

Equilibrium point defect and charge carrier concentrations in a material determined through calculation of the self-consistent Fermi energy

J. Buckeridge^{a,*}

^a*University College London, Department of Chemistry, 20 Gordon Street, London WC1H 0AJ, United Kingdom*

Abstract

A concise procedure to determine the self-consistent Fermi energy and defect and carrier concentrations in an extended crystalline system is presented. It is assumed that the formation enthalpies of a set of variously charged point defects in thermodynamic equilibrium are known, as well as the density of electronic states in the defect-free system. By applying the constraint of overall charge neutrality, the self-consistent Fermi energy is determined using an iterative searching routine. The procedure is incorporated within a Fortran code ‘SC-FERMI’: the input consists of the defect formation energies, density of sites where they can form, and the degeneracy of each charge state; the material band gap; and the calculated density of states of the pristine system. The output is the self-consistent Fermi energy, the total concentrations of each defect as well as the concentration of its individual charge states, and the free carrier concentrations. Furthermore, the procedure facilitates fixing the concentration of one or more defects and determining the resulting self-consistent Fermi energy and concentrations of other defects (performed using the related code ‘FROZEN-SC-FERMI’), thus modelling ‘frozen-in’ defects which may form by kinetic, rather than thermodynamic, processes. One can fix the total concentration or the concentration of a particular charge state; it is also possible to introduce new defects with a fixed concentration, but here the charge state must be specified. The background theory is discussed in some detail, and the operation of the program is demonstrated by some

*Corresponding author.

E-mail address: j.buckeridge@ucl.ac.uk

examples.

Keywords: Crystal defects, Defect concentrations, Self-consistent Fermi energy, Formation energies

PROGRAM SUMMARY

Manuscript Title: Equilibrium point defect and charge carrier concentrations in a material determined through calculation of the self-consistent Fermi energy

Authors: J. Buckeridge

Program Title: SC-FERMI

Journal Reference:

Catalogue identifier:

Licensing provisions: MIT licence

Programming language: FORTRAN 90

Computer: Any computer with a FORTRAN 90 compiler

Operating system: Any OS with a FORTRAN 90 compiler

RAM: 2 megabytes

Number of processors used: one

Keywords: Crystal defects, Defect concentrations, Self-consistent Fermi level, Formation energies

Classification: 16.1 Structure and properties, 23 Statistical Physics and Thermodynamics

Nature of problem:

To determine the self-consistent Fermi energy and equilibrium defect and carrier concentrations given a set of point defect formation energies in a crystalline system, assuming the constraint of charge neutrality.

Solution method:

The concentrations of each defect in each charge state are calculated, as are the free carrier concentrations. These concentrations are functions of the Fermi energy. The code, using an iterative search algorithm, determines the Fermi energy that satisfies the charge neutrality constraint (the self-consistent Fermi energy). The defect and carrier concentrations at that Fermi energy are then reported, as well as the Fermi energy itself.

Restrictions:

Thermodynamic equilibrium is assumed. The defect formation enthalpies and electronic density of states of the pristine system must be known.

Additional comments:

The concentrations of defects can be fixed to a particular value, thus modelling 'frozen-in' defects formed by e.g. kinetic processes. This procedure is facilitated by the related program, FROZEN-SC-FERMI, which is identical to SC-FERMI apart

from the additional defect concentration fixing routine.

Running time: Less than one second.

1. Introduction

Defects form in almost all conditions and strongly affect the structural, elastic, optical and electronic properties of crystalline and amorphous systems. [1, 2, 3] The physical properties arising from the presence of defects are notoriously difficult to attribute to individual species experimentally; unless the defects are present in very high concentrations, most often one has to infer their structure and energy states by secondary means (see e.g. Refs. [4, 5, 6, 7]). Computational approaches, however, allow one to construct a system into which a specific defect can be inserted and its properties calculated, and are therefore crucial to understanding the rôle defects play. [8, 9, 10] The process of point defect formation, in which an atom or group of atoms is either removed from a lattice site, replaced by another atom or group, or embedded in an interstitial site in equilibrium with reservoirs of the relevant atoms in elementary form (or in other competing compounds) and charge balanced by the formation of free electron or hole carriers, is described by a defect reaction. [11, 12, 13, 14, 15, 16] The enthalpy of this reaction, commonly referred to as the ‘defect formation energy’ can now be routinely calculated using a number of accurate approaches, with the workhorse being density functional theory (DFT). [17]

The defect formation energy allows one to determine the most favourable defects, and therefore most likely to form, in a given system under particular environmental and thermodynamic conditions. [17] Furthermore, by analysing how the formation energies vary with the electron chemical potential or Fermi energy (E_F), the nature of the most dominant charge carrier can be inferred, [18, 16, 19] as well as likely compensation schemes when the system is doped in an attempt to influence carrier concentrations. [20, 21, 22, 23, 24, 25] It is also possible, however, to determine equilibrium defect and carrier concentrations at temperature T , given the formation energies of the defects and density of states (DOS) of the pristine system, [26] by imposing the constraint of overall charge neutrality in the system. [27, 9] As the concentrations are functions of E_F , one can determine the value of E_F where the concentration of negatively charged defects and electrons equals that of holes and positively charged defects, i.e. the equilibrium concentra-

tions that satisfy charge neutrality. This value of E_F is referred to as the ‘self-consistent Fermi energy’.

Calculations of defect concentrations as a function of T provide useful data for experimentalists, particularly since the formation energies depend on environmental conditions through the chemical potentials of the atoms exchanged with chemical reservoirs; variations in chemical potential and the resulting effect on the concentrations allows one to quantify the influence of the environment on the defect chemistry. [10] A drawback is that the analysis depends on the assumption of thermodynamic equilibrium during the defect formation process, a condition which is often not met in real experiments. It is nevertheless possible to assume that concentrations of certain species exist within the system, which remain fixed or vary with T in a pre-defined way, and recalculate the self-consistent E_F subject to the charge neutrality constraint. In this way one can model ‘frozen in’ defects that may form away from equilibrium and remain in the system due to kinetic barriers. [9] For example, one can compute the defect concentrations that form at elevated temperatures, and assume some or all persist via kinetic means when the system is cooled down rapidly, and then determine the equilibrium concentrations of other defects and charge carriers given the presence of these ‘frozen in’ defects.

In this paper a simple but effective FORTRAN 90 code to calculate the self-consistent Fermi energy and equilibrium defect concentrations, given a temperature, set of defect formation energies and total density of states of the pristine system is presented, named SC-FERMI. [28, 29, 30, 31] A further routine, which allows the user to fix concentrations of defects or add more defects of a particular charge, is also described (FROZEN-SC-FERMI). This problem has, of course, been solved before (for some examples see Refs. [32, 9, 33, 34, 35, 36, 37, 38]), but the underlying theory has only been discussed briefly in other works, usually in relation to a particular (often simplified) problem, and, to the best of my knowledge a straightforward computer code to treat this problem has not been made available previously to the wider community. The method of solution presented here is based on a linear search algorithm, with appropriate checks for both very large and very small numbers. The implementation of the procedure and the power of the analysis available from the results is demonstrated through a thorough examination of a set of defect formation energies derived from calculations of the defect properties of the wide gap semiconductor GaN.

The background theory on equilibrium concentrations, originating in sta-

tistical physics, is presented in Section 2, the algorithm is described in Section 3, examples of usage are demonstrated in Section 4, and a summary is given in Section 5.

2. Theory

Let us assume that there exists an extended, solid-state system comprised of one or more atomic species in which defects may form, in thermodynamic equilibrium with its environment. The electronic structure of the material is described by the density of states per unit volume $\rho(E)$, a function of energy E , and both electron (n_0) and hole (p_0) carrier concentrations can be present through thermal activation across the band gap E_g . The concentrations are given by: [39]

$$n_0 = \int_{E_g}^{\infty} f_e(E)\rho(E)dE; \quad (1)$$

$$p_0 = \int_{-\infty}^0 f_h(E)\rho(E)dE; \quad (2)$$

where $f_e(E) = [\exp((E_F - E)/kT) + 1]^{-1}$ is the Fermi-Dirac distribution function and $f_h(E) = 1 - f_e(E)$ (k is Boltzmann's constant). Note that the valence band maximum (VBM) is set as the zero of the energy scale, and that it is possible that $E_g = 0$, i.e. gapless materials can be included in this analysis [30, 40]. For gapless materials, such as metals and semimetals, one must take considerable care in the modelling of charged defects, which will be screened by the surrounding charge density so that the combined defect plus screening charge is neutral. How the screening is achieved for particular defects in such systems needs to be analysed carefully; in metals defects can be ionised, [41] which may affect the Fermi energy and thus their concentrations could be analysed using SC-FERMI, but modelling such defects accurately can be challenging. [30] Of course, the details of how the defects are simulated only affects the input to SC-FERMI.

Above 0 K, in thermodynamic equilibrium the system contains a set of defects $\{X\}$; each member X of this set can exist in a range of charge states q , [42] each with a particular Gibb's free energy of formation $\Delta G(X^q) = \Delta H(X^q) - T\Delta S(X^q)$, where ΔH refers to enthalpy and ΔS to entropy. For simplicity, the defects are assumed to form in the dilute limit and are hence non-interacting. In the majority of realistic situations involving point defect

formation in solids, the entropic contribution to $\Delta G(X^q)$ is at least one order of magnitude lower than the enthalpic contribution. [8, 9, 43] Moreover, although defect formation can contribute a defect volume and pressure in a crystal, [43] the effect is small and typically ignored. It is therefore quite common to ignore the entropic term and equate $\Delta G(X^q)$ with the formation energy (rather than enthalpy) of X^q , $E_f(X^q)$, although there have been suggestions that some results from this approach can be misleading. [44, 45, 46] This approximation is applied in the following derivation, in common with many computational analyses of defects in solids. [17] It should be noted that one can still include the full enthalpic and entropic contributions to defect formation in the equations below; [47] it is not done so here for ease of notation.

The formation energy, computed from the relevant defect reaction, is given by

$$E_f(X^q) = \Delta E(X^q) + \sum_i n_i \mu_i + qE_F, \quad (3)$$

where $\Delta E(X^q)$ is the energy difference between the system with the defect and without and n_i is the number of atoms i with chemical potential μ_i removed ($n_i > 0$) or added ($n_i < 0$) when forming X .

The thermal transition level from charge state q to the neutral state $\epsilon(q/0)$ is defined as that value of E_F for which $E_f(X^q) = E_f(X^0)$:

$$q\epsilon(q/0) = \Delta E(X^0) - \Delta E(X^q). \quad (4)$$

Defect X has a total concentration C_{X^T} , and each of its charge states has a concentration C_{X^q} . Each defect X can form on crystal sites of density N_X , and each of its charge states has a degeneracy g_{X^q} . In the neutral state, the concentration C_{X^0} is given by:

$$C_{X^0} = N_X g_{X^0} \exp(-E_f(X^0)/kT). \quad (5)$$

This result is a well-known consequence of statistical thermodynamics [26, 47] that follows from consideration of the configurational entropy associated with the formation of X^0 and the minimisation of the resulting change of free energy as a function of the number of X^0 formed. [48] Taking into account the law of mass action for a process whereby X^0 transforms to X^q , one can derive an expression for C_{X^q} equivalent to Eq. 17 below. [12, 13] An alternative derivation, however, resulting from counting the average number of electrons (and holes) occupying defect states, is presented here. This

method demonstrates the connection between defect formation and electron counting in a solid and, indeed, is fully quantum statistical mechanical in nature. It has not been presented in this generalised form previously to my knowledge, and therefore should provide an instructive example of the statistics of defect formation.

