8

Energy Conversion: Solid-State Lighting

E. Kioupakis[1,2]*, P. Rinke*[1,3]*, A. Janotti*[1]*, Q. Yan*[1]*, and C.G. Van de Walle*[1]
[1]*Materials Department, University of California, Santa, Barbara, USA*
[2]*Department of Materials Science and Engineering, University of Michigan, Ann Arbor, USA*
[3]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany*

8.1 Introduction to Solid-State Lighting

Artificial illumination is a catalyst for development and improves the quality of life, but it also consumes a large fraction of the world's energy resources [1]. Lighting has been one of the earliest applications of electric power and today utilizes approximately 20% of the world's grid-based electricity production. The inefficiency of existing light sources that waste most of the power they consume is the reason for this large share. The worldwide demand for lighting is increasing rapidly, in particular in the developing world, and is projected to compound its strain on available energy sources. Therefore, there is a strong motivation for improving the efficiency of light sources and mitigating their effect on worldwide energy consumption and global climate change.

Historically, electric light bulbs have been of the incandescent type. Although this technology was developed more than 100 years ago, it is still in use today. Incandescent light bulbs operate by heating a tungsten wire in a vacuum bulb to a temperature of 3500 K. As a result of resistive heating, the wire glows white hot and emits a broad black-body radiation spectrum that covers the entire visible

Computational Approaches to Energy Materials, First Edition. Edited by Aron Walsh, Alexey A. Sokol and C. Richard A. Catlow.

range. A variation is the halogen lamp, which contains a mix of halogen and inert gases in the bulb. The halogen redeposits the evaporated tungsten back on to the wire, which allows the bulb to operate at a higher temperature. However, the efficiency of incandescent light bulbs is remarkably low, of the order of 5%, and their lifetimes are typically on the order of a thousand hours. This is the main reason why governments around the world are implementing legislation to phase out incandescent light bulbs for general lighting in the near future.

Fluorescent light bulbs are a more efficient and long-lasting alternative, but they also have their limitations. The light bulb is made of a fluorescent tube that contains an inert gas and small amounts of mercury. As the electric current passes through the gas, it excites the mercury atoms, causing them to emit ultraviolet radiation, which gets converted to visible light by the tube's fluorescent coating. Fluorescent light bulbs have efficiencies of the order of 20% and typical lifetimes of 10 000 hours. This is much better than for incandescent bulbs, but still leaves considerable room for improvement. In addition, the mercury contained in the light bulb poses a severe health hazard and needs to be safely recycled or disposed.

Solid-state lighting based on light-emitting diodes (LEDs) is a promising recent development and has the potential to reshape the lighting industry [2]. LEDs can reach energy-conversion efficiencies exceeding 50%, significantly reducing the energy cost of lighting. They can also achieve lifetimes as long as 50 000 hours, therefore reducing maintenance costs and the amount of electronic waste associated with lighting, and do not contain toxic elements. LEDs have long found applications as indicator lights, for example, in watches and calculators. Currently they are used for display backlights in portable computers and cellphones, and are also available for bicycle and car headlights. The total cost of ownership of LED bulbs is currently comparable with that of fluorescent ones and their market share for general lighting applications is increasing rapidly.

Semiconductors are the materials of choice for solid-state lighting applications [3]. Light emission from a semiconductor was first observed in SiC in 1907. The rapid developments in groups III–V compound semiconductor research in the 1960s and 1970s led to the fabrication of efficient LEDs in the long-wavelength part of the visible spectrum. However, white-light emission for lighting applications requires LEDs emitting in the blue or violet part of the spectrum. Research efforts initially focused on ZnSe, but defects in the material deteriorated the device performance and limited their lifetimes. A breakthrough happened in the early 1990s, when Shuji Nakamura developed the first viable blue LEDs based on group III nitrides [4], which are the dominant short-wavelength LED materials to this day.

The major group III nitrides used for LEDs, namely AlN, GaN, and InN, have been the subject of extensive research [5]. All three crystallize in the hexagonal wurtzite structure with four atoms in the primitive unit cell. Their electronic band gaps, which determine the light-emission wavelength, range from 6.2 eV for AlN to 3.4 eV for GaN and 0.7 eV for InN. Therefore, alloying GaN with InN yields

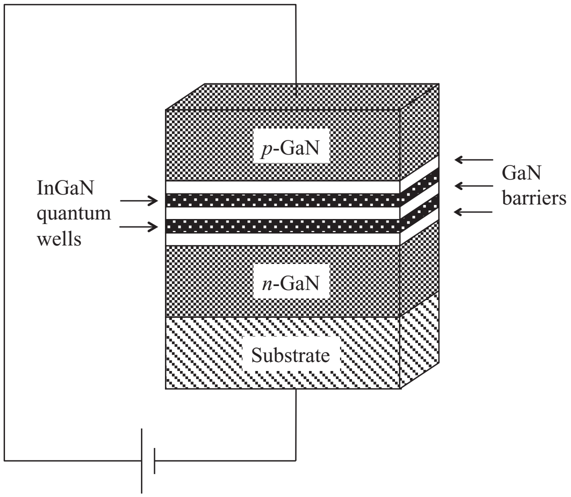


Figure 8.1 Schematic illustration of a typical nitride light-emitting diode. The active region, consisting of InGaN quantum wells and GaN barriers, is sandwiched between the n- and p-type GaN layers. The device is usually grown on top of a foreign substrate, typically sapphire or SiC.

a material with a direct band gap that is tunable in the entire visible range—a remarkable property that cannot be achieved with any other materials system at the moment. Doping GaN with Si and Mg yields *n*- and *p*-type-doped materials, respectively, that can be used for the fabrication of diodes.

A typical LED structure is shown in Figure 8.1. Bulk nitrides are difficult to fabricate, so devices are usually grown on a foreign substrate, typically sapphire (Al_2O_3) or SiC. The *n*-type Si:GaN layer is deposited first, followed by the growth of InGaN quantum wells and GaN barriers, and finally the *p*-type Mg:GaN layer. Not shown in Figure 8.1 are the transparent oxide top contact and the AlGaN electron-blocking layer that is commonly employed to enhance electron capture in the quantum wells. Connecting the device to a power source injects electrons and holes from the adjacent doping layers into the quantum wells, where they recombine and produce light at the wavelength corresponding to the band gap of the alloy material.

There are several approaches to achieve white-light emission appropriate for indoor lighting applications with nitride LEDs [2]. The simplest approach is to combine a blue LED with a yellow phosphor coating, which downconverts the high-energy blue photons to other visible wavelengths and produces a broad emission spectrum. However, the color quality that can be achieved with this approach is limited. Combining an ultraviolet LED with a mix of phosphors produces white light more suitable for indoor use but is less efficient. Another approach is to combine three or more LEDs of different colors, which is the best option for efficiency and color quality but comes at a higher cost.

The growth of nitride LEDs along nonpolar or semipolar crystallographic directions is a recent development that can improve the efficiency of devices and is a subject of active research [6]. Nitride LEDs are typically grown along the (0001) crystallographic direction of the hexagonal wurtzite structure, also known as the *c*-axis orientation. The nitride materials, however, are polar along this axis and, as a result of strained growth as well as spontaneous polarization, strong internal electric fields develop inside the InGaN quantum wells. These polarization fields keep the injected electrons and holes spatially separated and impede their efficient recombination, which affects the device efficiency. Different crystallographic growth orientations, however, can produce devices with reduced or even zero internal polarization fields, a fact that can substantially enhance their performance.

The study of nitride materials for light emission is a highly active research area with strong growth prospects. Current research focuses on improving the efficiency and reducing the cost of LEDs in order to deliver affordable light bulbs that can replace the existing solutions for general lighting. Growth of semipolar and nonpolar devices offers additional control over the device properties and is a promising area for new developments. Moreover, novel light-emitting devices have been fabricated based on even smaller nitride nanostructures, such as nanowires [7] or quantum dots [8]. These new developments guarantee that research interest in the field of nitride materials will remain strong.

First-principles computational approaches based on density-functional and many-body perturbation theory, introduced in Chapter 1, are a powerful tool set that has complemented and guided the experimental research efforts on nitride materials for LEDs. These methods yield equilibrium crystal structures of materials accurate to within 1% as well as accurate electronic properties, such as band gaps and carrier effective masses. In addition, these first-principles methods can be employed to understand and predict the properties of defects that are crucial for optoelectronic performance and device longevity. Finally, the radiative efficiency and the origin of the nonradiative losses in LEDs can be investigated and understood based on the results of theoretical calculations. In the following, we will present the corresponding computational methodology, along with specific results for group III nitrides used for solid-state lighting. These examples highlight the importance and predictive power of first-principles computational techniques for the study of technologically important materials and the performance of modern optoelectronic devices.

8.2 Structure and Electronic Properties of Nitride Materials

8.2.1 Density Functional Theory and Ground-State Properties

The most widely used computational method for electronic structure studies of materials properties is density functional theory (DFT, Section 1.3.3 and [9]). The central quantity of the theory is the total energy $E_{\text{tot}}[n]$, which is a functional of the electron density $n(r)$ and attains its minimum at the exact ground-state density. Using the Born–Oppenheimer approximation, the functional for the electronic energy can be divided into different contributions:

$$E_{\text{tot}}[n] = T_{\text{s}}[n] + E_{\text{ext}}[n] + E_{\text{H}}[n] + E_{\text{xc}}[n], \tag{8.1}$$

where $T_{\text{s}}[n]$ is the noninteracting kinetic energy, $E_{\text{ext}}[n]$ is the potential energy due to interaction with the ions, $E_{\text{H}}[n]$ is the classical electrostatic interaction (Hartree) term, and $E_{\text{xc}}[n]$ is the exchange–correlation energy.

