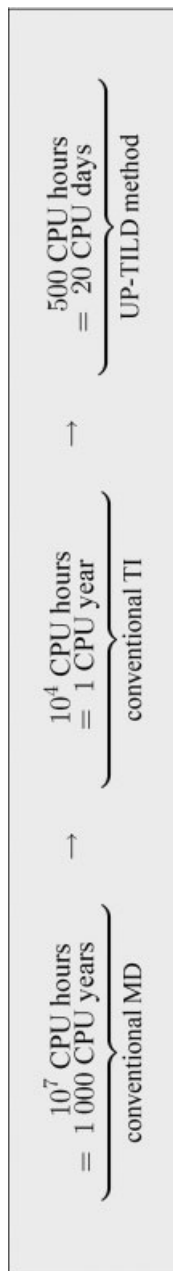


Figure 15.4 (online colour at: www.pss-b.com) CPU time reductions in calculating the anharmonic free energy as reported for the example of bulk aluminum including point defects (32 atomic fcc cell) [18]; MD: molecular dynamics; TI: thermodynamic integration; UP-TILD: upsampled TI using Langevin dynamics (see text for details).

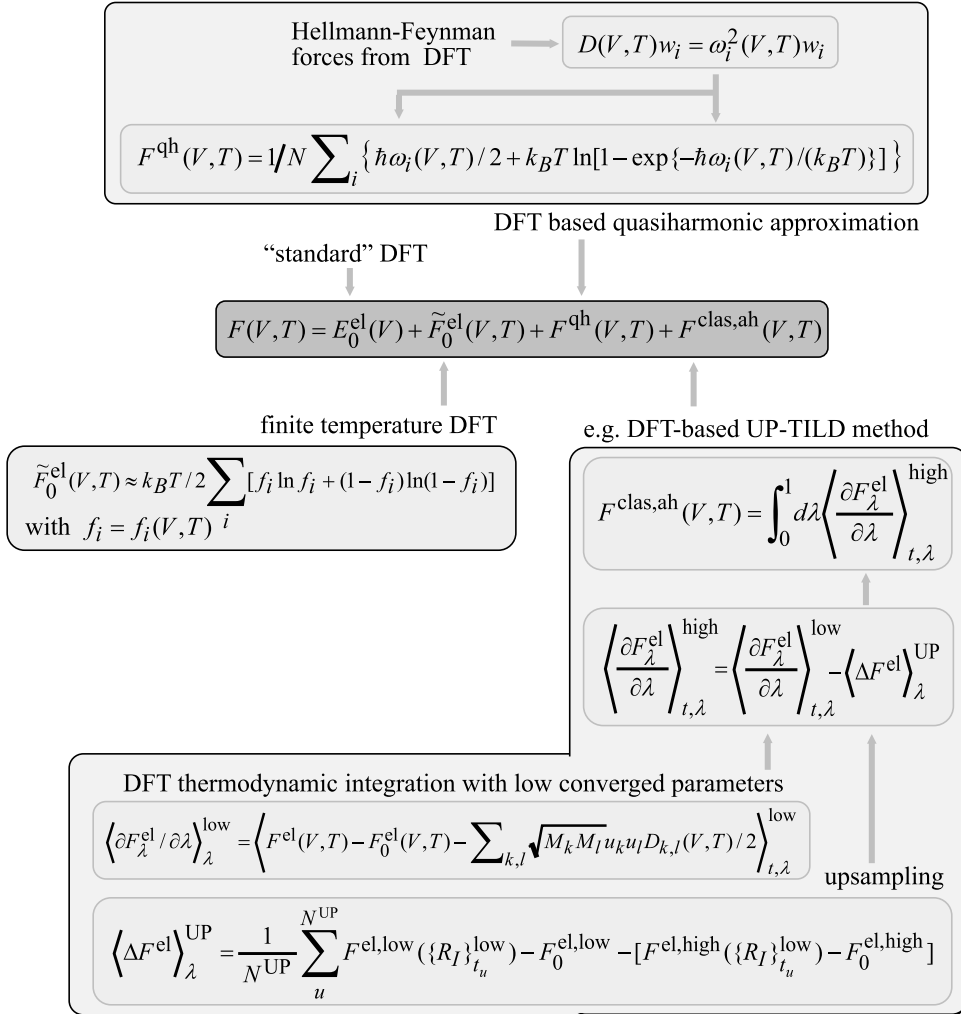


Figure 15.5 (online colour at: www.pss-b.com) Schematic illustration of the overall approach and its key equations to derive finite temperature contributions to the free energy presented in this article. The free energy surface $F(V, T)$ refers here either to the defect cell free

energy surface $F^{\text{d}}(V, T)$ or the perfect bulk free energy surface $F^{\text{p}}(V, T)$ as used in Section 15.2.1. The corresponding methods for each component are indicated. Definitions are given in Sections 2.2.2–2.2.4.

15.3

Results: Electronic, Quasiharmonic, and Anharmonic Excitations in Vacancy Properties

To discuss the performance of the methodology described in Section 15.2, we consider point defects in aluminum. The focus will be on vacancies, since self

interstitials are practically absent due to their high formation energy (3.4 eV) [31]. Computational details of the results presented in the following can be found in Refs. [18, 26].

Before applying the approach to defect properties, we will first consider bulk properties. This will allow a careful inspection and evaluation of the accuracy of the approach and the underlying exchange-correlation functionals. For the following discussion, we restrict on two properties that are highly sensitive to an accurate description of the free energy surface: The thermal expansion coefficient and the isobaric heat capacity. These quantities are first or higher order derivatives of the free energy surface being thus affected even by small changes in the