The problem we would like to solve is to determine the E_F which satisfies the constraint of charge neutrality in our system, given the presence of excess charge carriers and defects with non-zero charge states. To quantify the charge balance in the system, the electron and/or hole concentration that arises from the presence of charged defects therefore needs to be determined. A defect X will have several states associated with it that can be occupied by different numbers of electrons, or, alternatively, holes. The thermally averaged concentration of electrons $\langle n_X \rangle$ associated with X is calculated by summing the contributions of the different charge states q , dividing by the partition function and multiplying by the total concentration of X , C_{X^T} : [49]

$$\langle n_X \rangle = \frac{\sum_q N(X^q) \exp(-(E_{X^q} - E_F N(X^q))/kT)}{\sum_q \exp(-(E_{X^q} - E_F N(X^q))/kT)} C_{X^T}. \quad (6)$$

E_{X^q} is the energy of an electron in state q of X , $E_{X^q} = \Delta E(X^q) - \Delta E(X^0)$, while $N(X^q)$ is the number of electrons in state q of X . Each state contributes g_{X^q} terms to the summations in Eq. 6. One hole can be counted as -1 electron; $\langle n_X \rangle$ will therefore be positive for electrons and negative for holes, we have $N(X^q) = -q$ and the condition of charge neutrality can be expressed as

$$n_0 + \sum_X \langle n_X \rangle = p_0, \quad (7)$$

where we sum over all defects that form in the system. Taking into account the expression for the transition level $q\epsilon(q/0)$ in Eq. 4, we can write:

$$\langle n_X \rangle = \frac{\sum_q -q \frac{g_{X^q}}{g_{X^0}} \exp(q(\epsilon(q/0) - E_F)/kT)}{1 + \sum_{q \neq 0} \frac{g_{X^q}}{g_{X^0}} \exp(q(\epsilon(q/0) - E_F)/kT)} C_{X^T}. \quad (8)$$

The form of the exponent here indicates clearly the intuitive reasoning that, if the average occupation of a charged state q is high and hence has a large exponential term in Eq. 8, when $q > 0$ ($q < 0$) E_F lies above (below) the transition level. This outcome corresponds to the well-known picture of occupation of donor or acceptor levels in semiconductors, depending on the relative position of E_F to those levels. [39]

Let us now define the factor:

$$B_{X^q} \equiv \frac{g_{X^q}}{g_{X^0}} \exp(q(\epsilon(q/0) - E_F)/kT), \quad (9)$$

so that

$$\langle n_X \rangle = \frac{\sum_q -q B_{X^q}}{1 + \sum_{q \neq 0} B_{X^q}} C_{X^T} \equiv \sum_q -q C_{X^q}. \quad (10)$$

This result follows simply from the fact that each X^q contributes $-q$ electrons (counting holes as negative amounts of electrons, see above). We therefore have

$$C_{X^q} = \frac{B_{X^q}}{1 + \sum_{q \neq 0} B_{X^q}} C_{X^T}. \quad (11)$$

By definition,

$$C_{X^T} = \sum_{q \neq 0} C_{X^q} + C_{X^0}, \quad (12)$$

$$= \frac{\sum_{q \neq 0} B_{X^q}}{1 + \sum_{q \neq 0} B_{X^q}} C_{X^T} + C_{X^0}, \quad (13)$$

$$= C_{X^0} \left(1 + \sum_{q \neq 0} B_{X^q}\right). \quad (14)$$

It therefore follows that

$$C_{X^q} = C_{X^0} B_{X^q} \quad (15)$$

$$= C_{X^0} \frac{g_{X^q}}{g_{X^0}} \exp(q(\epsilon(q/0) - E_F)/kT). \quad (16)$$

Finally, noting from Eqs. 3 and 4 that $q(\epsilon(q/0) - E_F) = E_f(X^0) - E_f(X^q)$, we have

$$C_{X^q} = N_X g_{X^q} \exp(-E_f(X^q)/kT). \quad (17)$$

With reference to Eq. 5 and the law of mass action this result is not that surprising; indeed, the theory from which Eq. 5 is derived does not specify the charge of the defect species, but is applied to the neutral state to avoid problems of electron counting associated with charge neutrality. The above derivation demonstrates the relationship between the electrons introduced by a defect (that occupy its different states) and the total concentration of that

defect, thus taking into account directly the effect the defect concentration has on the charge balance in the system. We can rewrite Eq. 7 as

$$n_0 - \sum_X \sum_q q C_{X^q} = p_0. \quad (18)$$

As each term in Eq. 18 depends on E_F , a search algorithm can be used to calculate that value of E_F satisfying Eq. 18, given T .

3. Algorithm

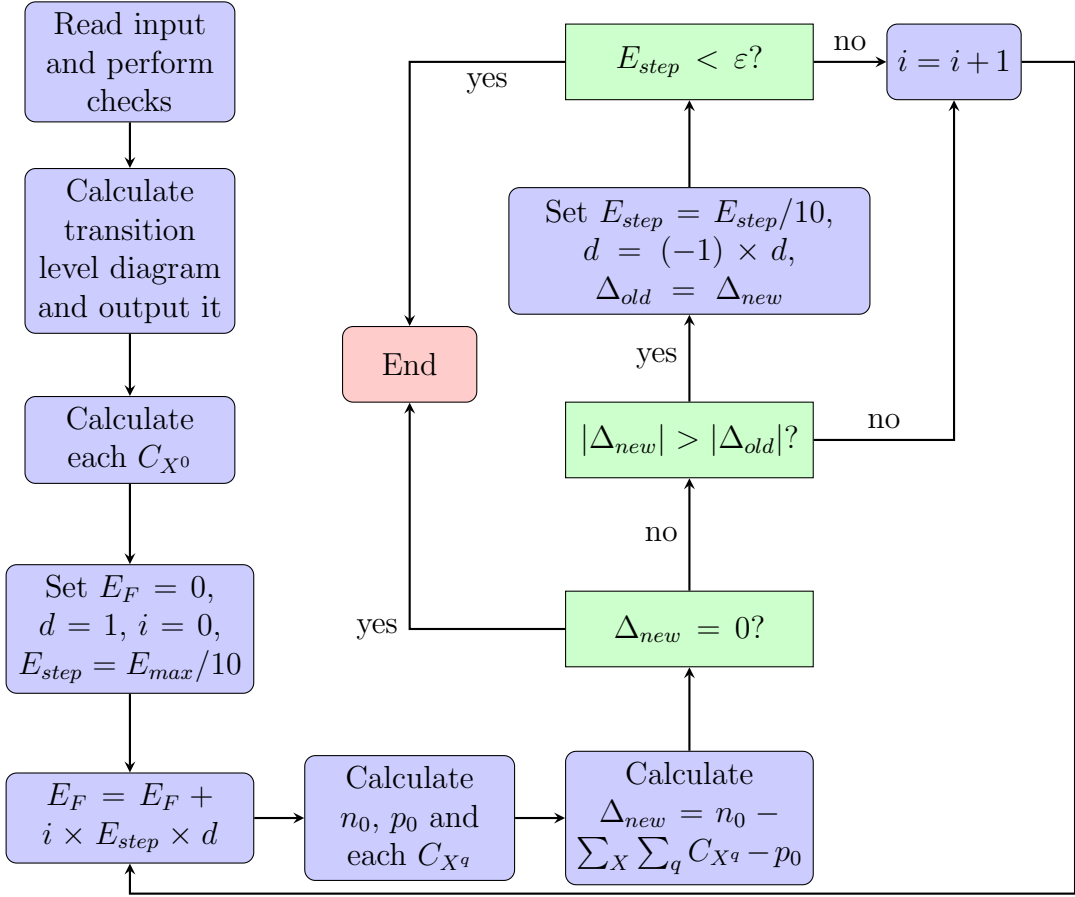


Figure 1: Flow chart of main algorithm.

A simple representation of the algorithm we have implemented to calculate the concentrations and self-consistent E_F is shown in Fig. 1. A detailed description now follows.

The initial step is to read in the input, which consists of $\rho(E)$, defined as the number of states per energy interval per unit cell, with the energy scale set to zero at the initial E_F (which will be the VBM for an insulator), E_g (which may be zero), the temperature T , the lattice vectors \mathbf{a} , \mathbf{b} and \mathbf{c} of the unit cell for which $\rho(E)$ was calculated and the number of electrons within that unit cell. For each defect of the set $\{X\}$, the name, the number of sites on which they can form within the unit cell, and the degeneracies g_{X^q} and formation energies $E_f(X^q)$ (at $E_F = 0$) of each of its charge states q are read in. By providing the $E_f(X^q)$, the user sets the environmental conditions in which defect formation is considered via the chemical potentials in Eq. 3. [50] Simple checks are performed on the input, such as making sure $\rho(E)$, T and E_g are non-negative and that the energy range within which $\rho(E)$ has been defined is broad enough to include conduction band states. From the lattice vectors, the cell volume is computed as $V = \mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$. The density of states is then integrated using Simpson's rule up to the initial Fermi energy (which, as stated above, for an insulator is set equal to the VBM and coincides with the zero of the energy scale) to check that the number of electrons and number of occupied states are compatible; if not, or if there are small differences, $\rho(E)$ is renormalised.

Once the input has been read in and checked, the defect formation energies are used to calculate the thermal transitions and output a transition level diagram, consisting of the formation energy of each defect in its lowest energy charge state, as a function of energy within the range for which $\rho(E)$ has been defined. These datasets are calculated as follows: for each defect, the lowest energy charge state is found at E_{min} , the lower limit of the energy range, using the values provided by the user of $E_f(X^q)$ at $E_F = 0$ and Eq. 3. Calling this charge state q , the energies E_{min} and $E_f(X^q)$ at $E_F = E_{min}$ are sent to output in two column format. The next step is to determine the q' for which $\epsilon(q/q')$ (see Eq. 4) lies closest to E_{min} , where $q' > q$. $\epsilon(q/q')$ and the formation energy $E_f(X^{q'})$ at $E_F = \epsilon(q/q')$ are then sent to output. This procedure is repeated, with q replaced by q' , until all charge states have been cycled through, resulting in a set of points defining the line of lowest formation energy for X as a function of E_F from E_{min} to E_{max} , the maximum energy in the range for which $\rho(E)$ is defined. The two column format can then be graphed to produce the transition level diagram.

After the transition level diagram has been determined, the neutral concentrations C_{X0} are calculated using Eq. 5. Next, an initial value for E_F is defined, but also for E_{step} , i and d , which are parameters used to step E_F through the energy range in some direction (given by d) in an energy step size E_{step} . The search algorithm begins by stepping E_F by an amount $i \times E_{step} \times d$ (for the initial step $i = 0$). n_0 , p_0 and the various C_{Xq} are then determined using Eqs. 1, 2 and 17, respectively, from which the difference from charge neutrality, $\Delta_{new} = n_0 - \sum_X \sum_q C_{Xq} - p_0$ is calculated. First we check if $\Delta_{new} = 0$ (zero here is approximated by $10^{-12} \times (n_0 - \sum_X \sum_q C_{Xq} + p_0)$); if so then the algorithm finishes, if not we check if $\Delta_{new} > \Delta_{old}$ (for the first iteration, Δ_{old} is not defined, so this check is assumed to be false), and, if false, we increment i and return to the stage where E_F is stepped along the energy range. If true, however, it indicates that we are progressing E_F in the wrong direction in order to find where the charge neutrality condition is satisfied. We therefore multiply d by -1 , which changes the search direction, and reduce E_{step} by an order of magnitude, allowing E_F to be stepped along the energy range by a finer one-dimensional grid point size. If, after these changes, E_{step} is below a certain value ε (we set $\varepsilon = 10^{-12}$ eV), the algorithm finishes. If not, i is iterated and the procedure returns to the stage where E_F is stepped along the energy range.

The algorithm therefore proceeds by a one-dimensional line search along the energy range to determine where E_F results in an n_0 , p_0 and set of C_{Xq} that most closely satisfy the charge neutrality condition given by Eq. 18. Once the algorithm finishes, n_0 , p_0 and all the $C_{X\tau}$ are calculated (see Eqs. 1, 2 and 14) and outputted, along with the final (self-consistent) E_F .

3.1. Large and small exponential terms

The procedure is quite stable, but care must be taken to avoid floating point number underflow and overflow. Underflow may become problematic at very low T , when all exponential terms are so small that they become stored as zero within the computer memory. When this problem arises, the search algorithm will fail as there will be a range of energy for which $\Delta_{new} = \Delta_{old} = 0$, with the true solution lying somewhere within this range. Our approach to overcome this problem is to take the midpoint of this range as the self-consistent E_F , which is a poor approximation but justifiable as equilibrium defect concentrations at very low T (below 10 K) are very rarely of interest. Overflow is less problematic, as large values of exponential terms occur far from where the charge neutrality condition is satisfied, so that the

algorithm by design avoids these energy ranges. By defining a maximum number as the limit to which the exponential terms can approach, set by the order of magnitude of the largest floating point number that can be stored in memory, overflow problems are avoided.

3.2. Fixed defect concentrations

The theory presented so far corresponds to thermodynamic equilibrium concentrations of defects and charge carriers. In real systems, however, many processes by which materials are grown or modified may occur far from equilibrium. In such cases, it is possible that defects form in non-equilibrium conditions and become ‘frozen in’ by kinetic barriers. [51] Their presence can then affect strongly the defect chemistry by altering the charge neutrality condition given in Eq. 18.

Modelling such ‘frozen in’ defects is possible using the procedure described above. If our system contains a set of defects $\{X\}$, for which we know the $E_f(X^q)$, then we can consider the following possibilities: one (or more) of the defects within $\{X\}$ has a fixed total concentration C_{XT} ; one (or more) of the defects has a fixed concentration C_{Xq} for one (or more) of its charge states q ; or a combination of both these cases. One must be cautious, however, in the application of the latter two possibilities, as the relative concentrations of different charge states of a defect depend strongly on the Fermi energy. Fixing the concentration of a certain charge state, while allowing the Fermi energy to change in order to find a self-consistent solution, corresponds to an unphysical situation. This functionality is included strictly for testing purposes. It may be the case that the user intends to analyse an assumption made in an experimental study regarding the dominant charge states of certain defects and the consequences for the electronic properties of the system in question. By fixing the concentrations of certain charges states, the outcomes of such assumptions can be tested for a given set of computed formation energies. Furthermore, one can consider the case where another defect Y , not belonging to $\{X\}$, has a fixed concentration C_{Yq} for a charge state q (here, as the $E_f(Y^q)$ are unknown, we must assume the charge state is also fixed). In this way, one can analyse how the balance in the concentrations of the defects X is affected by the presence of the charged defects Y^q , which would allow, for example, the compensation of ionised donors or acceptors in semiconductors to be calculated.