Since the exact form of the exchange–correlation functional, which contains all the complexity of the quantum many-body interactions, is unknown, suitable approximations have to be found for practical calculations. Common approximations are the local density approximation (LDA), which is based on the correlation energy of the electron gas calculated with quantum Monte Carlo [10], or the generalized gradient approximation (GGA), which includes higher order gradient corrections [11]. Recent developments include the incorporation of exact exchange in the so-called hybrid functionals, which mix LDA or GGA with exact Hartree–Fock exchange. More advanced functionals combine exact exchange with correlation in the random phase approximation (RPA) [12, 13]. The exact-exchange energy cancels the spurious self-interaction error present in the Hartree term exactly, while the RPA correlation energy is fully nonlocal and accurately includes long-range van der Waals interactions.

Practical DFT calculations typically proceed in the context of Kohn–Sham theory, which maps the real system of interacting particles onto a fictitious system of noninteracting particles that have the same ground-state electron density and total energy as the real one. The Kohn–Sham equation that describes the motion of these noninteracting particles in the effective Kohn–Sham potential $v_{\text{KS}}[n](r)$ is

$$\left[-\frac{\hbar^2\nabla^2}{2m} + v_{\text{KS}}[n](r)\right]\varphi_i(r) = \varepsilon_i\varphi_i(r), \tag{8.2}$$

where ε_i and $\varphi_i(r)$ are, respectively, the energy eigenvalues and wavefunctions of the Kohn–Sham orbitals. The Kohn–Sham potential is a local potential that depends on the density and is given by

$$v_{\text{KS}}[n](r) = v_{\text{ext}}(r) + v_{\text{H}}[n](r) + v_{\text{xc}}[n](r),$$

where $v_{\text{ext}}(r)$ is the applied external potential (e.g., the potential of the ions), $v_{\text{H}}[n](r)$ is the Hartree term, and $v_{\text{xc}}[n](r)$ is the exchange–correlation potential.

The electron density of the Kohn–Sham system, $n(r) = \sum_i^{\text{occ}} |\varphi_i(r)|^2$, which comes from the solution of (8.2) is then inserted in (8.1) to yield the total energy of the electron system. The energy in (8.1) depends parametrically on the position of the nuclei. These can then subsequently be optimized to minimize the total energy. This yields equilibrium lattice parameters and atomic geometries, while the curvature

Table 8.1 *Equilibrium lattice parameters (a, c, and u) and bulk modulus (B) of the wurtzite-structured nitride materials used for solid-state lighting devices, as calculated with various exchange–correlation functionals within the DFT and measured by experiment ([14, 15] and references therein)*

Material	Method	a/Å	c/Å	c/a	u	B/Mbar
AlN	Theory, LDA	3.057	4.943	1.617	0.3802	2.09
	Theory, GGA	3.113	5.041	1.619	0.3798	1.92
	Theory, HSE	3.102	4.971	1.606	0.3819	–
	Experiment	3.111	4.978	1.601	0.385	1.85–2.12
GaN	Theory, LDA	3.193	5.218	1.634	0.376	2.02
	Theory, GGA	3.245	5.296	1.632	0.3762	1.72
	Theory, HSE	3.182	5.173	1.626	0.3772	–
	Experiment	3.180	5.166	1.624	0.375	1.88–2.45
InN	Theory, LDA	3.544	5.762	1.626	0.377	1.40
	Theory, GGA	3.614	5.884	1.628	0.377	1.161
	Theory, HSE	3.548	5.751	1.621	0.3796	–
	Experiment	3.533	5.693	1.611	0.375	1.25

of the energy-versus-volume curve provides the bulk modulus of the material. In general, structural parameters calculated within DFT are quite accurate and usually lie within a small percentage of their corresponding experimental values.

DFT within the LDA or GGA for the exchange–correlation functional provides accurate structural properties for the nitride materials [14]. The lattice parameters of the wurtzite crystal structure for AlN, GaN, and InN are listed in Table 8.1. The calculated results are in good agreement with experiment and the deviations are on the order of the typical 1% error of these functionals. The structural parameters obtained from more advanced hybrid functionals, such as Heyd, Scuseria, and Ernzehof (HSE) discussed later in Section 8.2.3, are in even better agreement with experiment than LDA or GGA.

8.2.2 Electronic Excitations: *GW* and Exact Exchange

The computational description of electronic excitations was introduced in Chapter 2; however, we will approach the problem from a complementary point of view. Although DFT performs exceptionally well in predicting the ground-state properties of materials, it fails to accurately describe properties that involve electronic excitations. In contrast to the electron density and total energy, which represent the exact properties of the interacting system, the Kohn–Sham eigenvalues in (8.2) are primarily mathematical constructs that do not formally correspond to excitations of the real system. In practice, however, they are a good zeroth-order approximation [16] to the quasi-particle energies as measured in direct and inverse photoemission and are therefore frequently used to describe the band structure of materials. The most prominent deficiency of Kohn–Sham band structures is the significant underestimation (in LDA and GGA typically by 50%) of band gaps. This error is a feature of the Kohn–Sham approach [17, 18] and arises even if the exact exchange–correlation potential is used.

As introduced in Chapters 1 and 2, many-body perturbation theory within the *GW* approximation for the self-energy operator can formally account for electronic excitations and is the method of choice for calculating quasi-particle band structures in solids [19–26]. Within many-body perturbation theory, the band structure of a material is given by the solution of the quasi-particle equation:

$$\left[-\frac{\hbar^2\nabla^2}{2m} + v_{\text{ext}}(r) + v_{\text{H}}[n](r)\right]\psi_i(r) + \int \Sigma(r, r'; \varepsilon_i^{\text{QP}})\psi_i(r')\text{d}r' = \varepsilon_i^{\text{QP}}\psi_i(r),$$

where $\Sigma(r, r'; \varepsilon)$ is the self-energy operator. The self-energy is a non-Hermitian, non-local, and energy-dependent operator that contains all the complexity of exchange and correlation of the many-electron system. The *GW* approximation consists of expanding the self-energy operator in terms of Feynman diagrams with respect to the screened Coulomb interaction and keeping only the first-order term:

$$\Sigma \approx \Sigma_{GW} = iGW,$$

where G is the Green's function and W the screened Coulomb interaction. The dielectric function $\varepsilon(r, r'; \omega)$ needed to describe the screened interaction between charge carriers at r and r' is then given within the RPA. In order to properly describe screening in semiconductors, it is important to account for local field effects, that is, the explicit dependence on r and r', as well as the frequency dependence of the dielectric function. The calculation of the dielectric function makes the *GW* approach computationally very demanding, and several computational techniques and approximations have been developed to make these calculations tractable [26].

Band structure studies within the *GW* method usually start with a DFT calculation for the ground-state properties and proceed by applying the self-energy operator as a first-order perturbation to the Kohn–Sham system. The Green's function and the screened interaction are calculated from the Kohn–Sham orbitals and eigenvalues. The quasi-particle energies $\varepsilon_i^{\text{QP}}$ are given by the following perturbation theory expression in terms of matrix elements of the self-energy operator and exchange–correlation potential with respect to the Kohn–Sham orbitals:

$$\varepsilon_i^{\text{QP}} = \varepsilon_i + \left\langle \varphi_i \middle| \Sigma_{GW}(\varepsilon_i^{\text{QP}}) - v_{xc} \middle| \varphi_i \right\rangle.$$

The quasi-particle wavefunctions ψ_i can be calculated in terms of the Kohn–Sham states by diagonalizing the full self-energy matrix, including off-diagonal elements. However, in *sp*-bonded solids, the Kohn–Sham orbitals typically differ by less than 0.1% from the corresponding quasi-particle wavefunctions and can therefore be taken as a very good approximation for the latter [27]. The quasi-particle energies and wavefunctions could in principle also be used to update the Green's function and the screening until self-consistency is achieved. A more detailed discussion of quasi-particle wavefunctions and the advantages and disadvantages of self-consistency would go beyond the scope of this chapter and we refer the interested reader to the literature [21, 22, 24, 26]. *GW* calculations for bulk materials are therefore usually performed as a one-shot correction, a procedure denoted G_0W_0.

Although LDA and GGA are most commonly used as the starting point for G_0W_0 in solids [20–22], orbital-dependent functionals [16] may provide a better description for certain materials. For example, ground-state functionals that address the self-interaction error of DFT have proven to be particularly advantageous for materials with *d* and *f* electrons [24, 28–30]. For the nitrides this becomes most apparent for the band gap of InN, for which LDA and GGA result in an overlap between the conduction and the valence bands and thus an effectively metallic state [31, 32]. A parametrization derived from such a band structure would not appropriately reflect the properties of bulk InN.

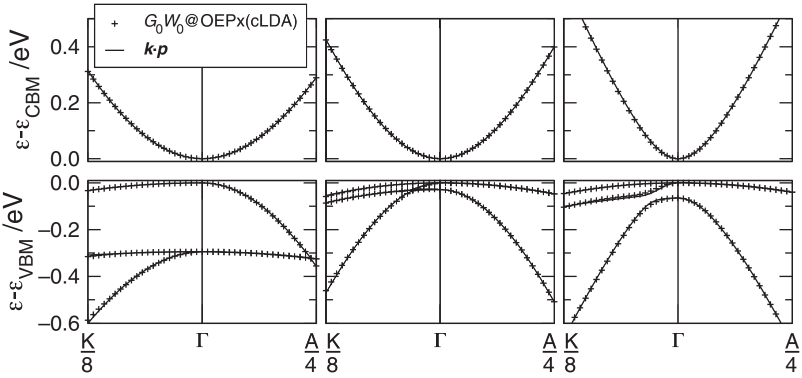
Orbital-dependent functionals have been used as the starting point of G_0W_0 calculations for group III nitrides and yield good agreement with experiment. In this context, the optimized effective potential approach applied to exact-exchange (OEPx) produces the variationally best local potential to the nonlocal Fock operator [33]. In OEPx(cLDA) standard LDA correlation is added to OEPx. Unlike in LDA and GGA, the self-interaction error in OEPx(cLDA) is greatly reduced (OEPx is fully self-interaction free) and correctly predicts InN to be semiconducting with the right band ordering in the wurtzite phase [34, 35]. Combining OEPx(cLDA) with G_0W_0, G_0W_0@OEPx(cLDA), yields band gaps for groups II and VI compounds, Ge, ScN, and group III nitrides that agree to better than 0.3 eV with experiment (Table 8.2) [24, 25, 32, 35, 36]. This is currently the best accuracy one can expect from a state-of-the-art first-principles theory that does not rely on fitting parameters.