free energy. As a consequence, to guarantee an unbiased comparison with experiment the error bar in the free energy has to be systematically kept < 1 meV/atom. This error bar is significantly lower than what is typically targeted at in defect calculations (≈ 0.1 eV) and particularly challenging to achieve at high temperatures. A major advantage of reaching this numerical accuracy is that the remaining discrepancy with experiment can be unambiguously related to deficiencies in the exchange-correlation functional. For the case considered here, fcc bulk aluminum, both of the most popular exchange-correlation functionals, the LDA and the GGA, show an excellent agreement with experiment for the expansion coefficient up to the melting point (Figure 15.6a) and for the heat capacity up to ≈ 500 K (Figure 15.6b). Above this temperature, the experimental scatter in the heat capacity becomes too large and prevents an unbiased comparison. Nonetheless, the *ab initio* results indicate a lower bound to experiment hinting at additional and apparently so far not controllable by experiment excitation mechanisms in the sample and/or experimental setup.

Having demonstrated the accuracy achievable by the method to describe bulk properties at finite temperatures, let us now turn to defect properties. Applying the same formalism as for the bulk and the constant pressure approach, the temperature and volume dependent isobaric defect formation free energy $F_P^f(V, T)$ can be calculated (see Section 15.2.1.2). Using it, the equilibrium defect concentration $c^{\text{eq}}(T, P) = \exp[-F_P^f(V^{\text{eq}}(T, P), T)/(k_B T)]$ is obtained. Here, $V^{\text{eq}}(T, P)$ is the equilibrium volume at temperature T and pressure P . The computed concentrations at finite temperatures and for zero pressure (the influence of atmospheric pressure is negligible) are shown together with available experimental data in Figure 15.7.