The procedure to treat such cases for ‘frozen in’ defects is as follows. If we have a defect X whose total concentration C_{XT} is fixed, as well as that

of some of its charge states q' ($C_{Xq'}$), then we first calculate the C_{Xq} , where $q \neq q'$, in analogy with Eq. 11:

$$C_{Xq} = \frac{B_{Xq}}{1 + \sum_{q' \neq 0} B_{Xq'}} [C_{X^T} - \sum_{q'} C_{Xq'}]. \quad (19)$$

The charge neutrality condition in Eq. 18 is then calculated, including the fixed concentrations $C_{Xq'}$ in the summation, and the procedure is continued as normal. Notice that the formation energies for the charge states q' need not be known; therefore one can add fixed concentrations for charge states not initially considered in the set $\{X\}$. In the same manner, one can introduce the fixed concentrations C_{Yq} of defects not in $\{X\}$. If $E_f(X^0)$ is known, then the concentration for the neutral case is determined as

$$C_{X^0} = C_{X^T} - \sum_{q'} C_{Xq'} - \sum_q C_{Xq}. \quad (20)$$

If $E_f(X^0)$ is not known, then the denominator in Eq. 19 is replaced with $\sum_{q' \neq 0} B_{Xq'}$.

Problems associated with very small exponentials (see Section 3.1) may occur during the procedure to compute the C_{Xq} when C_{X^T} is fixed. To mitigate such problems, the minimum $E_f(X_{\min}^q)$ at the current E_F is determined, and the energies that enter the computation of the B_q in Eq. 19 are shifted relative to that minimum energy. Doing so ensures that at least one term of the summation in the denominator of the right hand side of Eq. 19 has a factor of one, rather than $\exp(-E_f(X_{\min}^q)/kT)$, which avoids difficulties arising when $E_f(X_{\min}^q)/kT$ is large enough that the exponential is stored as zero in memory.

This methodology is a useful analytical tool to employ after the equilibrium concentrations for the set $\{X\}$ have been determined. One can then easily calculate the effect of ‘freezing’ particular defects and/or charge states, for example to model the formation of defects at an elevated temperature that persist after rapid quenching, test assumptions made regarding the persistence of particular defect charge states as well as introduce new defects with an assumed dominant charge state that change the defect chemistry of the system of interest.

4. Examples

To illustrate how the program works, a set of formation energies of intrinsic defects in the wide gap semiconductor GaN has been selected. The

purpose of this section is not to analyse the technological aspects of the material, discuss fundamental physical processes in GaN, explain experimental observations nor consider the implications of defect formation on potential device applications; rather, the usage of the program and the power of the analysis it provides is demonstrated. Some introduction to the topic, however, is required. GaN is a widely studied and technologically important material and a key component in blue light emitting diodes (LEDs). [52, 53] It has a fundamental band gap at low temperature of 3.5 eV, [54] and tends to be intrinsically *n*-type and easily *n*-dopable. [55] Achieving *p*-type samples, however, which are crucial for LEDs amongst other applications, is extremely challenging. [55] To date, doping with Mg has provided the only route to realising *p*-type GaN, but large concentrations of Mg are necessary for observable hole concentrations and the mechanism by which Mg incorporation promotes free hole formation is still a matter of debate. [56, 57] Key to understanding the conductivity properties and dopability of GaN is the intrinsic defect chemistry of the system. More details on defect formation in GaN can be found in, for example, Refs. [58, 59, 60].

In Table 1 a set of formation energies for intrinsic defects in GaN including a Ga vacancy (V_{Ga}), N vacancy (V_{N}), Ga interstitial (Ga_i), N antisite (N_{Ga}) and N interstitial (N_i), is presented. Each defect formation energy, calculated with $E_F = 0$ eV (the zero of the energy scale equals the VBM of the system), is given with the corresponding q and g_{Xq} . A comprehensive description of how these energies have been computed is given in Ref. [61]; here it suffices to say that the hybrid quantum mechanical/molecular mechanical (QM/MM) embedded cluster technique was employed, [62, 63, 64, 61, 65] using DFT with the hybrid functional B97-2. [66] The results are given for “anion-poor” conditions, i.e. those whereby an excess of Ga is present in the environment, so that in defect formation processes N atoms are exchanged with a reservoir of $\text{GaN}_{(s)}$, rather than $\text{N}_{2(g)}$.

It is worthwhile to note that, to demonstrate the usage of **SC-FERMI** and **FROZEN-SC-FERMI**, how the defect formation energies were computed is largely irrelevant. Indeed, the approach used to determine the values in Table 1 is non-standard for defects in wide-gap semiconductors and may in fact be somewhat controversial [60]. It is therefore paramount to stress that the current work does not aim to explain the defect physics of GaN, but to demonstrate the power of analysis available using **SC-FERMI** and **FROZEN-SC-FERMI**. The set of defect formation energies from Ref. [61] has been chosen, not necessarily for accuracy, but for the availability of all defect charge states over a

Fermi energy range greater than the band gap, which is required to demonstrate the full analytical power of the codes described here. It is shown below, however, that the current set in fact is quite similar to others determined using standard supercell techniques [58, 59]. (In the Supplementary Material, it is indicated that the conclusions reached using the current dataset would also be arrived at using the results available in Refs. [58, 59]). With this in mind, the key point is that a set of formation energies has been determined, and our task is now to compute the equilibrium carrier and defect concentrations at particular values of T , and analyse how such concentrations alter when particular defect concentrations are kept fixed.

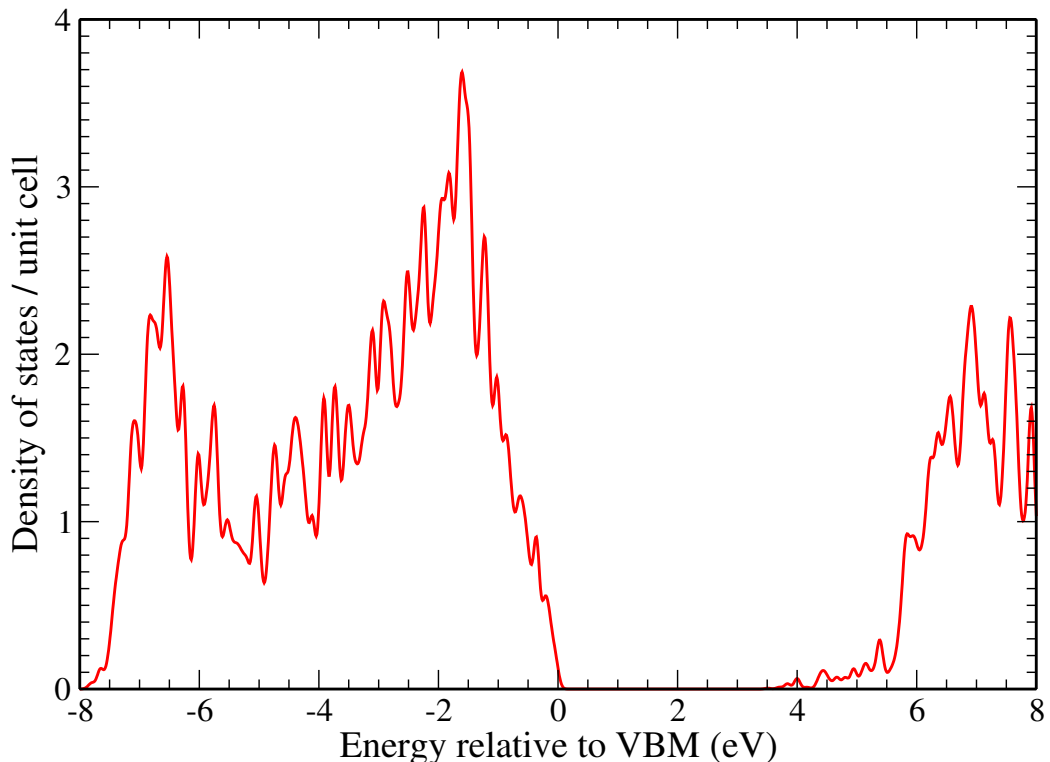


Figure 2: The calculated total density of states in pristine GaN (see Ref. [61]) as a function of energy relative to the valence band maximum (VBM). The density of states is given in units of number of states per energy interval per unit cell.

The values given in Table 1, along with $\rho(E)$ (shown in Fig. 2), E_g , T and the unit cell parameters of wurtzite GaN ($a = 3.181 \text{ \AA}$, $c = 5.185 \text{ \AA}$) [61]

Table 1: The formation energies $E_f(X^q)$ of the five lowest energy defects X in GaN in their various charge states q (each with degeneracy g_{X^q}), determined at $E_F = 0$ eV in anion-poor conditions. The values are taken from Ref. [61] and were determined using DFT with the B97-2 hybrid functional.

X	q	g_{X^q}	$E_f(X^q); E_F = 0$ (eV)
V_{Ga}	-3	1	14.75
	-2	6	11.84
	-1	6	9.23
	0	2	7.01
	1	2	5.61
V_{N}	3	1	-3.49
	2	2	-2.17
	1	1	0.91
	0	6	2.94
Ga_i	-1	6	6.70
	3	1	-0.86
	2	2	2.13
	1	1	4.44
N_{Ga}	0	2	9.03
	-3	4	21.29
	-2	2	15.64
	-1	4	12.43
	0	1	9.24
	1	6	7.24
	2	1	5.07
N_i	3	6	3.93
	-2	4	14.46
	-1	2	9.37
	0	8	5.89
	1	4	3.09
	2	8	1.98

comprise the input for SC-FERMI. In the following analysis, for simplicity the formation energies, unit cell parameters, $\rho(E)$ and E_g are assumed not to change with T . Of course, if known the trends of these properties with T can easily be incorporated by editing the input to the code.

4.1. Transition level diagram

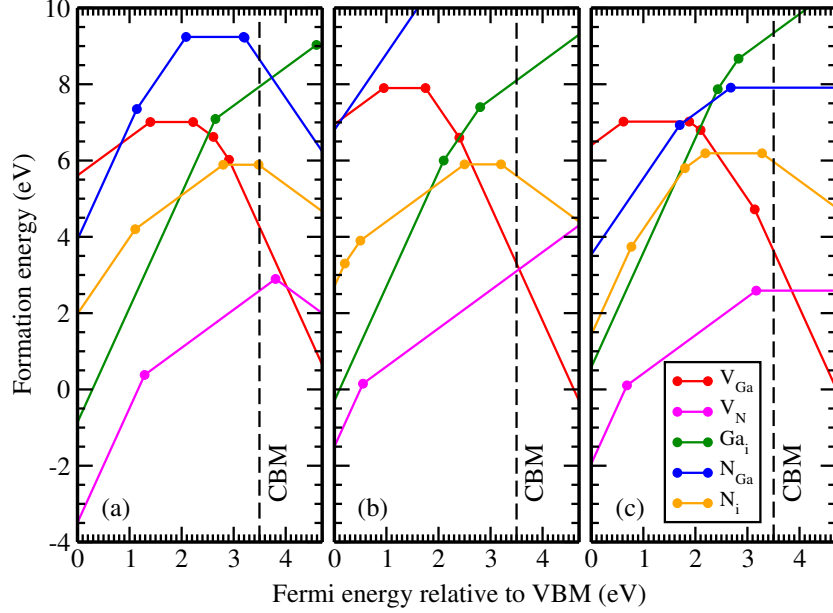


Figure 3: GaN intrinsic defect formation energies as a function of Fermi energy relative to the valence band maximum (VBM). The conduction band minimum (CBM) is indicated by the vertical dashed line. The slopes of the lines indicate the defect charge states; transitions from one charge state to another are represented by circles. The formation energies in (a) correspond to those in Table 1, while those in (b) are taken from Ref. [58] and in (c) from Ref. [59].

On execution, the program provides a file to plot the defect formation energies as a function of Fermi energy relative to the VBM. The corresponding plot is shown in Fig. 3(a). For comparison, the computed defect formation energies determined using plane-wave DFT are shown in Fig. 3(b), taken from Ref. [58] and in Fig. 3(c), taken from Ref. [59]. As can be seen, the main discrepancies between the results in Fig. 3(a) and the other studies are the energies of formation of the N_{Ga} and the $(3+ / +)$ transition of V_{N} . Neither of these differences, however, affect the results presented below in a significant way, as can be seen from the corresponding results presented in the Supplementary Material derived using the plane-wave DFT calculations from Refs. [58, 59]. The data presented in Fig. 3(a) has been chosen for further analysis as the full set of formation energies for each defect in all charge

states has been made available over a Fermi energy range greater than the band gap. For the data shown in Fig. 3(b) and (c), only particular charge states are available and the formation energies for E_F above the CBM are speculative, based on extrapolations from the values within the band gap.