The shallow cation semicore *d* levels of nitride materials strongly affect the calculated electronic structure [37]. First-principles approaches typically make use of pseudopotentials to describe the interaction between electrons and ions. The pseudopotential approximation significantly reduces the computational cost of plane-wave calculations because it freezes the inner shell electrons in the core and reduces the number of Fourier components needed to describe the resulting smooth valence wavefunctions. However, core-polarization effects are important

Table 8.2 *Band gaps of group III nitride materials as calculated with DFT and GW [15, 24] and measured by experiment*

Material	LDA	HSE	OEPx(cLDA)	G_0W_0@OEPx(cLDA)	Experiment
AlN	4.29	5.64	5.73	6.47	6.0–6.3
GaN	1.78	3.24	3.15	3.24	3.5
InN	–	0.68	1.00	0.69	0.65

All values are in eV.



(a) (b) (c)

Figure 8.2 *Electronic band structure of (a) AlN, (b) GaN, and (c) InN as calculated with G_0W_0@OEPx(cLDA) (symbols) and parameterized with a $k \cdot p$ Hamiltonian (lines).*

for materials with shallow *d* levels, such as group III nitrides. *GW* calculations starting from LDA for systems with semicore *d* electrons in general show poor agreement with experiment, unless the *s* and *p* orbitals of the semicore *d* shell are also treated explicitly in the valence [37, 38]. This accounts for the polarization of the ionic cores, but the strong spatial localization of these semicore orbitals substantially increases the cost of plane-wave calculations. On the other hand, pseudopotentials calculated with the OEPx(cLDA) method properly account for the exchange interaction between the localized semicore orbitals and allow the semicore *s* and *p* states to be frozen in the pseudopotential core [24, 25].

The electronic structure near the band edges determines the carrier effective masses and hence the performance of optoelectronic devices. *GW* calculations can also determine the band parameters of nitrides that can be used in device simulations. The G_0W_0@OEPx(cLDA) band structures of group III nitrides near the band edges are shown in Figure 8.2 [32]. A single state forms the lowest conduction band, but the top of the valence band involves three states—two degenerate heavy- and light-hole bands, and the crystal-field-split level. While the bands are parabolic near the Γ point, nonparabolicities develop away from Γ in particular for InN. The calculated crystal-field splittings and carrier effective masses are also in good agreement with available experimental data [15, 32].

Materials parameters can be derived from first-principles electronic structure methods for bulk phases, but the size and complexity of structures required for device simulations currently exceeds the capabilities of first-principles electronic structure tools. To bridge this gap, first-principles calculations can be used to parameterize simplified methods such as the $k \cdot p$ approach [39]. This methodology can then easily be implemented in device simulation software to determine the band structure as a function of position and operating conditions for the complex layered heterostructures used in modern optoelectronic devices. Figure 8.2 displays the band structures of group III nitrides in the vicinity of the Γ point as determined by a 4×4 $k \cdot p$ model. The band dispersion parameters of the model have been fitted to the G_0W_0@OEPx(cLDA) bands for the corresponding material [32], and capture both the nonparabolicity of the conduction band and the complexity of the valence band structure. This consistent set of band parameters can in turn be used to model the electronic properties of nitride alloys, for example, by interpolating from the pure phases.

8.2.3 Electronic Excitations: Hybrid Functionals

Another method that can simultaneously yield reliable structural and electronic properties of nitride semiconductors is hybrid density functional theory. Hybrid functionals mix the usual local exchange potentials, such as LDA or GGA, with the exact nonlocal Hartree–Fock exchange potential, while correlation is described within the LDA or GGA. This approach originated in quantum chemistry, and in particular the B3LYP functional, which combines Becke's exchange functional (B3) with Lee, Yang, and Parr (LYP) correlation, has been quite successful in describing structural properties and energetics of molecules [40]. Range-separated hybrid functionals can be used to study bulk semiconductors by introducing a screening length, as proposed by Heyd, Scuseria, and Ernzehof (HSE) [41,42]. The HSE functional divides the exchange energy into a short- and a long-range part. The short-range part mixes the GGA exchange of Perdew, Burke, and Ernzerhof (PBE) [11] with Hartree–Fock exchange in a 3:1 ratio. The long-range exchange-correlation potential is described by the PBE functional. The additional computational cost involved with HSE calculations, compared to the more usual functionals, stems from the calculation of four-center integrals for the Hartree–Fock potential, and typical HSE runs require at least an order of magnitude, more computational resources than standard DFT calculations.

Hybrid functionals produce accurate structural parameters while significantly reducing the band-gap error of local and semilocal DFT functionals. The equilibrium lattice parameters for group III nitrides as obtained with HSE are in better agreement with experiment than the corresponding LDA or GGA values, as shown in Table 8.1 [15]. In addition, the electronic band gaps calculated with HSE are in much better agreement with experiment than LDA or GGA and of comparable accuracy to G_0W_0@OEPx(cLDA) (Table 8.2). Hybrid functional calculations are thus a powerful technique that can simultaneously yield accurate structural and electronic properties for group III nitrides.

8.2.4 Band-gap Bowing and Band Alignments

Band structure calculations using the HSE hybrid functional have been performed to investigate the band-gap bowing of InGaN alloys [43]. When GaN with a band gap of 3.24 eV is alloyed with InN with a band gap of 0.68 eV, an alloy forms that can have a gap value anywhere between these two extremes. The equation that determines the alloy band gap is

$$E_g(\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}) = (1-x)E_g^{\mathrm{GaN}} + xE_g^{\mathrm{InN}} - bx(1-x),$$

where the first two terms represent the linear average between the two extrema (Vegard's law), while the nonlinear term expresses the band-gap bowing. The nonlinear behavior originates from composition fluctuations due to the statistical randomness in cation-site occupation either by In or Ga. The randomness leads to fluctuations of the local electrostatic potential arising from the different atomic potentials as well as to variations of the local microscopic structure (bond lengths and angles). The bowing parameter b quantifies the impact of these two effects on the band gap. HSE calculations have been employed to determine the band gaps of random InGaN alloy supercells of varying composition. The calculations reveal a strong band-gap bowing, which depends on In composition. The value of b is 1.10 eV if a single bowing parameter is used for the entire alloy range. There is, however, stronger bowing at low In content, that is, the regime usually employed for optoelectronic devices, and the bowing parameter is $b = 2.29$ eV at $x = 6.125\%$ and $b = 1.79$ eV at $x = 12.5\%$.

The band alignment between two different nitride materials, which is important for the operation of optoelectronic devices, can also be determined with hybrid functionals [43]. The natural band alignments between different nitrides are referenced to the vacuum level and reflect the relative band arrangement of the unstrained materials. These can be derived from separate bulk and surface calculations. The positions of the valence and conduction bands can be calculated with respect to the average electrostatic potential in the material from bulk calculations, while the alignment of the electrostatic potential relative to the vacuum level can be determined from a surface calculation in a nonpolar slab geometry. The calculations reveal that the valence band offset between GaN and InN is 0.62 eV, in good agreement with experiment, and varies linearly with alloy composition, while most of the band-gap bowing is manifested in the conduction band. This is expected since the valence band is composed of *N* orbitals, while the conduction band consists of randomly ordered cation orbitals. The natural band alignments account for bulk effects but do not consider the effect of interfaces between the two materials, which can be sizeable and must be studied explicitly for each specific interface geometry.

8.2.5 Strain and Deformation Potentials

One effect of the interface between dissimilar nitride materials, such as the InGaN quantum wells grown between GaN layers in heterostructures, is the development of strain. Vegard's law predicts that the equilibrium lattice constant of the alloy will be at a linearly interpolated value between the two extreme phases and will in general be different from either of them. As a result, nitride alloy layers that are pseudomorphically grown on GaN will be strained, which will affect their electronic properties. Strain changes the absolute position of the valence and conduction band extrema and hence modifies the quantum well depth and confinement of electrons and holes in the active region [44]. It also induces piezoelectric polarization, which greatly lowers the electron–hole overlap and hence the radiative recombination rate in quantum wells [45]. In addition, biaxial strain modifies the details of the valence band structure of nitride alloys in c-plane InGaN/GaN devices and changes the transition energies and effective masses of carriers [46]. Finally, biaxial strain in nonpolar and semipolar quantum wells drastically modifies the subband structure and wavefunction characteristics and can induce polarized light emission [47].

The effects of strain on the band structure of group III nitrides can be described by deformation potentials, defined as the linear coefficient in the expansion of the band energies with respect to strain. The concept of deformation potentials originates from $k \cdot p$ perturbation theory [48]. In this approach, the lowest conduction and topmost three valence bands are the eigenstates of an 8×8 Hamiltonian. The effects of a strain perturbation are described by two conduction band (a_{ct} and a_{cz}) and six valence band deformation potentials (D_1 to D_6). The deformation potentials a_{cz}, D_1, and D_3 are related to strain in the c-plane ($\varepsilon_{\perp} = \varepsilon_{xx} + \varepsilon_{yy}$), while a_{ct}, D_2, and D_4 are related to the out-of-plane component ($\varepsilon_{\parallel} = \varepsilon_{zz}$). The deformation potential D_5 determines the effect of anisotropic in-plane strain $\left|\varepsilon_{xx} - \varepsilon_{yy}\right|$ on the band structure, especially the splitting of the heavy- and light-hole valence bands, and the D_6 component accounts for shear strain (ε_{xy}, ε_{xz}, and ε_{yz}) in semipolar nitride materials.