As can be seen, an excellent agreement is obtained. This agreement is particularly impressive when considering that errors in the defect formation free energy scale exponentially in the concentration. An important finding is that LDA and GGA provide an upper and lower bound to the experimental data and may thus be used as empirical (but *ab initio* computable) error bars. This interesting and highly useful behavior is not restricted to Al but has been systematically observed for a wide range of metals when considering temperature dependent material properties [19]. Figure 15.7 provides also a direct insight into the relevance of effects due to the finite size of practical supercells as discussed in Section 15.2.1. As can be seen, correcting these effects using the constant pressure approach increases the defect concentration by almost an order of magnitude. Therefore, the application of the constant pressure approach is critical to achieve the desired accuracy. A further extension of the

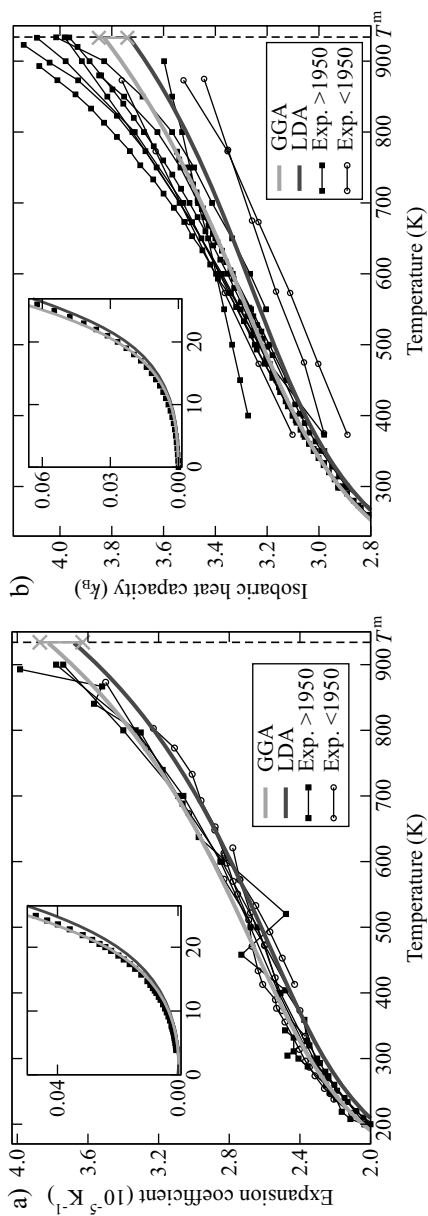


Figure 15.6 (online colour at: www.pss-b.com) (a) Thermal expansion coefficient and (b) isobaric heat capacity of aluminum including the electronic, quasiharmonic, anharmonic, and vacancy contribution compared to experiment. The melting temperature T^m of Al (933 K) is given by the vertical dashed line. At T^m , the crosses indicate the sum of all numerical errors (e.g., pseudopotential error or statistical inaccuracy; cf. Ref. [26]) in all

contributions for GGA. The LDA error is of the same order of magnitude. Experimental data older than 1950 are indicated by open circles (Refs. [32–35] for the expansion coefficient and [36–38] for the heat capacity). The remaining experimental results are indicated by filled squares (Refs. [39–44] and [45–53]). The inset shows the low temperature region with experimental data from Refs. [54, 51].

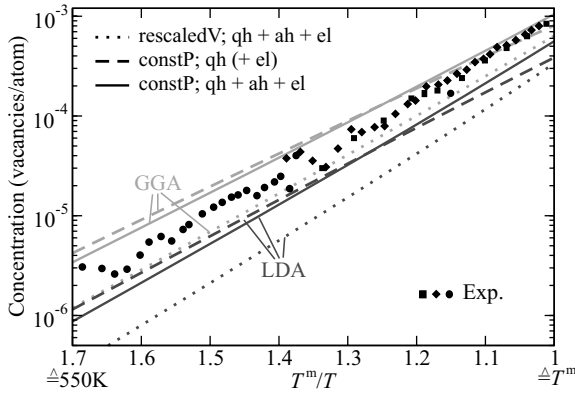


Figure 15.7 (online colour at: www.pss-b.com) Equilibrium vacancy concentration at zero pressure of aluminum as a function of the inverse temperature multiplied by the melting temperature T^m . Results for the rescaled volume and constant pressure approach are shown. The volume optimized approach yields concentrations which are identical to the

constant pressure results on the shown scale. The electronic contribution yields a negligible contribution (indicated by the parenthesis). The squares indicate experimental values from Ref. [55] (differential dilatometry). The diamonds/circles indicate experimental values from Ref. [56] (differential dilatometry/positron annihilation).

methodology to the volume optimized approach has a negligible effect on the vacancy concentration shown in Figure 15.7.