Some conclusions can be drawn immediately from Fig. 3(a). According to the computed formation energies, the dominant intrinsic defect will be V_N as it has the lowest formation energy for values of E_F across the band gap and up to 0.5 eV above the CBM, after which V_{Ga} becomes the lowest energy defect. As the dominant charge states of V_N are $3+$ up to $E_F = 1.30$ eV and $+$ for $1.30 < E_F < 3.81$ eV, the defect may act as a compensating centre for acceptor-like dopants, depending on its concentration. Also of note is the fact that, at $E_F < 1.19$ eV, $E_f(V_N)$ is negative. From Eq. 17, it is evident that a negative formation energy would result in a defect concentration greater than the concentration of sites on which that defect can form; a logical inconsistency indicating the breakdown of the assumption of non-interacting defects at the dilute limit. The self-consistent E_F therefore cannot be in this range. Although these conclusions can be drawn by simply studying the transition level diagram, SC-FERMI allows one to progress beyond this basic analysis and calculate the concentrations at different temperatures. In the following, all conclusions are based on the computed defect formation energies presented in Table 1. Once again, it is worth stating that the conclusions are not definitive with regard to the defect physics of GaN; instead they are what can be determined *given* a particular set of defect formation energies.

4.2. Equilibrium concentrations

In Fig. 4, the computed self-consistent E_F (upper panel) and equilibrium carrier and defect concentrations (lower panel) in GaN as a function of T are shown. As expected from examining the transition level diagram, the dominant defect is V_N , the concentration of which is compensated (to maintain charge neutrality) by electron carriers. E_F remains in the upper half of the band gap for the entire range of T used, as charge neutrality depends on the balance between $[V_N]$, which decreases with increasing E_F , and n_0 , which increases with E_F . Indeed, for the values of E_F determined over this range of T , the most favourable state is V_N^+ ; consequently, $[V_N] = n_0$ holds, as can be seen in the lower panel of Fig. 4. At $T > 1340$ K, p_0 rises above 10^{11} cm^{-3} , which can be attributed to thermal ionisation across the band gap. p_0 remains approximately five orders of magnitude below n_0 and $[V_N]$, however, and therefore has no visible effect on the charge neutrality condition of

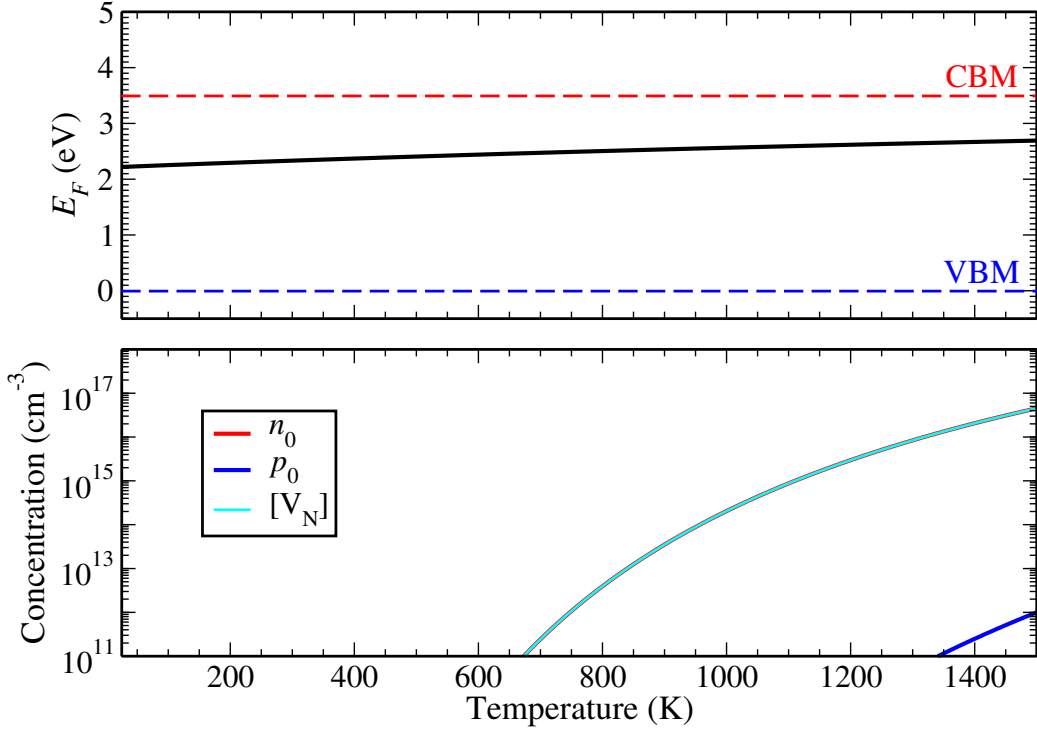


Figure 4: The self-consistent Fermi energy (upper panel) and carrier and defect concentrations (lower panel) as a function of temperature for intrinsic defects in GaN. In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively.

$$[V_N] = n_0.$$

In terms of material properties, these results demonstrate that V_N will form in significant concentrations that give rise to intrinsic n -type carrier concentrations in GaN (assuming that the calculated formation energies and $\rho(E)$ are sufficiently accurate). The n_0 , however, is somewhat lower than that measured in real samples, apart from at higher values of T . One can posit that the V_N form during synthesis, when T is far higher than room temperature (typically $900 - 1050^\circ \text{C}$), [55] and persist at lower T (or become ‘frozen in’) due to the presence of kinetic barriers. [9] This situation can then be analysed using FROZEN-SC-FERMI.

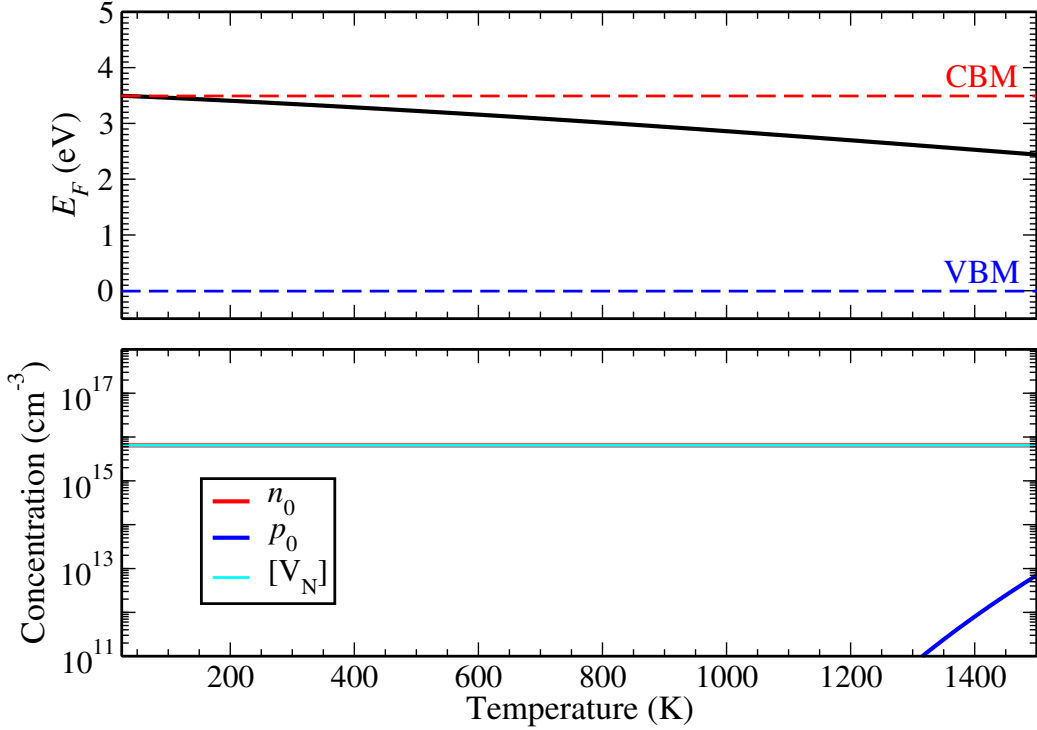


Figure 5: The self-consistent Fermi energy E_F (thick black line in upper panel), electron carrier n_0 (thick red line), hole carrier p_0 (thick blue line) and defect concentrations (all in lower panel) as a function of temperature for intrinsic defects in GaN, when the N vacancy concentration $[V_N]$ (thin cyan line) is kept fixed at $6.447 \times 10^{15} \text{ cm}^{-3}$. In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively.

4.3. Fixed concentrations of N vacancies

If the formation of V_N occurs at $T = 1273 \text{ K}$, and these defects persist at lower T , by fixing $[V_N] = 6.447 \times 10^{15} \text{ cm}^{-3}$ (the value calculated at $T = 1273 \text{ K}$) the concentrations of other defects and carriers that would be present under such circumstances can be determined. Fig. 5 shows the self-consistent E_F (upper panel) and carrier and defect concentrations (lower panel) as a function of T , with $[V_N] = 6.447 \times 10^{15} \text{ cm}^{-3}$ kept fixed. The situation is similar to that depicted in Fig. 4, i.e. charge neutrality is maintained via $[V_N] = n_0$, except here $[V_N]$ does not change with T , and, due to the enforced higher $[V_N]$ at $T < 1273 \text{ K}$, E_F is pushed closer to the CBM. The variation

in E_F with T , governed by maintaining $[V_N] = n_0$, follows the trend for a gas of fermions, which is unsurprising given the dependence of n_0 on the Fermi-Dirac distribution function (see Eq. 1). Again, at higher T , p_0 rises above 10^{11} cm^{-3} due to thermal activation.

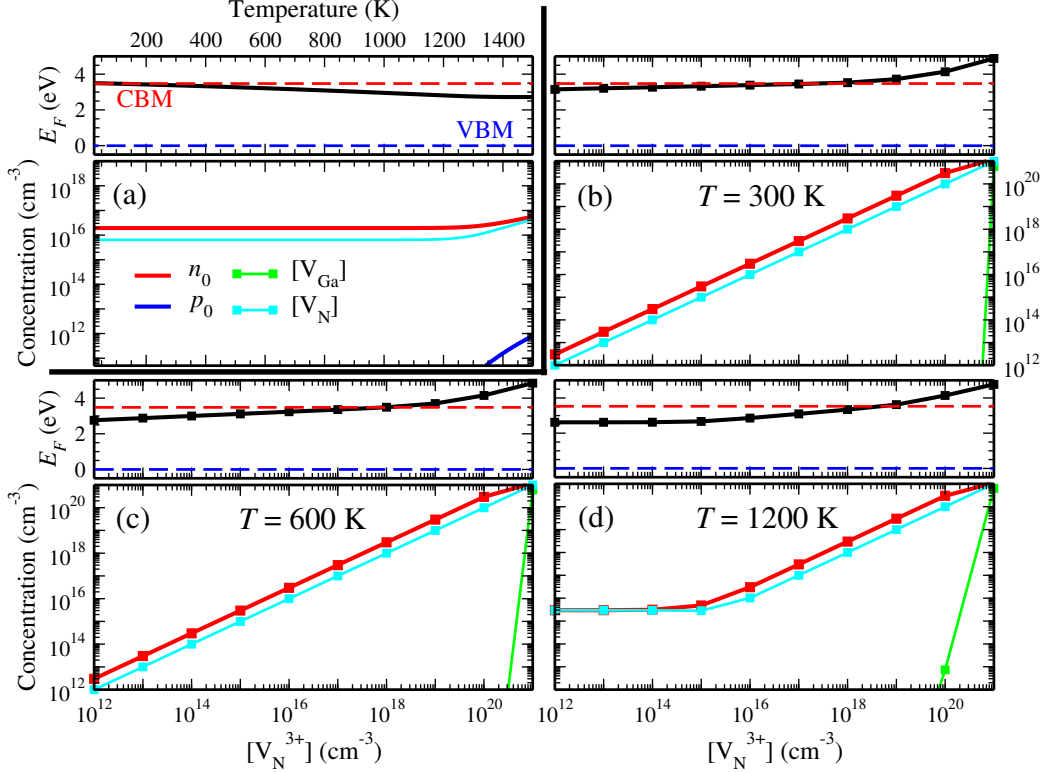


Figure 6: (a) The self-consistent Fermi energy E_F (thick black line in upper panel), electron carrier n_0 (thick red line), hole carrier p_0 (thick blue line) and defect concentrations (all in lower panel) as a function of temperature for intrinsic defects in GaN, when the concentration of the N vacancy in the 3+ charge state $[V_N^{3+}]$ is kept fixed at $6.447 \times 10^{15} \text{ cm}^{-3}$. In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively. (b)–(d) Similar to (a), but here the self-consistent Fermi energy and concentrations are shown as a function of $[V_N^{3+}]$, for temperature $T = 300 \text{ K}$ (b), $T = 600 \text{ K}$ (c) and $T = 1200 \text{ K}$ (d). The Ga vacancy concentration $[V_{Ga}]$ is indicated by the thin green line and the N vacancy concentration $[V_N]$ by the thin cyan line. In (b)–(d) the squares represent the data points; the lines are a guide for the eye.