The experimental determination of deformation potentials is difficult. The measured data for GaN are scattered over a wide range of values, while for InN only one set of measured data have been reported so far [49]. For AlN no data are available yet. Most of the experimental data so far have been obtained by measuring the change of optical transition energies under biaxial stress induced by the c-plane substrate. However, uniaxial and biaxial stress cannot be applied separately, and hence a combination of several deformation potentials results. The quasi-cubic approximation is then applied to extract the deformation potentials from the combined data. Another experimental approach is to carry out optical measurements under uniaxial stress. The advantage of this approach is that various strain conditions can be applied without having to resort to the quasi-cubic approximation. This technique has produced a complete set of deformation potentials for GaN [50].

A consistent set of deformation potentials for group III nitrides can be obtained from first-principles calculations [51, 52]. The band structure of each material is calculated within LDA, GGA, or HSE for the corresponding relaxed crystal structure under various applied strain conditions. In particular, the internal structural parameter u sensitively affects the magnitude of the crystal field splitting and must be fully relaxed. The deformation potentials obtained by fitting the analytical $k \cdot p$ eigenenergies to the calculated band structures at the Γ point are listed in Table 8.3 and compared to experiment. The band-gap-related deformation potentials

Table 8.3 *Deformation potentials for group III nitrides as calculated with various first-principles methods and measured by experiment*

	Method	$a_{cz} - D_1$	$a_{ct} - D_2$	D_3	D_4	D_5	D_6
AlN	LDA	−3.44	−11.39	8.97	−3.95	−3.36	–
	GGA	−3.39	−11.38	9.12	−4.01	−3.37	–
	HSE	−4.21	−12.07	9.22	−3.74	−3.30	−4.49
GaN	LDA	−4.56	−8.03	5.61	−3.03	−2.94	–
	GGA	−4.46	−8.08	5.83	−2.98	−3.13	–
	HSE	−6.02	−8.98	5.45	−2.97	−2.87	−3.95
	GW	−5.49	−8.84	5.80	−3.10	–	–
	Experiment	−9.6 to −3.1	−11.8 to −8.1	1.4 to 8.2	−4.1 to −0.7	−4.7 to −2.4	–
InN	HSE	−3.64	−4.58	2.68	−1.78	−2.07	−3.02

All values are in eV.

($a_{cz} - D_1$ and $a_{ct} - D_2$) for GaN and AlN calculated with LDA and GGA agree with each other but appreciably deviate from the HSE results. This discrepancy has been attributed to the band-gap problem of these functionals. For InN, the negative band gap predicted by the same functionals inhibits the determination of deformation potentials. In contrast, first-principles approaches that account for corrections to the band gap, such as HSE and G_0W_0@OEPx(cLDA), give accurate deformation potentials for group III nitrides. Cross-checks show that the deformation potentials of GaN obtained with HSE are in good agreement with G_0W_0@OEPx(cLDA) calculations (Table 8.3), which validates the reliability of the HSE approach in studying strain effects in nitride materials. It has been found that the transition energies depend nonlinearly on the lattice parameters, hence the equilibrium lattice parameters around which the linear expansion is constructed significantly affect the deformation potentials. The set of deformation potentials presented in Table 8.3 was therefore obtained by constraining the strain range to realistic conditions in the linear regime around the experimental lattice parameters.

A complete set of deformation potentials constitutes essential input for the modeling of nitride optoelectronic devices. It allows the prediction of subband structures under realistic strain conditions especially for light-emitting devices grown along nonpolar or semipolar crystallographic directions, for which the strain is more complicated compared to c-plane devices and crucial for the device performance.

8.3 Defects in Nitride Materials

Point defects and impurity atoms can strongly affect the properties of materials and the performance of optoelectronic devices. The intentional addition of donor and acceptor atoms to semiconductors yields *n*- and *p*-type doped materials, respectively, that are necessary for the fabrication of diodes. However, the free carrier density that can be achieved by doping is limited by compensating defects that spontaneously form at high doping concentrations. Moreover, certain materials exhibit unintentional doping due to either intrinsic defects or unintentional incorporation of foreign impurity atoms. In addition, defects mediate nonradiative recombination and loss in optoelectronic devices and cause device degradation. These examples highlight the need to study defects in materials and understand their effect on device performance.

8.3.1 Methodology

First-principles methods based on DFT have already been successful in the study of defects in semiconductors and insulators [53], but the band-gap error limits the predictive power of the commonly used functionals. The severe underestimation of band gaps by LDA and GGA may introduce large errors in calculated defect formation energies (and, thus, defect concentrations), transition levels, and optical and emission energies associated with defects or impurities. A predictive theory to describe defects must therefore incorporate an accurate treatment of excited-state materials properties.

Recent developments in hybrid functionals have led to great improvements over the traditional DFT approaches for the study of defects [54]. As discussed in Section 8.2.2, hybrid functionals such as HSE [41, 42] significantly improve the description of band gaps, and therefore, aid in the investigation of basic electronic properties of host semiconductors and the electronic and structural properties of defects and impurities. The main advantage of hybrid functionals over other excited-state methods, such as the *GW* approach, is that hybrid functionals also provide accurate total energies and atomic forces. These quantities are essential for calculating formation energies and lattice relaxations associated with defects, impurities, interfaces, and surfaces.

The HSE functional describes the band gaps with great accuracy for many semiconductor materials [55], but it may still lead to band gaps noticeably smaller than experimental values for some wide band-gap semiconductors and insulators if used with the standard mixing parameter of 25% (GGA exchange mixed with Hartree–Fock in a 3 : 1 ratio). In these cases, band-gap values can be brought into closer agreement with experiment by increasing the mixing parameter beyond the standard 25% value [56–58]. It has been shown that the mixing parameter that reproduces the experimental band-gap value also predicts defect levels that quantitatively agree with experimental observations [58, 59].

Isolated defects and impurities are typically calculated using the supercell approach, in which the defect is created in a section of the material that is periodically repeated in three spatial directions. Supercell calculations containing up to thousands of atoms are nowadays achievable within LDA and GGA, but present computational resources limit HSE calculations to a few hundred atoms. These methods can compute total energies and forces for a solid containing an isolated defect or impurity, and can fully take into account atomic relaxations around the lattice imperfection.

The formation energy is the key quantity that determines the stability and concentration of defects in the dilute limit, that is, when defect–defect interactions are negligible. The formation energy of a defect D in charge state q is defined as a total energy difference and is given by [53]

$$E^f(D^q) = E_{\text{tot}}(D^q) - E_{\text{tot}}(\text{host}) + \sum n_X \mu_X + q\varepsilon_F,$$

where $E_{\text{tot}}(D^q)$ and $E_{\text{tot}}(\text{host})$ are total energies calculated for supercells containing the defect in charge state q and the perfect host crystal, respectively, n_X is the number of atoms of species X that were removed from the crystal to form the defect, and μ_X is the chemical potential of the species X reservoir. In the case of a nitrogen vacancy in GaN, for example, X is nitrogen and $n_X = +1$. The chemical potential for species X is conventionally referenced to the energy per atom of the standard phase of the element. It is bound by the stability condition of the host and the requirement of avoiding the formation of secondary phases. In the case of GaN, μ_N is referenced to the total energy of an isolated N_2 molecule, and must satisfy the following condition:

$$\mu_{\text{N}} + \mu_{\text{Ga}} = \Delta H^f(\text{GaN}),$$

where $\Delta H^f(\text{GaN})$ is the formation enthalpy of GaN [53]. Charged defect calculations, for which $q \neq 0$, involve the exchange of one or more electrons with an electron reservoir. The chemical potential of the reservoir is the Fermi level ε_F, which is conventionally referenced to the valence band maximum of the host. In practical calculations, the valence band maximum is referenced to the averaged electrostatic potential, rendering it necessary to align the average potentials in the host and the defect-containing supercell. Additional corrections are necessary to correct for finite-size effects in charged defect supercells. Calculations that employ periodic boundary conditions represent the charged defect as an infinite array of charges in a uniform neutralizing background. The interactions between defects in neighboring supercells can lead to errors on the order of 1 eV in the formation energy of defects, especially for highly charged defects ($|q| > 2$). Several approaches have been developed to correct for interactions between defects in neighboring supercells [60,61]. The most straightforward, but computationally expensive, is to perform calculations for progressively increasing supercell sizes and extrapolate the results to infinite limit. A more rigorous approach for performing charge-state corrections, based on an exact treatment of the electrostatic problem, has recently been proposed [62], circumventing the need to perform calculations for various supercell sizes.

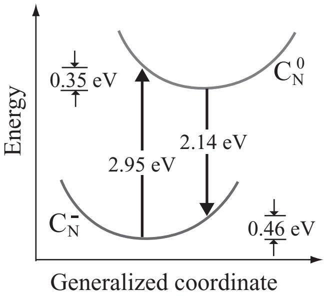


Figure 8.3 *Calculated configuration–coordinate diagram for the C_N impurity in GaN using the HSE hybrid functional [59]. The optical absorption by the negatively charged defect has a zero phonon line at 2.60 eV and a peak at 2.95 eV. The emission process ($C_N^0 + e \rightarrow C_N^- + \hbar\omega$) has a zero-phonon line at 2.60 eV and a peak at 2.14 eV, in the yellow part of the visible spectrum.*

The electrical activity of defects, which determines for instance whether a defect is a deep or shallow donor or acceptor, is defined by the position of transition levels relative to the band edges. The thermodynamic transition levels of a defect can be derived from formation energies [53]:

$$\varepsilon(q/q') = \frac{E^f(D^q) - E^f(D^{q'})}{q' - q}.$$

For Fermi-level values below the transition level $\varepsilon(q/q')$ the defect is stable in charge state q, while for values above the transition level it is stable in charge state q'. The difference between thermodynamic and optical transition levels of defects is that the former involve local lattice relaxations while the latter do not. The optical transition levels are associated with optical absorption and luminescence. Since optical transitions occur instantaneously compared to the time scale associated with lattice relaxations, they can be represented with vertical lines in a configuration–coordinate diagram (Figure 8.3). The optical transition energy between two charge states is the difference in formation energies of the corresponding charged defects calculated at the initial-state lattice geometry. It is also necessary to specify whether during the optical transition the defect exchanges electrons with the valence or with the conduction band. If the transition involves placing an electron in the conduction band, the Fermi level ε_F is taken at the conduction band minimum, while if a hole is lifted from the valence band to the defect level, then the Fermi level is placed at the top of the valence band.