To analyze the temperature dependence of the defect concentration in more detail, it is convenient to express the formation free energy in terms of the $T = 0$ K enthalpy and the entropy of formation. To provide a direct comparison with experimental data where only a rather small temperature window is available, we restrict on the experimental temperature interval and obtain both quantities from a linear regression in the $\log-1/T$ plot (Figure 15.7). The results are summarized in Table 15.2. A surprising finding is that including anharmonicity has a major effect on the entropy and enthalpy: For LDA the entropy of formation increases from $0.2k_B$ (quasiharmonic) to $2.2k_B$ (anharmonic) and the enthalpy from 0.65 to 0.78 eV. These numbers are in excellent agreement with experimentally derived data of $2.4k_B$ and 0.75 eV.

It is interesting to note that the substantial deviations due to anharmonic effects have little consequence on the absolute defect concentrations (see Figure 15.7). The changes in entropy and enthalpy largely compensate each other in this quantity. An important conclusion from this result is that the defect formation enthalpy derived from high temperature data (the only region where experimental data is available) can be substantially different from the true $T = 0$ K formation enthalpy. For the system considered here the difference is ≈ 0.1 eV. To guarantee an accurate comparison between theory and (high temperature) experimental data it is therefore critical to not restrict to the formation enthalpy (as usually done) but to use also the experimental entropy to compute and compare defect concentrations.

Table 15.2 The extrapolated formation energy E^f and averaged entropy of formation S^f for various approaches and combinations of the free energy contributions used for the calculation of vacancy properties of aluminum. ai/ep indicates values for the coupled *ab initio*-empirical potentials approach from Ref. [16]. The further values (also the experimental) are obtained by fitting the vacancy concentrations over the temperature range given in Figure 15.7 to the function $\exp[-(E^f - TS^f)/(k_B T)]$. The notation is as in Figure 15.7.

| | E^f (eV) | | S^f (k_B) | |
|----------------------|------------|------|-----------------|-----|
| | LDA | GGA | LDA | GGA |
| constP; qh | 0.65 | 0.58 | 0.2 | 0.1 |
| constP; qh + el | 0.65 | 0.58 | 0.2 | 0.1 |
| constP; qh + ah + el | 0.78 | 0.68 | 2.2 | 1.5 |
| volOpt; qh + ah + el | 0.78 | 0.68 | 2.2 | 1.5 |
| constV; qh + ah + el | 0.85 | 0.75 | 2.5 | 1.9 |
| ai/ep [16] | 0.78 | 0.61 | 1.6 | 1.3 |
| experiment | 0.75 | | 2.4 | |

15.4

Conclusions

In this paper, we gave a brief overview about the challenges one encounters when including temperature effects in defect calculations beyond configurational entropy. The paper outlines strategies to compute all relevant free energy contributions arising due to electronic and vibronic (quasiharmonic and anharmonic) excitations. A major focus has been devoted to numerical performance since even on today's most powerful supercomputers such calculations quickly approach the limits of available resources when attempting a full *ab initio* description. As an example which numerical accuracy can be achieved, a simple yet instructive defect system has been considered: vacancies in fcc bulk Al. The results nicely illustrate that both bulk and defect properties in the absence of any band gap problem can be described with a surprising accuracy and predictive power. An interesting result of this study is that anharmonic contributions have a rather small effect on the vacancy concentration in the experimentally accessible temperature window, but have a drastic effect on the averaged/extrapolated entropy and enthalpy of formation. To achieve an accuracy of better than 0.1 eV the inclusion of anharmonic contributions is mandatory. With the advent of powerful approaches to overcome the band gap problem in semiconductor defect calculations, we believe that inclusion of finite temperature effects (possibly on the LDA/GGA level) becomes critical.