The results displayed in Fig. 5 were obtained by assuming the total con-

centration of V_N was fixed to the value determined at $T = 1273$ K. If instead, one would like to consider the situation where those V_N formed at higher T have a particular dominant charge state, say $3+$, which persists at lower T , $[V_N^{3+}]$ should then be kept fixed (rather than $[V_N^T]$). The results for such a situation are shown in Fig. 6. In Fig. 6(a), the self-consistent E_F (upper panel) and carrier and defect concentrations (lower panel) as a function of T , with $[V_N^{3+}] = 6.447 \times 10^{15} \text{ cm}^{-3}$ kept fixed, are shown. Here, the charge neutrality condition is satisfied through $3[V_N] = n_0$, up to about $T = 1200$ K; above this value the total $[V_N]$ begins to increase, and gets closer to n_0 . In Fig. 6(b)–(d), the self-consistent E_F (upper panels) and defect and carrier concentrations (lower panels) are shown when T is kept constant and the fixed value of $[V_N^{3+}]$ is allowed to vary. For (b) $T = 300$ K and (c) $T = 600$ K, $3[V_N] = n_0$ holds for values of $[V_N^{3+}]$ up to 10^{20} cm^{-3} ; above this concentration n_0 is much closer to $[V_N]$, and a significant $[V_{Ga}]$ is observed. In (d), where $T = 1200$ K, the results are quite different. For $[V_N^{3+}] < 10^{15} \text{ cm}^{-3}$, n_0 is equal (or very close) to $[V_N]$; above 10^{15} cm^{-3} n_0 becomes closer to $3[V_N]$. Again, at $[V_N^{3+}] = 10^{21} \text{ cm}^{-3}$, n_0 is closer to $[V_N]$ and there is a significant $[V_{Ga}]$ present.

The puzzling aspects of some of these results can easily be resolved by looking at the concentrations of the individual charge states of V_N , which are given in the output from **FROZEN-SC-FERMI**. Fig. 7 shows the total concentration of nitrogen vacancies, $[V_N^T]$, as well as the concentrations of the individual charge states (see Table 1). Fig. 7(a) corresponds to Fig. 6(a), that is, the same constraints apply, but now just the V_N concentrations are plotted; Fig. 7(b) corresponds to Fig. 6(d). As a guide, in the upper panels the positions of the relevant transition levels $\epsilon(3+/+)$ and $\epsilon(+/-)$ are shown relative to the VBM and CBM. In Fig. 6(a), where the results are shown for a fixed $[V_N^{3+}] = 6.447 \times 10^{15} \text{ cm}^{-3}$ over a range of values of T , an increase in $[V_N]$ was observed for $T > 1200$ K. In Fig. 7(a), it is evident that this increase is due to a thermally induced increase in $[V_N^+]$, which is the lowest energy charge state for the computed range of E_F shown in the upper panel. The other charge states of V_N do not play a rôle here. In Fig. 6(d), the results correspond to the situation where $T = 1200$ K is constant, and the fixed $[V_N^{3+}]$ is varied. Fig. 7(b) shows that, for $[V_N^{3+}] < 10^{15} \text{ cm}^{-3}$, the total concentration is dominated by $[V_N^+]$ (as $+$ is the lowest energy charge state for the corresponding values of E_F), which is much greater than the (fixed) $[V_N^{3+}]$. Above 10^{15} cm^{-3} , $[V_N^{3+}]$ becomes greater than $[V_N^+]$ and dominates the total concentration, so that $[V_N^{3+}] = [V_N^T]$. At the same time, E_F is pushed

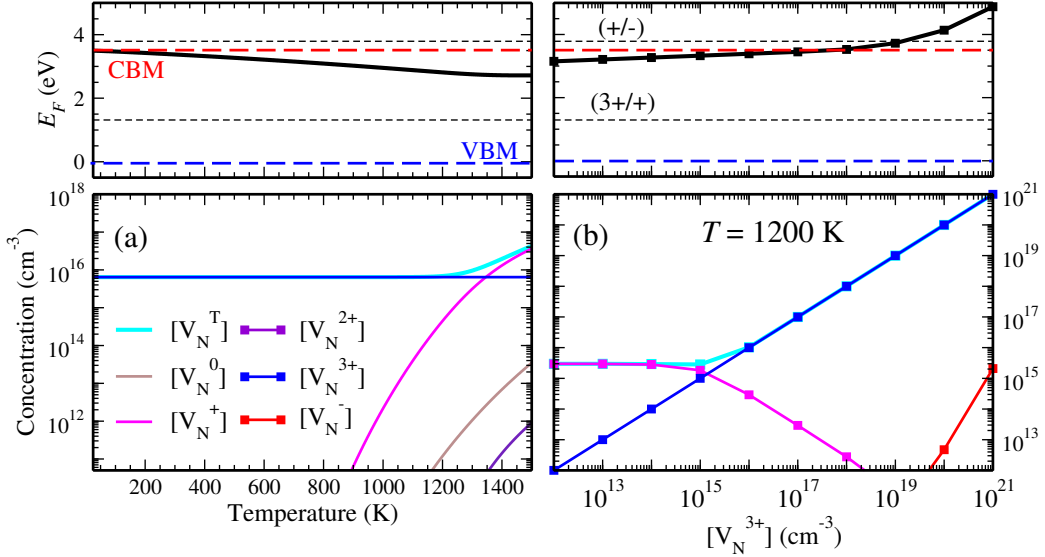


Figure 7: (a) The self-consistent Fermi energy E_F (thick black line in upper panel), total N vacancy concentration $[V_N^T]$ (thick cyan line) and concentrations of the individual charge states: $[V_N^-]$ (thin red line), $[V_N^0]$ (thin brown line), $[V_N^+]$ (thin magenta line), $[V_N^{2+}]$ (thin indigo line) and $[V_N^{3+}]$ (thin blue line), as a function of temperature, while $[V_N^{3+}]$ is kept fixed at $6.447 \times 10^{15} \text{ cm}^{-3}$. In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively, while the thin black dashed lines represent the V_N $(3+/+)$ and $(+/-)$ transition energies. (b) Similar to (a), but here E_F and the V_N concentrations are shown as a function of $[V_N^{3+}]$, for temperature $T = 1200$ K. The squares represent the data points; the lines are a guide for the eye.

to higher values (due to the increase in n_0 to maintain charge neutrality), which increases the formation energy of V_N^+ ; consequently, $[V_N^+]$ decreases. At $[V_N^{3+}] = 10^{19} \text{ cm}^{-3}$, E_F crosses $\epsilon(+/-)$. At higher values of $[V_N^{3+}]$, an increase in $[V_N^-]$ is therefore observed and, at $[V_N^{3+}] = 10^{21} \text{ cm}^{-3}$, in Fig. 6(d) it can be seen that n_0 becomes closer to $[V_N^-]$. This result follows as, at this point E_F rises close to 5 eV, where $E_f(V_{\text{Ga}}^{3-})$ is low enough that significant concentrations of this defect form, which compensate the positive V_N^{3+} , thus pushing down n_0 . From this analysis, a complex balance between charge states and compensation mechanisms is thus revealed for the case where a concentration of V_N is ‘frozen in’ to the system.

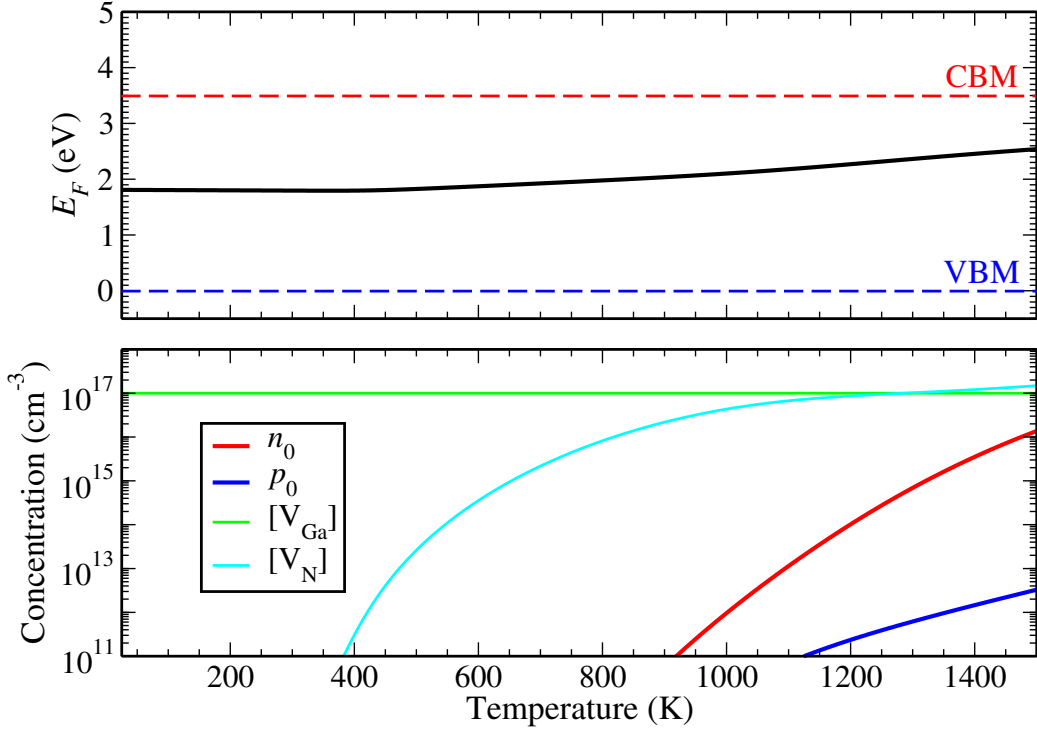


Figure 8: The self-consistent Fermi energy E_F (thick black line in upper panel), electron carrier n_0 (thick red line), hole carrier p_0 (thick blue line) and defect concentrations (all in lower panel) as a function of temperature for intrinsic defects in GaN, when the Ga vacancy concentration $[V_{\text{Ga}}]$ (thin green line) is kept fixed at 10^{17} cm^{-3} . In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively. The N vacancy concentration $[V_{\text{N}}]$ is indicated by the thin cyan line.

4.4. Fixed concentrations of Ga vacancies

Although, according to the results shown in Fig. 3, Ga vacancies have high formation energies for E_F in the band gap or close to the CBM, these defects have been proposed to be present in significant concentrations and to affect the luminescence properties of the material. [5, 67, 68] Their formation may occur under strongly non-equilibrium conditions, or arise due to the presence of complex and/or extended defects. Whatever their origin, using FROZEN-SC-FERMI one can analyse their effect on the concentrations of carriers and other defects. Fig. 8 shows the self-consistent E_F (upper panel) and carrier and defect concentrations (lower panel) as a function of T , with

$[V_{\text{Ga}}] = 10^{17} \text{ cm}^{-3}$ kept fixed. For much of the range of T , the material is intrinsic, with no significant concentration of a majority carrier; the charge neutrality constraint results in E_F remaining at values where V_{Ga}^0 dominates. For $T > 1000 \text{ K}$, however, E_F is increased to values where negative charge states of V_{Ga} become favourable, which are compensated by suitable concentrations of V_{N} . At these elevated temperatures, where E_F moves closer to the CBM, n_0 begins to increase, while p_0 also increases due to thermal excitation across the band gap. Both n_0 and p_0 , however, remain quite low compared with the concentrations of the vacancies.