8.3.2 Example: C in GaN

The carbon impurity in GaN presents an example of application of first-principles techniques to study defects and impurities in the nitride materials. The role of

C in GaN has been put under scrutiny since the early explorations of GaN for LEDs. Ion implantation with various elements demonstrated that C was associated with yellow luminescence in GaN [63]. However, the microscopic structure of the C impurity—whether it was substituting for a Ga or a N atom or sitting on an interstitial site, or possibly associated with a point defect—remained inconclusive [64, 65]. Based on atomic size and valence considerations, it was thought that C substituting on the N site (C_N) would act as a shallow acceptor in GaN [66, 67], yet no experimental evidence of *p*-type GaN using C as an acceptor exists. Intentional C doping actually leads to semi-insulating GaN layers, which are used in AlGaN/GaN heterojunction field-effect transistors [68]. The semi-insulating behavior was tentatively explained by the incorporation of the C_N acceptor and the C_{Ga} donor in equal concentrations [68, 69].

First-principles calculations employing hybrid functionals provide an accurate description of C impurities in GaN and can explain all the experimental observations, including the source of the yellow luminescence. Previous DFT-LDA/GGA calculations found C_N to be a shallow acceptor with a transition level in the range of 200–300 meV [69–72]. The band-gap error in LDA/GGA calculations introduces large errors in the calculated transition levels, precluding definite conclusions on the exact position of the C acceptor level with respect to the valence band. In addition, hole localization tends to be underestimated in LDA or GGA, a problem that is also alleviated by the use of a hybrid functional [73]. Calculations employing the HSE hybrid functional [59] show that C is likely to incorporate on the N site and gives rise to an acceptor $\varepsilon(0/-)$ level located at 0.90 eV above the valence band maximum (Figure 8.4). C substituting on the Ga site (C_{Ga}) acts as a shallow donor, in accord with previous results [72, 74]. These results indicate that C_N itself can explain the semi-insulating behavior of C-doped GaN without the necessity of invoking equal incorporation of C_N acceptors and C_{Ga} donors.

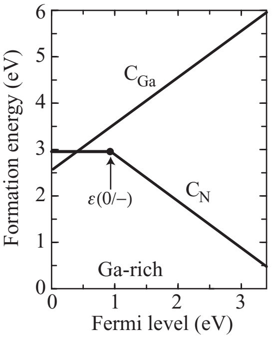


Figure 8.4 *Calculated formation energy versus Fermi level for the C impurity in GaN under Ga-rich conditions using the HSE hybrid functional [59]. The transition level $\varepsilon(0/-)$ for the C_N acceptor is indicated.*

The HSE-calculated optical transition levels of the C_N acceptor in GaN are represented in the configuration–coordinate diagram in Figure 8.3. An absorption peak at 2.95 eV is predicted for the optical transition from the negatively charged acceptor to the conduction band ($C_N^- + \hbar\omega \rightarrow C_N^0 + e$). The inverse process ($C_N^0 + e \rightarrow C_N^- + \hbar\omega$) exhibits an emission peak at 2.14 eV, which is in the yellow part of the visible spectrum. The difference between absorption and emission energies originates from the large differences in local lattice relaxations around C_N in the negative and neutral charge states. These results indicate that C_N can explain both the semi-insulating behavior of C-doped GaN and the yellow luminescence.

8.4 Auger Recombination and Efficiency Droop Problem of Nitride LEDs

8.4.1 Efficiency Droop

The peak efficiencies of nitride LEDs far exceed those of incandescent and fluorescent light bulbs. However, the remarkable efficiency of LEDs decreases dramatically when they are operating at the high power that general lighting requires (Figure 8.5). This phenomenon is called *efficiency droop* and is a universal feature of nitride LEDs [75]. Moreover, droop becomes progressively worse for devices operating at longer wavelengths, an effect known as the *green gap problem*. This efficiency loss limits the light output power and increases the operating cost of LEDs. To compensate for the light output reduction, more individual LEDs per light bulb need to be employed, which increases the total cost. The efficiency droop problem prevents LED light bulbs from taking over the lighting market. Hence, there is a strong incentive to understand its origin and devise engineering solutions to mitigate its impact.

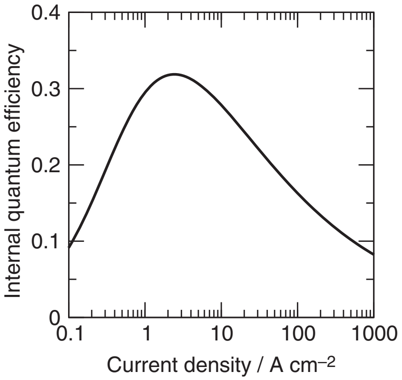
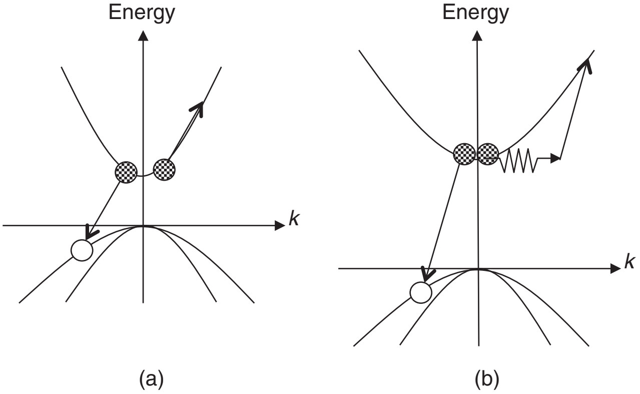


Figure 8.5 *Typical internal quantum efficiency versus injected current density for a nitride LED. The curve is generated from representative experimental data [76]. The efficiency exhibits a peak at low current but decreases dramatically at the high power needed for lighting.*

Several loss mechanisms have been proposed as the origin of the droop, but the topic is still a subject of active research and intense debate. The most prominent hypotheses for the loss include electron leakage from the quantum wells [77], defect-related mechanisms [78, 79], or Auger recombination [80–83]. It is difficult to distinguish the contribution by each source of loss experimentally and to identify the precise microscopic mechanism responsible for the droop. Therefore, insight gained from first-principles calculations can provide valuable input on the importance of the various loss mechanisms and their effect on the efficiency of nitride LEDs.

8.4.2 Auger Recombination

Auger recombination is one of the more plausible theories that have been put forward as the cause of the efficiency droop. In the Auger process, an electron and a hole recombine in a nonradiative manner, while the energy released by the recombination is transferred via electron–electron scattering to a third carrier that gets excited to a higher conduction band or lower valence band state (Figure 8.6). The third carrier involved in the transition can either be an electron (electron–electron–hole process) or a hole (hole–hole–electron). Since three carriers are involved in each Auger recombination event, the overall recombination rate scales with the third power of the free carrier density, $R_{\text{Auger}} = Cn^3$, and becomes dominant at high injected current densities. On the other hand, the radiative recombination rate increases with the square of the free carrier density, $R_{\text{rad}} = Bn^2$, while Shockley–Read–Hall nonradiative recombination at defects depends linearly on



(a) (b)

Figure 8.6 Schematic illustration of Auger recombination of free carriers in semiconductors. (a) During the direct Auger process, an electron and a hole recombine nonradiatively while transferring the excess energy to a third carrier. (b) Indirect Auger recombination is a higher order process that is assisted by a scattering mechanism, which provides additional momentum and facilitates the energy transfer. Indirect Auger recombination is the dominant nonradiative recombination mechanism in nitride LEDs at high power.

the carrier concentration, $R_{\mathrm{SRH}} = An$. As a result, the internal quantum efficiency η_{IQE}, defined as the ratio of emitted photons over total recombined electron–hole pairs,

$$\eta_{\mathrm{IQE}} = \frac{R_{\mathrm{rad}}}{R_{\mathrm{SRH}} + R_{\mathrm{rad}} + R_{\mathrm{Auger}}} = \frac{Bn^2}{An + Bn^2 + Cn^3}, \tag{8.3}$$

decreases at high injected carrier densities and explains the efficiency reduction in Figure 8.5. The Auger mechanism was first proposed after photoluminescence experiments showed that efficiency droop behavior described by equation (8.3) occurs even for optically excited samples [80], excluding any carrier leakage effects. The efficiency droop was subsequently observed for bulk GaN samples [84], indicating that the mechanism for the droop is intrinsic to nitride materials and not a consequence of the heterostructure. Experimentally measured values for the Auger coefficient of InGaN are on the order of 10^{-31}–10^{-30} cm^6 s^{-1} and are large enough to explain the observed efficiency droop in LEDs [75].

However, the significance of Auger recombination has been put into question due to an apparent discrepancy between experiment and calculations. Theoretical investigations that focused on the direct Auger recombination process (Figure 8.6a) found an Auger coefficient of the order of 10^{-34} cm^6 s^{-1} [85, 86]. This value is three to four orders of magnitude smaller than the experimental results and cannot account for the measurements. This is expected because the Auger coefficient of a direct band-gap semiconductor decays exponentially with increasing band gap and will therefore be small for wide band-gap materials such as the nitrides. Moreover, the direct Auger recombination rate is very sensitive to temperature, a fact that is at variance with experimental measurements [81]. These observations rule out direct Auger recombination as the cause of the droop.