References

- 1 Estreicher, S.K., Backlund, D., and Gibbons, T. M. (2010) *Thin Solid Films*, 518, 2413.
- 2 Al-Mushadani, O.K. and Needs, R.J. (2003) *Phys. Rev. B*, 68, 235205.

- 3 Estreicher, S.K., Sanati, M., West, D., and Ruymgaart, F. (2004) *Phys. Rev. B*, **70**, 125209.
- 4 Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2003) *J. Chem. Phys.*, **118**, 8207.
- 5 Heyd, J., Peralta, J.E., Scuseria, G.E., and Martin, R.L. (2005) *J. Chem. Phys.*, **123**, 174101.
- 6 van Schilfgaarde, M., Kotani, T., and Faleev, S. (2006) *Phys. Rev. Lett.* **96**, 226402.
- 7 Rinke, P., Qteish, A., Neugebauer, J., and Scheffler, M. (2008) *Phys. Status Solidi B*, **245**, 929.
- 8 Gillan, M. J. (1989) *J. Phys.: Condens. Matter*, **1**, 689.
- 9 Pawellek, R., Fähnle, M., Elsässer, C., Ho, K.M., and Chan, C.T. (1991) *J. Phys.: Condens. Matter*, **3**, 2451.
- 10 Polatoglou, H.M., Methfessel, M., and Scheffler, M. (1993) *Phys. Rev. B*, **48**, 1877.
- 11 Korhonen, T., Puska, M.J., and Nieminen, R.M. (1995) *Phys. Rev. B*, **51**, 9526.
- 12 Turner, D.E., Zhu, Z.Z., Chan, C.T., and Ho, K.M. (1997) *Phys. Rev. B*, **55**, 13842.
- 13 Satta, A., Willaime, F., and de Gironcoli, S. (1998) *Phys. Rev. B*, **57**, 11184.
- 14 Satta, A., Willaime, F., and de Gironcoli, S. (1999) *Phys. Rev. B*, **60**, 7001.
- 15 Carling, K. M., Wahnström, G., Mattsson, T. R., Mattsson, A. E., Sandberg, N., and Grimvall, G. (2000) *Phys. Rev. Lett.*, **85**, 3862.
- 16 Carling, K. M., Wahnström, G., Mattsson, T. R., Sandberg, N., and Grimvall, G. (2003) *Phys. Rev. B*, **67**, 054101.
- 17 Lucas, G. and Schaublin, R. (2009) *Nucl. Instrum. Methods B*, **267**, 3009.
- 18 Grabowski, B., Ismer, L., Hickel, T., and Neugebauer, J. (2009) *Phys. Rev. B*, **79**, 134106.
- 19 Grabowski, B., Hickel, T., and Neugebauer, J. (2007) *Phys. Rev. B*, **76**, 024309.
- 20 Keer, H. V. (1993) *Principles of the Solid State* (Wiley, New York).
- 21 Varotsos, P. A. and Alexopoulos, K. D. (1986) *Thermodynamics of Point Defects and Their Relation with Bulk Properties, Defects in Solids*, Vol. 14 (North-Holland, Amsterdam).
- 22 Born, M. and Oppenheimer, J. R. (1927) *Ann. Phys. (Leipzig)* **389**, 457.
- 23 Cao, J. and Berne, B. J. (1993) *J. Chem. Phys.*, **99**, 2902.
- 24 Suzuki, M. (1977) *Commun. Math. Phys.*, **57**, 193.
- 25 Mermin, N. D. (1965) *Phys. Rev.*, **137**, A1441.
- 26 Grabowski, B. (2009) Towards ab initio assisted materials design: DFT based thermodynamics up to the melting point, OT *PhD thesis*, Universität Paderborn.
- 27 Haile, J. M. (1997) *Molecular Dynamics Simulation: Elementary Methods* (Wiley, New York).
- 28 Ramirez, R., Herrero, C. P., Antonelli, A., and Hernández, E. R. (2008) *J. Chem. Phys.*, **129**, 064110.
- 29 Wu, Z. and Wentzcovitch, R. M. (2009) *Phys. Rev. B*, **79**, 104304.
- 30 Wu, Z. (2010) *Phys. Rev. B*, **81**, 172301.
- 31 Denteneer, P. J. H. and Soler, J. M. (1991) *J. Phys.: Condens. Matter*, **3**, 8777.
- 32 Honda, K. and Okubo, Y. (1924) *Sci. Rep. Tohoku Imp. Univ., Ser. 1* **13**, 101 (cited from Ref. [42]).
- 33 Uffelmann, F. L. (1930) *Philos. Mag.*, **10**, 633 (cited from Ref. [42]).
- 34 Nix, F. C. and MacNair, D. (1941) *Phys. Rev.*, **60**, 597.
- 35 Wilson, A. J. C. (1942) *Proc. Phys. Soc.*, **54**, 487.
- 36 Eastman, E. D., Williams, A. M., and Young, T. F. (1924) *J. Am. Chem. Soc.*, **46**, 1178.
- 37 Umino, S. (1926) *Sci. Rep. Tohoku Imp. Univ., Ser. 1* **15**, 597 (cited from Ref. [52]).
- 38 Avramescu, A. (1939) *Z. Tech. Phys. (Leipzig)* **20**, 213.
- 39 Nicklow, R. M. and Young, R. A. (1963) *Phys. Rev.*, **129**, 1936 (cited from Ref. [42]).
- 40 Strelkov, P. G. and Novikova, S. I. (1957) *Prib. Tekh. Eksp. (USSR)* **5**, 105 (cited from Ref. [42]).
- 41 Pathak, P. D. and Vasavada, N. G. (1970) *J. Phys. C, Solid State Phys.*, **3**, L44 (cited from Ref. [42]).
- 42 Touloukian, Y. S., Kirby, R. K., Taylor, R. E., and Desai, P. D. (1975) *Thermal Expansion: Metallic Elements and Alloys, Thermophysical Properties of Matter*, Vol. 12 (IFI/Plenum, New York).