Fixing $[V_{\text{Ga}}]$ results in, for $T < 1000 \text{ K}$, the neutral state dominating. It may be of interest to analyse whether this defect may be electronically active, and promote hole formation in the system, which may be done by fixing a particular charge state, say V_{Ga}^{3-} . The results obtained with $[V_{\text{Ga}}^{3-}] = 10^{17} \text{ cm}^{-3}$ kept fixed, while T is varied, are displayed in Fig. 9(a), with the self-consistent E_F shown in the upper panel and the carrier and defect concentrations in the lower panel. Figs. 9(b)–(d) shows the results when the value of the fixed $[V_{\text{Ga}}^{3-}]$ is varied and T is kept constant at (b) $T = 100 \text{ K}$, (c) $T = 600 \text{ K}$ and (d) $T = 1200 \text{ K}$. Rather than promoting hole formation, it is found that the V_{Ga} are compensated by the formation of V_{N} . In Fig. 9(a), for $T < 300 \text{ K}$, the charge neutrality condition is determined by $[V_{\text{Ga}}] = [V_{\text{N}}]$, while for $T > 400 \text{ eV}$, the condition is $3[V_{\text{Ga}}] = [V_{\text{N}}]$. This change in the charge neutrality condition is also seen in the differences between the results in Fig. 9(b) and those in Fig. 9(c), where T is below and above 400 eV , respectively; although, at values of $[V_{\text{Ga}}^{3-}] > 10^{18} \text{ cm}^{-3}$, in (c) it is observed that $[V_{\text{N}}]$ moves closer to $[V_{\text{Ga}}]$, rather than remaining three times larger. In Fig. 9(d), where T is much higher at 1200 K , for values of $[V_{\text{Ga}}^{3-}] < 10^{15} \text{ cm}^{-3}$, the equilibrium concentration of V_{N} is higher (see Fig. 4) than that of the fixed $[V_{\text{Ga}}^{3-}]$, and hence the dominant carriers are compensating electrons, with $n_0 = [V_{\text{N}}]$. For $[V_{\text{Ga}}^{3-}] > 10^{15} \text{ cm}^{-3}$, however, the V_{Ga} begin to dominate while being compensated by $[V_{\text{N}}]$ and their presence pushes E_F into the band gap, reducing n_0 . Initially the condition $3[V_{\text{Ga}}] = [V_{\text{N}}]$ holds, but at $[V_{\text{Ga}}^{3-}] = 10^{21} \text{ cm}^{-3}$ the condition becomes closer to $[V_{\text{Ga}}] = [V_{\text{N}}]$. Some thermally excited holes are also evident in this range of $[V_{\text{Ga}}^{3-}]$.

The changes in the charge neutrality condition described above with reference to Fig. 9 can be easily understood by looking at the concentrations of the individual charge states of V_{N} , and the variation in E_F with respect to the $\epsilon(3+/+)$ transition level, as a function of temperature and of $[V_{\text{Ga}}^{3-}]$. The self-consistent E_F (upper panels), and the total V_{N} concentration, as well as

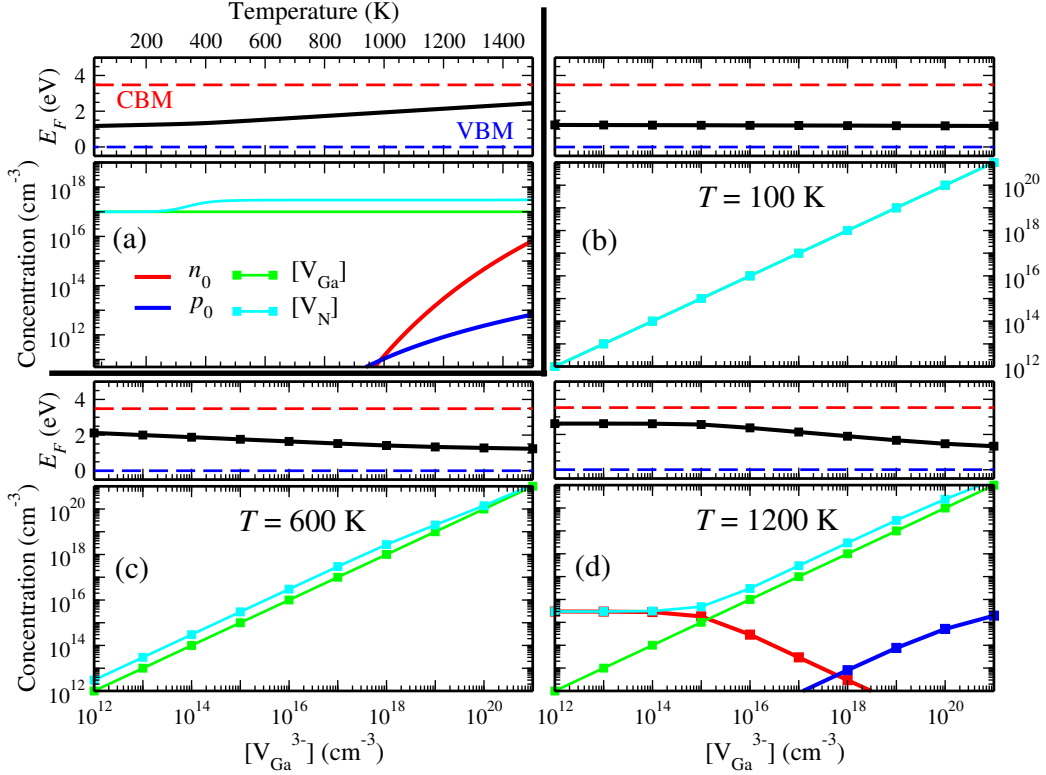


Figure 9: (a) The self-consistent Fermi energy E_F (thick black line in upper panel), electron carrier n_0 (thick red line), hole carrier p_0 (thick blue line) and defect concentrations (all in lower panel) as a function of temperature for intrinsic defects in GaN, when the concentration of the Ga vacancy in the 3- charge state $[V_{Ga}^{3-}]$ is kept fixed at 10^{17} cm^{-3} . In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively. (b)–(d) Similar to (a), but here the self-consistent Fermi energy and concentrations are shown as a function of $[V_{Ga}^{3-}]$, for temperature $T = 100 \text{ K}$ (b), $T = 600 \text{ K}$ (c) and $T = 1200 \text{ K}$ (d). The Ga vacancy concentration $[V_{Ga}]$ is indicated by the thin green line and the N vacancy concentration $[V_N]$ by the thin cyan line. In (b)–(d) the squares represent the data points; the lines are a guide for the eye. Note that in the lower panel in (b), the cyan line lies directly on top of the green line.

that of its charge states and of V_{Ga}^{3-} (lower panels) are shown in Fig. 10, where (a) corresponds to Fig. 9(a), and (b) corresponds to Fig. 9(d). In Fig. 10(a) it is immediately obvious that the change in the charge neutrality condition that occurs at about $T = 300 \text{ K}$ arises from E_F crossing $\epsilon(3 + /+)$, which

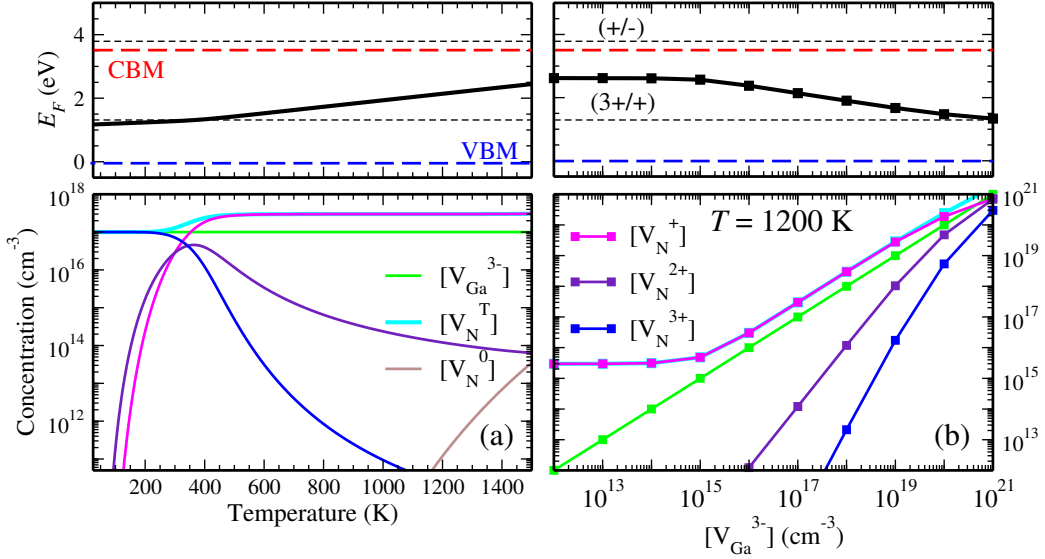


Figure 10: (a) The self-consistent Fermi energy E_F (thick black line in upper panel), total N vacancy concentration $[V_N^T]$ (thick cyan line) and concentrations of the individual charge states: $[V_N^0]$ (thin brown line), $[V_N^+]$ (thin magenta line), $[V_N^{2+}]$ (thin indigo line) and $[V_N^{3+}]$ (thin blue line), as a function of temperature, while $[V_{Ga}^{3-}]$ (thin green line) is kept fixed at 10^{17} cm⁻³. In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively, while the thin black dashed lines represent the V_N (3+/+) and (+/-) transition energies. (b) Similar to (a), but here E_F and the V_N concentrations are shown as a function of $[V_{Ga}^{3-}]$, for temperature $T = 1200$ K. The squares represent the data points; the lines are a guide for the eye.

changes the dominant charge state of V_N from 3+ to +, thus changing the concentration of V_N required to compensate the fixed $[V_{Ga}^{3-}]$. Interestingly, one can see that, around the point where $\epsilon(3+/+)$ is crossed, significant concentrations of V_N^{2+} are observed. At no value of E_F is 2+ the ground state; by just considering the transition level diagram (Fig. 3) this charge state would not be considered important, but the results in Fig. 10(a) demonstrate that it can occur in quite significant concentrations. V_N^{2+} can also be seen to play an important rôle in Fig. 10(b); for $[V_{Ga}^{3-}] < 10^{20}$ cm⁻³, E_F remains in the upper half of the band gap and the dominant charge state of V_N is +, with $[V_N^+] = [V_N^T]$. At $[V_{Ga}^{3-}] = 10^{21}$ cm⁻³, however, where $[V_N]$ moves closer to $[V_{Ga}^{3-}]$ (as discussed above), E_F approaches $\epsilon(3+/+)$, and $[V_N^+]$, $[V_N^{2+}]$ and $[V_N^{3+}]$ all contribute to the compensation of the V_{Ga}^{3-} .

4.5. Fixed concentrations of ionised donors and acceptors

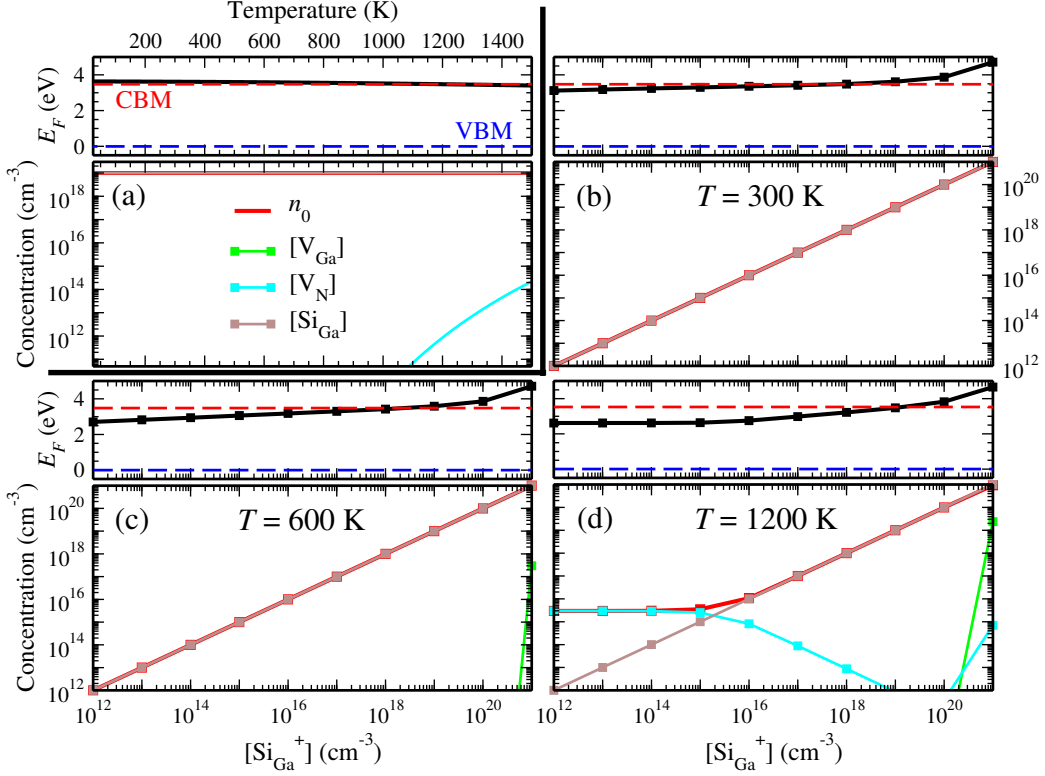


Figure 11: (a) The self-consistent Fermi energy E_F (thick black line in upper panel), electron carrier n_0 (thick red line) and defect concentrations (all in lower panel) as a function of temperature for intrinsic defects in GaN, when the concentration of an ionised Si donor in the + charge state $[\text{Si}_{\text{Ga}}^+]$ (thin brown line) is kept fixed at 10^{19} cm^{-3} . In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively. (b)–(d) Similar to (a), but here the self-consistent Fermi energy and concentrations are shown as a function of $[\text{Si}_{\text{Ga}}^+]$, for temperature $T = 300 \text{ K}$ (b), $T = 600 \text{ K}$ (c) and $T = 1200 \text{ K}$ (d). The Ga vacancy concentration $[\text{V}_{\text{Ga}}]$ is indicated by the thin green line and the N vacancy concentration $[\text{V}_{\text{N}}]$ by the thin cyan line. In (b)–(d) the squares represent the data points; the lines are a guide for the eye.