Higher order indirect Auger processes are also possible, however, and can potentially be of importance in nitride materials. Indirect Auger recombination requires the assistance of a microscopic carrier-scattering mechanism to provide additional momentum and facilitate the Auger transition (Figure 8.6b). As a result, indirect processes can become the dominant Auger recombination mechanism in wide band-gap materials and can be strong enough to explain the efficiency droop. It is therefore important to investigate these indirect Auger processes theoretically and to evaluate their effect on LED performance.

Several carrier-scattering mechanisms can potentially be at work in nitride materials. One mechanism is the interaction of electrons and holes with lattice vibrations. Electron–phonon coupling is a universal carrier-scattering mechanism in every material and responsible for several well-known electronic phenomena such as resistivity, superconductivity, and indirect optical absorption. This coupling is particularly strong in the nitrides because of the ionic nature of the material as well as the nature of the nitrogen valence orbitals. Since the $2p$ orbitals are the innermost p orbitals of an atom, they feel the core potential almost unscreened and are thus very sensitive to atomic vibrations [87]. Another scattering mechanism is alloy scattering, due to the disordered distribution of In and Ga atoms in the InGaN quantum wells where carriers recombine in devices. Finally, the long-range Coulomb potential of charged defects can scatter charge carriers and enable Auger recombination.

8.4.3 Computational Methodology

First-principles calculations are a powerful set of computational tools that can elucidate the impact of indirect Auger recombination on the performance of nitride LEDs. These computational techniques allow explicit investigations of microscopic phenomena such as electron–phonon coupling and alloy scattering and their effect on the Auger recombination rate. Moreover, Auger recombination in these wide band-gap materials generates hot carriers at excited states that are very high in energy and cannot be described with near-band-edge models such as the $k \cdot p$ theory. First-principles methods, in contrast, can accurately provide the full band structure throughout the Brillouin zone, and can readily describe these high-energy states of interest. First-principles approaches are therefore a suitable and predictive computational formalism for the description of indirect Auger recombination in the nitride materials.

The transition probability rate for Auger recombination is calculated from Fermi's golden rule [88]:

$$R_{\text{Auger}} = 2\frac{2\pi}{\hbar} \sum_{1234} P \, |M_{1234}|^2 \, \delta(\varepsilon_1 + \varepsilon_2 - \varepsilon_3 - \varepsilon_4),$$

where we used composite band and wave-vector indices $i = (n_i,\mathrm{k}_i)$, and P is a statistics factor that ensures transitions are only allowed from occupied to empty states,

$$P = f_1 f_2 (1 - f_3)(1 - f_4),$$

where f_1 is the Fermi–Dirac occupation number of state i. The perturbation Hamiltonian is the screened Coulomb interaction between charge carriers, and the matrix elements involve a direct M^d and an exchange M^x term (Figures 8.7a and b) to

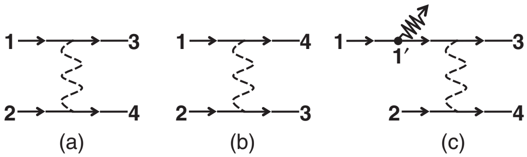


Figure 8.7 Electron–electron scattering diagrams for the (a) direct and (b) exchange terms of direct Auger recombination. (c) One of the 16 diagrams involved in the phonon-assisted Auger process.

account for fermion exchange symmetry:

$$|M_{1234}|^2 = \left|M^d_{1234} - M^x_{1234}\right|^2 + \left|M^d_{1234}\right|^2 + \left|M^x_{1234}\right|^2,$$

where $M^d_{1234} = M^x_{1234} = \iint \psi_1^*(r_1)\psi_2^*(r_2)W(r_1, r_2)\psi_3(r_1)\psi_4(r_2)\mathrm{d}r_1\mathrm{d}r_2$ are matrix elements of the screened Coulomb interaction W. The band energies ε_i and wavefunctions $\psi_i(r)$ are readily calculated with DFT. The dielectric function and its wave-vector dependence needed for W can be described with a model [89] or, alternatively, can be calculated entirely from first principles within the RPA [20].

The transition rate equations for indirect Auger recombination are slightly modified to account for the additional phonon emission or absorption process. In this case the probability rate is given by second-order Fermi's golden rule:

$$R^{\text{Indirect}}_{\text{Auger}} = 2\frac{2\pi}{\hbar}\sum_{1234;\text{vq}}\left(n_{\text{vq}} + \frac{1}{2} \pm \frac{1}{2}\right)P\left|\tilde{M}_{1234;\text{vq}}\right|^2 \delta(\varepsilon_1 + \varepsilon_2 - \varepsilon_3 - \varepsilon_4 \mp \hbar\omega_{\text{vq}}),$$

where n_{vq} are the phonon Bose occupation numbers, $\hbar\omega_{\text{vq}}$ are the phonon frequencies, and the upper or lower sign corresponds to phonon emission or absorption, respectively. The perturbation matrix elements $\tilde{M}_{1234;\text{vq}}$ are also slightly modified to account for the electron–phonon coupling interaction. The overall rate involves 16 terms to account for all possible orderings of electron–electron scattering and electron–phonon coupling diagrams. For example, the matrix element that corresponds to the diagram of Figure 8.7c is given by

$$\tilde{M}_{1234;v\text{q}} = \sum_{1'} \frac{g_{11';v} M^d_{1'234}}{\varepsilon_{1'} - \varepsilon_1 \pm \hbar\omega_{v\text{q}}},$$

where $g_{11';\text{v}}$ is the electron–phonon coupling matrix element from state 1 to state $1'$ by phonon mode v.

8.4.4 Results

First-principles calculations have been applied to determine the alloy-scattering-assisted Auger recombination rate in InGaN alloys [90]. A common method to study alloys is to employ the virtual crystal approximation, which models the alloy potential as an interpolated average between the two pure phases. This approximation, however, ignores the effect of atomistic disorder that gives rise to alloy scattering and enables Auger recombination. In order to capture these effects, calculations need to explicitly model the composition fluctuations and randomness of the alloy on the atomic scale. This can be accomplished with quasi-random alloy crystal structures, which are periodic alloy supercells of the desired composition with an atomic distribution that has been optimized to reproduce the short-range correlation function of the fully random alloy [91]. The calculations were done for the 32-atom structure of 25% composition and interpolated for other alloy configurations by considering the dependence of the alloy-scattering potential on

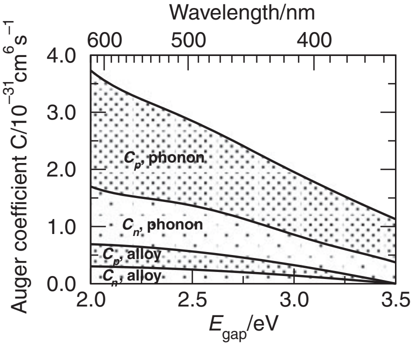


Figure 8.8 *Indirect Auger recombination coefficients for the electron–electron–hole (C_n) and hole–hole–electron (C_p) processes due to electron–phonon coupling and alloy scattering as a function of the InGaN alloy band gap.*

composition, $V^2 \propto x(1-x)$ [88]. Alloy scattering was found to have a significant effect on Auger recombination. The calculated Auger coefficients shown in Figure 8.8 are significantly enhanced compared to the direct Auger process and increase for increasing In mole fraction.

Phonon-assisted Auger recombination has also been studied with first-principles methods and can account for the efficiency droop of nitride LEDs [90]. The vibrational parameters, such as the phonon frequencies and electron–phonon coupling matrix elements, can be calculated with density functional perturbation theory for arbitrary phonon wave vectors in the first Brillouin zone [92]. The electron wavefunctions and energies were calculated for GaN with DFT and the LDA functional. The LDA energy eigenvalues were corrected with a scissors shift that was adjusted to account for varying In mole fractions and to model various alloy compositions. The calculated phonon-assisted Auger coefficients are shown in Figure 8.8. The values for the phonon-assisted case are also strongly enhanced compared to direct Auger recombination and can explain the experimentally measured values from droop experiments. Moreover, the corresponding Auger coefficients increase for higher In content and for increasing temperature, since more phonons are present at elevated temperatures. Phonon-assisted Auger is also possible even at absolute zero temperature because of the phonon-emission processes.

It is informative to examine which vibrational modes are the ones that primarily contribute to phonon-assisted Auger recombination. Group III nitrides are ionic materials, for which the longitudinal-optical (LO) modes are known to be strongly coupled to charge carriers [88]. However, an analysis of the contribution of the various modes to the Auger coefficient reveals that they are not the dominant ones. Significant contributions arise from the acoustic as well as the other optical phonon modes. This is to be expected because phonon-assisted Auger recombination for these wide band-gap materials excites charge carriers to higher electronic states, which happen to be near the edges of the Brillouin zone. These transitions involve scattering by short-range phonons. Although the LO phonons are indeed strongly coupled to electrons, they are so only for long-range scattering, while for short-range scattering all the phonon modes are important. Therefore, a first-principles analysis that accounts for short-range phonon effects is needed to properly describe phonon-assisted Auger for the nitrides.

Scattering by charged defects can also enable indirect Auger recombination, but the resulting values (assuming realistic defect concentrations) are much weaker than those due to electron–phonon coupling or alloy scattering [90]. The interaction of carriers with charged defects can be modeled with a screened Coulomb potential that accounts for dielectric screening by the bulk material as well as metallic screening by the free carriers [88]. However, the calculated charged defect-assisted Auger coefficient values are weak and become significant only for high defect concentrations, of the order of 10^{20} cm^{-3}, which are not encountered in actual devices. This kind of indirect Auger process can therefore be ruled out as a significant source of energy loss in LEDs.