- 43 von Guérard, B., Peisl, H., and Zitzmann, R. (1974) *Appl. Phys.*, **3**, 37, numerical differentiation of the given experimental expansion data; strongly scattering points were taken out.
- 44 Kroeger, F. R. and Swenson, C. A. (1977) *J. Appl. Phys.*, **48**, 853.
- 45 Pochapsky, T. E. (1953) *Acta Metall.* **1**, 747.
- 46 McDonald, R. A. (1967) *J. Chem. Eng. Data* **12**, 115.
- 47 Brooks, C. R. and Bingham, R. E. (1968) *J. Phys. Chem. Solids*, **29**, 1553.
- 48 Leadbetter, A. J. (1968) *J. Phys. C, Solid State Phys.*, **1**, 1481.
- 49 Schmidt, U., Vollmer, O., and Kohlhaas, R. (1970) *Z. Naturforsch. A* **A25**, 1258.
- 50 Marchidan, D. I. and Ciopec, M. (1970) *Rev. Roum. Chim.*, **15**, 1005 (cited from Ref. [52]).
- 51 Ditmars, D. A., Plint, C. A., and Shukla, R. C. (1985) *Int. J. Thermophys.*, **6**, 499.
- 52 Desai, P. D. (1987) *Int. J. Thermophys.* **8**, 621.
- 53 Takahashi, Y., Azumi, T., and Sekine, Y. (1989) *Thermochim. Acta*, **139**, 133.
- 54 McLean, K. O., (1969) Low Temperature Thermal Expansion of Copper, Silver, Gold, and Aluminum, OT *PhD thesis*, Iowa State University (cited from Ref. [42]).
- 55 Simmons, R. O. and Balluffi, R. W. (1960) *Phys. Rev.*, **117**, 52.
- 56 Hehenkamp, T. (1994) *J. Phys. Chem. Solids*, **55**, 907.