The results so far demonstrate how the set of intrinsic point defect formation energies presented in Table 1 can be used to compute equilibrium concentrations and analyse the effects of fixing particular total defect concentrations and concentrations of certain charge states of defects. FROZEN-SC-FERMI,

however, can also be used to study the effect other charged defects, not belonging to the original set and hence having unknown formation energies, have on the charge balance and consequent carrier and defect concentrations. Such an approach can be used to analyse, for example, how intrinsic defects respond to the presence of ionised impurities, such as donors and acceptors in semiconductors. In GaN, a commonly observed donor is Si substituting on a Ga site, Si_{Ga} , [55, 5, 69] which formally donates a single electron to the conduction band. Shown in Fig. 11 are the calculated self-consistent E_F (upper panels) and defect and carrier concentrations (lower panels) that result from a fixed concentration of a singly-ionised donor, denoted Si_{Ga}^+ , where in (a) T is varied and $[\text{Si}_{\text{Ga}}^+] = 10^{19} \text{ cm}^{-3}$ is fixed, and in (b)–(d) the value of the fixed $[\text{Si}_{\text{Ga}}^+]$ is varied while T is kept constant at (b) $T = 300 \text{ K}$, (c) $T = 600 \text{ K}$ and (d) $T = 1200 \text{ K}$. In some respects, the results shown here are trivial. As there are no compensating negative defects with low formation energies, for the full range of T in Fig. 11(a) the condition $n_0 = [\text{Si}_{\text{Ga}}^+]$ holds, i.e. the donors are fully activated. The same situation applies in Figs. 11(b) and (c), although in (c) at $[\text{Si}_{\text{Ga}}^+] = 10^{21} \text{ cm}^{-3}$, where E_F gets pushed up close to 5 eV, a small compensating concentration of V_{Ga} is observed. In Fig. 11(d), due to the elevated T , at values of $[\text{Si}_{\text{Ga}}^+] < 10^{15} \text{ cm}^{-3}$, there is a substantial concentration of V_{N} that is compensated by n_0 . for $[\text{Si}_{\text{Ga}}^+] > 10^{15} \text{ cm}^{-3}$, $n_0 = [\text{Si}_{\text{Ga}}^+]$ and E_F increases with $[\text{Si}_{\text{Ga}}^+]$, until at $[\text{Si}_{\text{Ga}}^+] = 10^{21} \text{ cm}^{-3}$, where, similar to in (c), E_F gets close to 5 eV and some compensating V_{Ga} are observed ($[\text{V}_{\text{N}}]$ also increases here, after initially decreasing, due to the dominance of the $-$ state, which becomes more favourable as E_F increases, see Fig. 3).

The situation that arises when a fixed concentration of acceptors is introduced is quite different. In Fig. 12, the calculated self-consistent E_F (upper panels) and defect and carrier concentrations (lower panels) that result from a fixed concentration of singly-ionised acceptors, denoted Mg_{Ga}^- , are shown, where in (a) T is varied and $[\text{Mg}_{\text{Ga}}^-] = 10^{19} \text{ cm}^{-3}$ is fixed, and in (b)–(d) the value of the fixed $[\text{Mg}_{\text{Ga}}^-]$ is varied while T is kept constant at (b) $T = 300 \text{ K}$, (c) $T = 600 \text{ K}$ and (d) $T = 1200 \text{ K}$. As there are low-formation-energy defects with positive charge states for E_F throughout the band gap, the most significant of which is the V_{N} , the negatively charged ionised acceptors are fully compensated and no significant hole concentrations are observed. In Figs. 12(a)–(c) charge neutrality is maintained by $[\text{Mg}_{\text{Ga}}^-] = [\text{V}_{\text{N}}]$ or $[\text{Mg}_{\text{Ga}}^-] = 3[\text{V}_{\text{N}}]$, depending on the value of E_F . This result implies that, given the formation energies of intrinsic point defects in Table 1, even if all

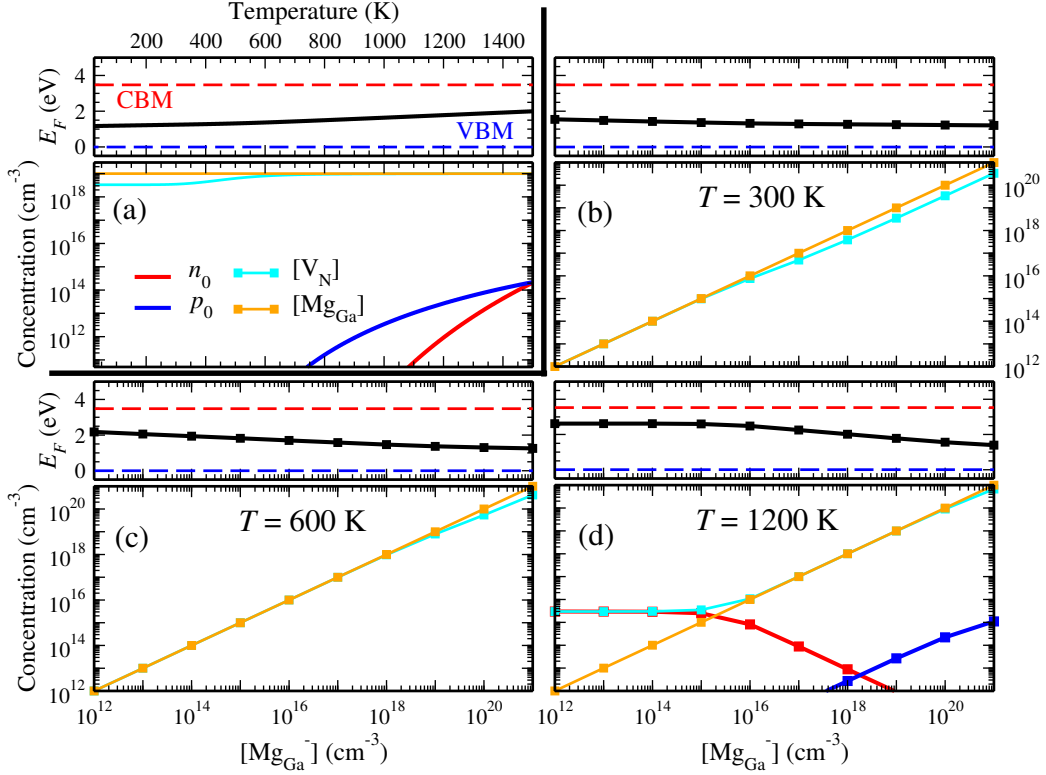


Figure 12: (a) The self-consistent Fermi energy E_F (thick black line in upper panel), electron carrier n_0 (thick red line), hole carrier p_0 (thick blue line) and defect concentrations (all in lower panel) as a function of temperature for intrinsic defects in GaN, when the concentration of an ionised Mg acceptor in the $-$ charge state $[\text{Mg}_{\text{Ga}}^-]$ (thin orange line) is kept fixed at 10^{19} cm^{-3} . In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively. (b)–(d) Similar to (a), but here the self-consistent Fermi energy and concentrations are shown as a function of $[\text{Si}_{\text{Ga}}^+]$, for temperature $T = 300 \text{ K}$ (b), $T = 600 \text{ K}$ (c) and $T = 1200 \text{ K}$ (d). The Ga vacancy concentration $[\text{V}_{\text{Ga}}]$ is indicated by the thin green line and the N vacancy concentration $[\text{V}_{\text{N}}]$ by the thin cyan line. In (b)–(d) the squares represent the data points; the lines are a guide for the eye.

the acceptors introduced to the system are fully ionised, they remain unactivated. Mg is the only known p -type dopant in GaN, and the associated acceptor state has an ionisation energy of at least $\sim 0.2 \text{ eV}$, [56, 5, 70] meaning that significant thermal activation is required. The analysis here, based on the point defect energies, however, indicates that simple substitution can

not result in p -type conductivity; instead non-equilibrium processes, possibly involving more complex defects, which suppress V_N formation in some way must dominate. In Fig. 12(d), at $T = 1200$ K it is observed that significant $[V_N]$ is present even for $[Mg_{Ga}^-] < 10^{15}$ cm^{-3} , which gives rise to $n_0 = [V_N]$. For $[Mg_{Ga}^+] > 10^{15}$ cm^{-3} , E_F is pushed further into the gap, n_0 decreases and $[Mg_{Ga}^-] = [V_N]$.

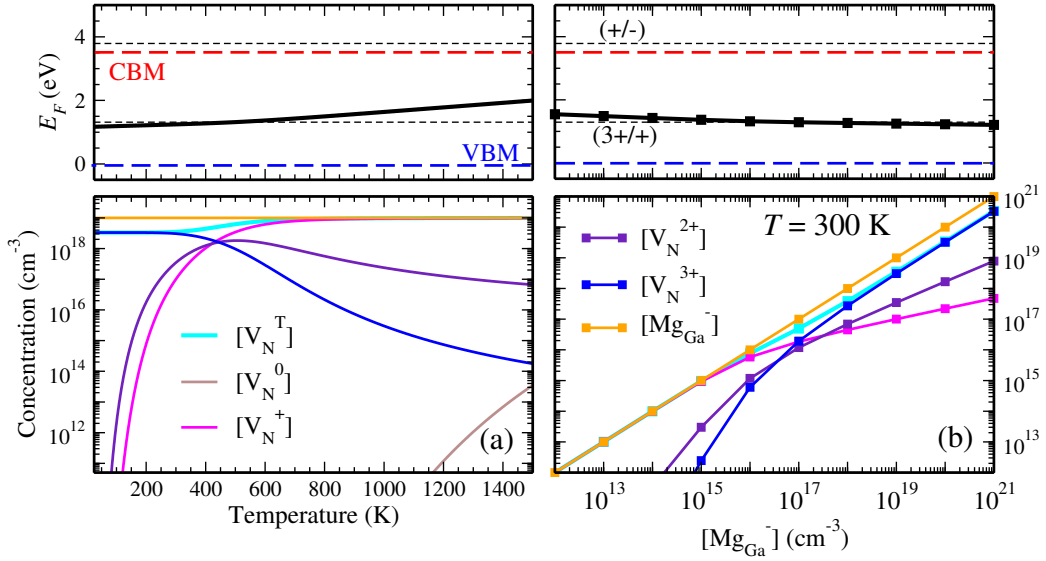


Figure 13: (a) The self-consistent Fermi energy E_F (thick black line in upper panel), total N vacancy concentration $[V_N^T]$ (thick cyan line) and concentrations of the individual charge states: $[V_N^0]$ (thin brown line), $[V_N^+]$ (thin magenta line), $[V_N^{2+}]$ (thin indigo line) and $[V_N^{3+}]$ (thin blue line), as a function of temperature, while $[Mg_{Ga}^-]$ (thin orange line) is kept fixed at 10^{19} cm^{-3} . In the upper panel, the blue and red dashed lines indicate the energies of the valence band maximum (VBM) and conduction band minimum (CBM), respectively, while the thin black dashed lines represent the V_N $(3+/-)$ and $(+/-)$ transition energies. (b) Similar to (a), but here E_F and the V_N concentrations are shown as a function of $[Mg_{Ga}^-]$, for temperature $T = 300$ K. The squares represent the data points; the lines are a guide for the eye.

The change from the condition $[Mg_{Ga}^-] = 3[V_N]$ to $[Mg_{Ga}^-] = [V_N]$ in Fig. 12(a), and *vice versa* in Fig. 12(b) and (c), is shown in more detail in Fig. 13(a) (corresponding to Fig. 12(a)) and Fig. 13(b) (corresponding to Fig. 12(b)). The upper panels show the self-consistent E_F and the lower panels show the concentrations of the total and individual charge state concentrations of V_N , as well as $[Mg_{Ga}^-]$, as (a) T is varied and $[Mg_{Ga}^-] = 10^{19}$

cm^{-3} is kept fixed, and (b) as the fixed $[\text{Mg}_{\text{Ga}}^-]$ is varied with $T = 300$ K is kept constant. In Fig. 13(a), it is clear that the transition from $[\text{Mg}_{\text{Ga}}^-] = 3[\text{V}_{\text{N}}]$ to $[\text{Mg}_{\text{Ga}}^-] = [\text{V}_{\text{N}}]$ at about 400 K occurs as E_F crosses $\epsilon(3 + /+)$ and the dominant charge state changes from $+$ to $3+$. Again, it is worth noting that the $2+$ charge state, the importance of which is not evident from Fig. 3, contributes significantly to $[\text{V}_{\text{N}}^T]$, particularly at values of T close to the transition point. The same effect is evident in Fig. 13(b), but in this case, as $[\text{Mg}_{\text{Ga}}^-]$ increases, E_F moves downwards within the band gap, crossing $\epsilon(3 + /+)$ at about $[\text{Mg}_{\text{Ga}}^-] = 10^{15} \text{ cm}^{-3}$. Therefore the transition goes from $[\text{Mg}_{\text{Ga}}^-] = [\text{V}_{\text{N}}]$ to $[\text{Mg}_{\text{Ga}}^-] = 3[\text{V}_{\text{N}}]$, and the importance of the $2+$ state is once again in evidence.