The cumulative Auger recombination coefficient is plotted in Figure 8.8. The overall value ranges from 1.5×10^{-31} $\mathrm{cm}^6\,\mathrm{s}^{-1}$ for violet devices to 3.0×10^{-31} $\mathrm{cm}^6\,\mathrm{s}^{-1}$ for green and is consistent with the experimentally measured range of values. This first-principles result provides evidence that indirect Auger recombination assisted by alloy disorder and electron–phonon scattering is the primary LED efficiency loss mechanism at high power. This finding also suggests strategies to mitigate the impact of Auger recombination on devices. Since Auger is an intrinsic loss mechanism, it is hard to eliminate entirely. However, the device efficiency can improve if the carrier density in the quantum wells is kept as low as possible. This can be achieved by either growing thicker quantum wells or using a larger number of them and spreading the charge carriers over a larger volume. Another approach is to employ nonpolar or semipolar growth directions and eliminate or reduce the polarization fields that keep electrons and holes apart inside the quantum wells. This will enhance the overall recombination rate and reduce the steady-state carrier density in the quantum wells.

8.5 Summary

We discussed recent developments in first-principles computational methods for the study of nitride materials employed for solid-state lighting. We also presented examples that showed the wide range of applications of first-principles calculations in this field, ranging from the basic structural and electronic properties of the nitride materials to the effects of strain, defects, and nonradiative recombination on the optoelectronic device performance. First-principles methods are, therefore, a powerful explanatory and predictive computational tool that can assist and guide the experimental development of efficient solid-state optoelectronic devices and can help reduce the impact of general lighting on the world's energy resources.

Acknowledgments

We acknowledge fruitful discussions with K. Delaney, M. Scheffler, P.G. Moses, J. Speck, and C. Weisbuch. EK was supported as part of the Center for Energy Efficient Materials, an Energy Frontier Research Center funded by the U.S. DOE, BES under Award No. DE-SC0001009. PR acknowledges the support of the Deutsche Forschungsgemeinschaft, the UCSB-MPG Exchange Program and the NSF-IMI Program (Grant No. DMR04-09848). Additional support was provided by the NSF (Award No. DMR-0906805) and by the UCSB Solid State Lighting and Energy Center. Computational resources were provided by the CNSI Computing Facility under NSF Grant No. CHE-0321368, the DOE NERSC facility, and Teragrid.

References

1. Humphreys, C.J. (2008) Solid-state lighting, *MRS Bull.* **33**, 459–470.
2. Pimputkar, S., Speck, J.S., DenBaars, S.P. and Nakamura, S. (2009) Prospects for LED lighting, *Nat. Photonics* **3**, 180–182.
3. Schubert, E.F. (2006) *Light-Emitting Diodes*, Cambridge University Press, New York.
4. Nakamura, S., Senoh, M. and Mukai, T. (1993) P-GaN/n-InGaN/n-GaN double-heterostructure blue-light-emitting diodes, *Japan. J. Appl. Phys.* **32**, L8–L11.
5. (2007) *Nitride Semiconductor Devices: Principles and Simulation* (ed. J. Piprek), Wiley-VCH Verlag GmbH, Berlin.
6. Speck, J.S. and Chichibu, S.F. (2009) Nonpolar and semipolar group III materials, *MRS Bull.* **34**, 304–309.
7. Guo, W., Banerjee, A., Bhattacharya, P. and Ooi, B.S. (2011) InGaN/GaN disk-in-nanowire white light emitting diodes on (001) silicon, *Appl. Phys. Lett.* **98**, 193102.
8. Zhang, M., Bhattacharya, P. and Guo, W. (2010) InGaN/GaN self-organized quantum dot green light emitting diodes with reduced efficiency droop, *Appl. Phys. Lett.* **97**, 011103.
9. Martin, R.M. (2004) *Electronic Structure, Basic Theory and Practical Methods*, Cambridge University Press, Cambridge.
10. Ceperley, D.M. and Alder, B. (1980) Ground state of the electron gas by a stochastic method, *Phys. Rev. Lett.* **45**, 566–569.
11. Perdew, J., Burke, K. and Ernzerhof, M. (1996) Generalized gradient approximation made simple, *Phys. Rev. Lett.* **77**, 3865–3868.
12. Harl, J. and Kresse, G. (2009) Accurate bulk properties from approximate many-body techniques, *Phys. Rev. Lett.* **103**, 056401.
13. Ren, X., Tkatchenko, A., Rinke, P. and Scheffler, M. (2011) Beyond the random-phase approximation for the electron correlation energy: the importance of single excitations, *Phys. Rev. Lett.* **106**, 153003.
14. Stampfl, C. and Van de Walle, C.G. (1999) Density-functional calculations for III-V nitrides using the local-density approximation and generalized gradient approximation, *Phys. Rev. B* **59**, 5521–5535.

15. Yan, Q., Rinke, P., Winkelnkemper, M. *et al.* (2011) Band parameters and strain effects in ZnO and group-III nitrides, *Semicond. Sci. Technol.* **26**, 014037.
16. Kümmel, S. and Kronik, L. (2008) Orbital-dependent density functionals: theory and applications, *Rev. Mod. Phys.* **80**, 3–60.
17. Perdew, J. and Levy, M. (1983) Physical content of the exact Kohn-Sham orbital energies: band gaps and derivative discontinuities, *Phys. Rev. Lett.* **51**, 1884–1887.
18. Sham, L. and Schluüter, M. (1983) Density-functional theory of the energy gap, *Phys. Rev. Lett.* **51**, 1888–1891.
19. Hedin, L. and Lundqvist, S. (1970) Effects of electron-electron and electron-phonon interactions on the one-electron states of solids, *Solid State Phys.* **23**, 1–181.
20. Hybertsen, M.S. and Louie, S.G. (1986) Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies, *Phys. Rev. B* **34**, 5390–5413.
21. Aulbur, W.G., Jönsson, L. and Wilkins, J.W. (2000) Quasiparticle calculations in solids, *Solid State Phys.* **54**, 1–218.
22. Onida, G., Reining, L. and Rubio, A. (2002) Electronic excitations: density-functional versus many-body Green's-function approaches, *Rev. Mod. Phys.* **74**, 601–659.
23. Louie, S.G. (2006) Predicting materials and properties: theory of the ground and excited state, in *Conceptual Foundations of Materials: A Standard Model for Ground- and Excited-State Properties* (eds S.G. Louie and M.L. Cohen), Elsevier, Amsterdam.
24. Rinke, P., Qteish, A., Neugebauer, J. and Scheffler, M. (2008) Exciting prospects for solids: Exact-exchange based functionals meet quasiparticle energy calculations, *Phys. Status Solidi B* **245**, 929–945.
25. Rinke, P., Qteish, A., Neugebauer, J. *et al.* (2005) Combining GW calculations with exact-exchange density-functional theory: an analysis of valence-band photoemission for compound semiconductors, *New J. Phys.* **7**, 126.
26. Giantomassi, M., Stankovski, M., Shaltaf, R. *et al.* (2011) Electronic properties of interfaces and defects from many-body perturbation theory: recent developments and applications, *Phys. Status Solidi B* **248**, 275–289.
27. Bruneval, F., Vast, N. and Reining, L. (2006) Effect of self-consistency on quasiparticles in solids, *Phys. Rev. B* **74**, 045102.
28. Fuchs, F., Furthmüller, J., Bechstedt, F. *et al.* (2007) Quasiparticle band structure based on a generalized Kohn-Sham scheme, *Phys. Rev. B* **76**, 115109.
29. Jiang, H., Gomez-Abal, R., Rinke, P. and Scheffler, M. (2009) Localized and itinerant states in lanthanide oxides united by GW@LDA + U, *Phys. Rev. Lett.* **102**, 126403.
30. Bechstedt, F., Fuchs, F. and Kresse, G. (2009) Ab-initio theory of semiconductor band structures: new developments and progress, *Phys. Status Solidi B* **246**, 1877–1892.
31. Janotti, A. and Van de Walle, C.G. (2007) Absolute deformation potentials and band alignment of wurtzite ZnO, MgO, and CdO, *Phys. Rev. B* **75**, 121201.
32. Rinke, P., Winkelnkemper, M., Qteish, A. *et al.* (2008) Consistent set of band parameters for the group-III nitrides AlN, GaN, and InN, *Phys. Rev. B* **77**, 075202.
33. Städele, M., Majewski, J.A., Vogl, P. and Görling, A. (1997) Exact Kohn-Sham exchange potential in semiconductors, *Phys. Rev. Lett.* **79**, 2089.
34. Qteish, A., Al Sharif, A.I., Fuchs, M. *et al.* (2005) Role of semicore states in the electronic structure of group-III nitrides: An exact-exchange study, *Phys. Rev. B* **72**, 155317.