The results in this section demonstrate the power of the analysis available with `SC-FERMI` and `FROZEN-SC-FERMI`. Although some conclusions can be drawn simply from the transition level diagram, the programs facilitate quantitative analysis of intrinsic carrier concentrations and compensation mechanisms, but also provide insights not available from the defect formation energies alone, such as the importance of particular non-ground state configurations and division of defect concentration amongst charge states close to transition levels. Moreover, how carrier compensation changes with T or with impurity concentrations can be determined easily, while the dopability of a material can be quantified given the possible compensating defects. The above analysis was applied to a wide-gap semiconductor, but could equally well be applied to conventional semiconductors, insulators, narrow gap systems and (semi)metals.

5. Conclusions

A program to compute the self-consistent Fermi energy and equilibrium carrier and defect concentrations has been presented and the relevant theory has been described. The input required consists of the defect charge states, degeneracies, possible formation sites in the unit cell and formation energies; the density of states of the pristine system; the unit cell lattice vectors used to determine the volume; and the temperature. The concentrations and self-consistent Fermi energy, as well as a transition level diagram, comprise the output. An additional functionality, whereby certain defect concentrations (either total concentrations, or concentrations of particular charge states) can be fixed and the resulting self-consistent Fermi energy due to the new charge balance determined, is included and described. The operation of the

program is demonstrated using some relevant examples. The equations used to compute defect concentrations were derived using a fully quantum statistical mechanical approach, for the first time in such a generalised manner to the best of my knowledge. This code will benefit those analysing defect formation in a wide range of systems, where formation energies can be reliably computed.

Acknowledgment

The author acknowledges funding from EPSRC grant EP/K016288/1. I would like to thank S. R. Bishop, C. R. A. Catlow and A. A. Sokol for useful discussions. I would also like to thank Z. Xie and A. A. Sokol for allowing me to use some of our calculated formation energies of defects in GaN in this work.

- [1] A. M. Stoneham, *Theory of Defects in Solids*, Oxford University Press, New York, NY, 1975.
- [2] C. Wagner, *Ann. Rev. Mater. Sci.* 7 (1977) 1–24.
- [3] H. J. Queisser, E. E. Haller, *Science* 281 (1998) 945–950.
- [4] G. D. Watkins, *Phys. Solid State* 41 (1999) 746–750.
- [5] M. A. Reshchikov, H. Morkoç, *J. Appl. Phys.* 97 (2005) 061301.
- [6] F. Tuomisto, I. Makkonen, *Rev. Mod. Phys.* 85 (2013) 1583–1631.
- [7] B. Abbey, *JOM* 65 (2013) 1183–1201.
- [8] P. M. Fahey, P. B. Griffin, J. D. Plummer, *Rev. Mod. Phys.* 61 (1989) 289–384.
- [9] C. G. Van de Walle, J. Neugebauer, *J. Appl. Phys.* 95 (2004) 3851.
- [10] L. L. Baranowski, P. Zawadzki, S. Lany, E. S. Toberer, A. Zakutayev, *Semicond. Sci. Technol.* 31 (2016) 123004.
- [11] F. A. Kröger, H. J. Vink, volume 3 of *Solid State Physics*, Academic Press, 1956, pp. 307 – 435.
- [12] F. Kröger, H. Vink, *J. Phys. Chem. Solids* 5 (1958) 208 – 223.

- [13] R. F. Brebrick, *J. Phys. Chem. Solids* 4 (1958) 190 – 195.
- [14] R. Brebrick, *J. Phys. Chem. Solids* 18 (1961) 116 – 128.
- [15] F. A. Kröger (Ed.), *The Chemistry of Imperfect Crystals*, North-Holland, Amsterdam, 1974.
- [16] W. Walukiewicz, *Phys. Rev. B* 37 (1988) 4760–4763.
- [17] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, C. G. Van de Walle, *Rev. Mod. Phys.* 86 (2014) 253–305.
- [18] G. A. Baraff, E. O. Kane, M. Schlüter, *Phys. Rev. B* 21 (1980) 5662–5686.
- [19] S. B. Zhang, J. E. Northrup, *Phys. Rev. Lett.* 67 (1991) 2339–2342.
- [20] S. B. Zhang, S.-H. Wei, A. Zunger, *J. Appl. Phys.* 83 (1998) 3192–3196.
- [21] S. B. Zhang, S.-H. Wei, A. Zunger, *Phys. Rev. Lett.* 84 (2000) 1232–1235.
- [22] S.-H. Wei, *Comput. Mater. Sci.* 30 (2004) 337 – 348.
- [23] A. Walsh, C. R. A. Catlow, M. Miskufova, A. A. Sokol, *J. Phys.: Condens. Matter* 23 (2011) 334217.
- [24] C. R. A. Catlow, A. A. Sokol, A. Walsh, *Chem. Commun.* 47 (2011) 3386–3388.
- [25] A. Walsh, J. Buckeridge, C. R. A. Catlow, A. J. Jackson, T. W. Keal, M. Miskufova, P. Sherwood, S. A. Shevlin, M. B. Watkins, S. M. Woodley, A. A. Sokol, *Chem. Mater.* 25 (2013) 2924–2926.
- [26] G. Mandel, *Phys. Rev.* 134 (1964) A1073–A1079.
- [27] D. B. Laks, C. G. Van de Walle, G. F. Neumark, P. E. Blöchl, S. T. Pantelides, *Phys. Rev. B* 45 (1992) 10965–10978.
- [28] J. Buckeridge, <https://sourceforge.net/projects/sc-fermi>, 2016.
- [29] J. Buckeridge, <https://github.com/projects/sc-fermi.git>, 2016.

- [30] J. Buckeridge, D. Jevdokimovs, C. R. A. Catlow, A. A. Sokol, *Phys. Rev. B* 94 (2016) 180101.
- [31] F. H. Taylor, J. Buckeridge, C. R. A. Catlow, *Chem. Mater.* 28 (2016) 8210–8220.
- [32] D. B. Laks, C. G. Van de Walle, G. F. Neumark, S. T. Pantelides, *Phys. Rev. Lett.* 66 (1991) 648–651.
- [33] S. Lany, A. Zunger, *Phys. Rev. Lett.* 98 (2007) 045501.
- [34] K. K. Chin, *J. Semicond.* 32 (2011) 062001.
- [35] J. Ma, S.-H. Wei, T. A. Gessert, K. K. Chin, *Phys. Rev. B* 83 (2011) 245207.
- [36] J.-H. Yang, W.-J. Yin, J.-S. Park, S.-H. Wei, *Sci. Rep.* 5 (2015) 16977.
- [37] J.-H. Yang, W. K. Metzger, S.-H. Wei, *Appl. Phys. Lett.* 111 (2017) 042106.
- [38] Y.-N. Wu, X.-G. Zhang, S. T. Pantelides, *Phys. Rev. Lett.* 119 (2017) 105501.
- [39] C. Kittel, H. Kroemer, *Thermal Physics*, W. H. Freeman and Company, second edition, 1980.
- [40] J. Buckeridge, D. Jevdokimovs, C. R. A. Catlow, A. A. Sokol, *Phys. Rev. B* 93 (2016) 125205.
- [41] A. A. Abrikosov, *Introduction to the Theory of Normal Metals*, Academic Press, New York, NY and London, UK, 1972.
- [42] A. Walsh, A. A. Sokol, J. Buckeridge, D. O. Scanlon, C. R. A. Catlow, *J. Phys. Chem. Lett.* 8 (2017) 2074–2075.
- [43] A. Walsh, A. A. Sokol, C. R. A. Catlow, *Phys. Rev. B* 83 (2011) 224105.
- [44] B. Grabowski, T. Hickel, J. Neugebauer, *Phys. Status Solidi B* 248 (2010) 1295–1308.

- [45] B. Grabowski, T. Hickel, J. Neugebauer, Formation Energies of Point Defects at Finite Temperatures, Wiley-VCH, Verlag GmbH, pp. 259–284.
- [46] B. Cheng, M. Ceriotti, Phys. Rev. B 97 (2018) 054102.
- [47] E. G. Seebauer, M. C. Kratzer, Fundamentals of Defect Ionization and Transport, Springer London, London, pp. 5–37.
- [48] R. A. Swalin, Classification of Defects in Crystals, John Wiley and Sons, New York, second edition, pp. 259–266.
- [49] N. W. Ashcroft, N. D. Mermin, Solid State Physics, Holt, Rinehart and Winston, New York, pp. 561–589.
- [50] J. Buckeridge, D. O. Scanlon, A. Walsh, C. R. A. Catlow, Comput. Phys. Commun. 185 (2014) 330 – 338.
- [51] R. Z. Bachrach, B. S. Krusor, J. Vac. Sci. Technol. 18 (1981) 756–764.
- [52] S. Nakamura, M. Krames, Proc. IEEE 101 (2013) 2211–2220.
- [53] T.-Y. Seong, J. Han, H. Amano, H. Morkoç (Eds.), III-Nitride Based Light Emitting Diodes and Applications, Springer, Berlin, 2013.
- [54] B. Monemar, Phys. Rev. B 10 (1974) 676–681.
- [55] H. Morkoç (Ed.), Handbook of Nitride Semiconductors and Devices, volume 1, Wiley-VCH, Weinheim, 2008.
- [56] J. Buckeridge, C. R. A. Catlow, D. O. Scanlon, T. W. Keal, P. Sherwood, M. Miskufova, A. Walsh, S. M. Woodley, A. A. Sokol, Phys. Rev. Lett. 114 (2015) 016405.
- [57] G. Miceli, A. Pasquarello, Phys. Rev. B 93 (2016) 165207.
- [58] I. C. Diallo, D. O. Demchenko, Phys. Rev. Applied 6 (2016) 064002.
- [59] G. Miceli, A. Pasquarello, Microelectron. Eng. 147 (2015) 51 – 54. Insulating Films on Semiconductors 2015.
- [60] J. L. Lyons, C. G. Van de Walle, npj Comput. Mater. 3 (2017) 12.

- [61] Z. Xie, Y. Sui, J. Buckeridge, C. R. A. Catlow, T. W. Keal, P. Sherwood, A. Walsh, M. R. Farrow, D. O. Scanlon, S. M. Woodley, A. A. Sokol, arXiv:1803.06273 (2018).
- [62] A. A. Sokol, S. T. Bromley, S. A. French, C. R. A. Catlow, P. Sherwood, *Int. J. Quantum Chem.* 99 (2004) 695–712.
- [63] A. A. Sokol, S. A. French, S. T. Bromley, C. R. A. Catlow, H. J. J. van Dam, P. Sherwood, *Faraday Discuss.* 134 (2007) 267–282.
- [64] C. R. A. Catlow, J. Buckeridge, M. R. Farrow, A. J. Logsdail, A. A. Sokol, *Quantum Mechanical/Molecular Mechanical (QM/MM) Approaches*, WileyVCH Verlag, Weinheim, Germany, pp. 647–680.
- [65] J. Buckeridge, C. R. A. Catlow, M. R. Farrow, A. J. Logsdail, D. O. Scanlon, T. W. Keal, P. Sherwood, S. M. Woodley, A. A. Sokol, A. Walsh, *Phys. Rev. Materials* 2 (2018) 054604.
- [66] P. J. Wilson, T. J. Bradley, D. J. Tozer, *J. Chem. Phys.* 115 (2001) 9233–9242.
- [67] L. J. L., A. Audrius, J. Anderson, V. de Walle Chris G., *Phys. Status Solidi B* 252 (2015) 900–908.
- [68] Z. Xie, Y. Sui, J. Buckeridge, A. A. Sokol, T. W. Keal, A. Walsh, arXiv:1802.10222 (2018).
- [69] Z. Xie, Y. Sui, J. Buckeridge, C. R. A. Catlow, T. W. Keal, P. Sherwood, A. Walsh, D. O. Scanlon, S. M. Woodley, A. A. Sokol, *Phys. Status Solidi A* 214 (2017) 1600445.
- [70] S. Lany, A. Zunger, *Appl. Phys. Lett.* 96 (2010) 142114.