35. Rinke, P., Qteish, A., Winkelnkemper, M. *et al.* (2006) Band gap and band parameters of InN and GaN from quasiparticle energy calculations based on exact-exchange density-functional theory, *Appl. Phys. Lett.* **89**, 161919.
36. Qteish, A., Rinke, P., Neugebauer, J. and Scheffler, M. (2006) Exact-exchange-based quasiparticle energy calculations for the band gap, effective masses, and deformation potentials of ScN, *Phys. Rev. B* **74**, 245208.
37. Tiago, M.L., Ismail-Beigi, S. and Louie, S.G. (2004) Effect of semicore orbitals on the electronic band gaps of Si, Ge, and GaAs within the GW approximation, *Phys. Rev. B* **69**, 125212.
38. Rohlfing, M., Krüger, P. and Pollmann, J. (1995) Quasiparticle band structure of CdS, *Phys. Rev. Lett.* **75**, 3489–3492.
39. Kane, E.O. (1982) *Band Theory and Transport Properties, Handbook on Semiconductors*, vol. 1 (ed. W. Paul), North-Holland, Amsterdam.
40. Becke, A.D. (1993) A new mixing of Hartree-Fock and local density-functional theories, *J. Chem. Phys.* **98**, 1372–1377.
41. Heyd, J., Scuseria, G.E. and Ernzerhof, M. (2003) Hybrid functionals based on a screened Coulomb potential, *J. Chem. Phys.* **118**, 8207–8215.
42. Heyd, J., Scuseria, G.E. and Ernzerhof, M. (2006) Erratum: 'Hybrid functionals based on a screened Coulomb potential' [*J. Chem. Phys.* **118**, 8207 (2003)], *J. Chem. Phys.* **124**, 219906.
43. Moses, P.G. and Van de Walle, C.G. (2010) Band bowing and band alignment in InGaN alloys, *Appl. Phys. Lett.* **96**, 021908.
44. Van de Walle, C.G. and Martin, R.M. (1989) 'Absolute' deformation potentials: Formulation and ab initio calculations for semiconductors, *Phys. Rev. Lett.* **62**, 2028.
45. Romanov, A.E., Baker, T.J., Nakamura, S. and Speck, J.S. (2006) Strain-induced polarization in wurtzite III-nitride semipolar layers, *J. Appl. Phys.* **100**, 023522.
46. Shan, W., Hauenstein, R.J., Fischer, A.J. *et al.* (1996) Strain effects on excitonic transitions in GaN: deformation potentials, *Phys. Rev. B* **54**, 13460.
47. Ueda, M., Funato, M., Kojima, K. *et al.* (2008) Polarization switching phenomena in semipolar $In_xGa_{1-x}N$/GaN quantum well active layers, *Phys. Rev. B* **78**, 233303.
48. Bir, G.L. and Pikus, G.E. (1974) *Symmetry and Strain-Induced Effects in Semiconductors*, John Wiley & Sons, Inc., New York.
49. Gil, B., Moret, M., Briot, O. *et al.* (2009) InN excitonic deformation potentials determined experimentally, *J. Cryst. Growth* **311**, 2798.
50. Ishii, R., Kaneta, A., Funato, M. *et al.* (2010) All deformation potentials in GaN determined by reflectance spectroscopy under uniaxial stress: Definite breakdown of the quasicubic approximation, *Phys. Rev. B* **81**, 155202.
51. Yan, Q., Rinke, P., Scheffler, M. and Van de Walle, C.G. (2009) Strain effects in group-III nitrides: Deformation potentials for AlN, GaN, and InN, *Appl. Phys. Lett.* **95**, 121111.
52. Yan, Q., Rinke, P., Scheffler, M. and Van de Walle, C.G. (2010) Role of strain in polarization switching in semipolar InGaN/GaN quantum wells, *Appl. Phys. Lett.* **97**, 181102.
53. Van de Walle, C.G. and Neugebauer, J. (2004) First-principles calculations for defects and impurities: Applications to III-nitrides, *J. Appl. Phys.* **95**, 3851.

54. Van de Walle, C.G. and Janotti, A. (2011) Advances in electronic structure methods for defects and impurities in solids, *Phys. Status Solidi B* **248**, 19–27.
55. Marsman, M., Paier, J., Stroppa, A. and Kresse, G. (2008) Hybrid functionals applied to extended systems, *J. Phys. Condens. Matter.* **20**, 064201.
56. Oba, F., Togo, A., Tanaka, I. *et al.* (2008) Defect energetics in ZnO: a hybrid Hartree-Fock density functional study, *Phys. Rev. B* **77**, 245202.
57. Clark, S.J., Robertson, J., Lany, S. and Zunger, A. (2010) Intrinsic defects in ZnO calculated by screened exchange and hybrid density functionals, *Phys. Rev. B* **81**, 115311.
58. Lyons, J.L., Janotti, A. and Van de Walle, C.G. (2009) Why nitrogen cannot lead to *p*-type conductivity in ZnO, *Appl. Phys. Lett.* **95**, 252105.
59. Lyons, J.L., Janotti, A. and Van de Walle, C.G. (2010) Carbon impurities and the yellow luminescence in GaN, *Appl. Phys. Lett.* **97**, 152108.
60. Makov, G. and Payne, M.C. (1995) Periodic boundary conditions in ab initio calculations, *Phys. Rev. B* **51**, 4014.
61. Shim, J., Lee, E.-K., Lee, Y.J. and Nieminen, R.M. (2005) Density-functional calculations of defect formation energies using supercell methods: defects in diamond, *Phys. Rev. B* **71**, 035206.
62. Freysoldt, C., Neugebauer, J. and Van de Walle, C.G. (2009) Fully ab initio finite-size corrections for charged-defect supercell calculations, *Phys. Rev. Lett.* **102**, 016402.
63. Ogino, T. and Aoki, M. (1980) Mechanism of yellow luminescence in GaN, *Japan. J. Appl. Phys.* **19**, 2395.
64. Pankove, J.I. and Moustakas, T.D. (eds) (1998) *Gallium Nitride (GaN) I, Semiconductors and Semimetals*, vol. 50, Academic Press, San Diego, CA.
65. Pankove, J.I. and Moustakas, T.D. (eds) (1999) *Gallium Nitride (GaN) II, Semiconductors and Semimetals*, vol. 57, Academic Press, San Diego, CA.
66. Fischer, S., Wetzel, C., Haller, E.E. and Meyer, B.K. (1995) On p-type doping in GaN-acceptor binding energies, *Appl. Phys. Lett.* **67**, 1298.
67. Reshchikov, M.A. and Morkoç, H. (2005) Luminescence properties of defects in GaN, *J. Appl. Phys.* **97**, 061301.
68. Armstrong, A., Arehart, A.R., Moran, B. *et al.* (2004) Impact of carbon on trap states in n-type GaN grown by metalorganic chemical vapor deposition, *Appl. Phys. Lett.* **84**, 374.
69. Seager, C.H., Wright, A.F., Yu, J. and Götz, W. (2002) Role of carbon in GaN, *J. Appl. Phys.* **92**, 6553.
70. Neugebauer, J. and Van de Walle, C.G. (1996) Gallium vacancies and the yellow luminescence in GaN, *Appl. Phys. Lett.* **69**, 503.
71. Bogusławski, P. and Bernholc, J. (1997) Doping properties of C, Si, and Ge impurities in GaN and AlN, *Phys. Rev. B* **56**, 9496.
72. Ramos, L.E., Furthmüller, J., Leite, J.R. *et al.* (2002) Carbon-based defects in GaN: doping behaviour, *Phys. Status Solidi B* **234**, 864.
73. Lany, S. and Zunger, A. (2009) Polaronic hole localization and multiple hole binding of acceptors in oxide wide-gap semiconductors, *Phys. Rev. B* **80**, 085202.
74. Wright, A.F. (2002) Substitutional and interstitial carbon in wurtzite GaN, *J. Appl. Phys.* **92**, 2575.
75. Piprek, J. (2010) Efficiency droop in nitride-based light-emitting diodes, *Phys. Status Solidi A* **207**, 2217–2225.

76. Laubsch, A., Sabathil, M., Baur, J. *et al.* (2010) High-power and high-efficiency InGaN-based light emitters, *IEEE Trans. Electron Dev.* **57**, 79–97.
77. Kim, M.-H., Schubert, M.F., Dai, Q. *et al.* (2007) Origin of efficiency droop in GaN-based light-emitting diodes, *Appl. Phys. Lett.* **91**, 183507.
78. Monemar, B. and Sernelius, B.E. (2007) Defect related issues in the 'current roll-off' in InGaN based light emitting diodes, *Appl. Phys. Lett.* **91**, 181103.
79. Hader, J., Moloney, J.V. and Koch, S.W. (2010) Density-activated defect recombination as a possible explanation for the efficiency droop in GaN-based diodes, *Appl. Phys. Lett.* **96**, 221106.
80. Shen, Y.C., Mueller, G.O., Watanabe, S. *et al.* (2007) Auger recombination in InGaN measured by photoluminescence. *Appl. Phys. Lett.* **91**, 141101.
81. Laubsch, A., Sabathil, M., Bergbauer, W. *et al.* (2009) On the origin of IQE-'droop' in InGaN LEDs, *Phys. Status Solidi C* **6**, S913–S916.
82. Zhang, M., Bhattacharya, P., Singh, J. and Hinckley, J. (2009) Direct measurement of auger recombination in $In_{0.1}Ga_{0.9}N$/GaN quantum wells and its impact on the efficiency of $In_{0.1}Ga_{0.9}N$/GaN multiple quantum well light emitting diodes, *Appl. Phys. Lett.* **95**, 201108.
83. David, A. and Grundmann, M.J. (2010) Droop in InGaN light-emitting diodes: A differential carrier lifetime analysis, *Appl. Phys. Lett.* **96**, 103504.
84. David, A. and Gardner, N.F. (2010) Droop in III-nitrides: comparison of bulk and injection contributions, *Appl. Phys. Lett.* **97**, 193508.
85. Hader, J., Moloney, J.V., Pasenow, B. *et al.* (2008) On the importance of radiative and Auger losses in GaN-based quantum wells, *Appl. Phys. Lett.* **92**, 261103.
86. Bertazzi, F., Goano, M. and Bellotti, E. (2010) A numerical study of Auger recombination in bulk InGaN, *Appl. Phys. Lett.* **97**, 231118.
87. Cardona, M. (2005) Electron-phonon interaction in tetrahedral semiconductors, *Solid State Commun.* **133**, 3–18.
88. Ridley, B.K. (1999) *Quantum Processes in Semiconductors*, Oxford University Press, Oxford.
89. Cappellini, G., Del Sole, R., Reining, L. and Bechstedt, F. (1993) Model dielectric function for semiconductors, *Phys. Rev. B* **47**, 9892–9895.
90. Kioupakis, E., Rinke, P., Delaney, K.T. and Van de Walle, C.G. (2011) Indirect Auger recombination as a cause of efficiency droop in nitride light-emitting diodes, *Appl. Phys. Lett.* **98**, 161107.
91. Shin, D., Arróyave, R., Liu, Z.-K. and Van de Walle, A. (2006) Thermodynamic properties of binary hcp solution phases from special quasirandom structures, *Phys. Rev. B* **74**, 1–13.
92. Baroni, S., de Gironcoli, S., Dal Corso, A. and Giannozzi, P. (2001) Phonons and related crystal properties from density-functional perturbation theory, *Rev. Mod. Phys.* **73**, 515–562.