# Polarization-induced doping
## and its application to light-emitting diodes

Julien JAMAUX

Master Thesis

Department of Materials Engineering

McGill University

Montreal,Quebec

2014-09

Requirements Statement

Copyright Statement

# ACKNOWLEDGEMENTS

I would like first to express my deepest gratitude to my supervisor, Pr. Kirk H. Bevan, for giving me the opportunity to carry out this research, and more specifically for his patience, his availability, his guidance and his encouragements.

I would also like to give my sincere thanks to the members of our group as well as my office friends, without which these months would have seemed much longer. I would particularly like to thank Bin Ouyang, Asif Iqbal, Daniel Marchand, Sazzad Hossain and Zi Wang for the personal interactions we have had over the last year.

I would like to acknowledge Pr. Ishiang Shih for taking the time to read and correct my thesis in meticulous detail.

Finally, I would like to thank my family and friends for their support during these months as a M.Eng student here at McGill.

# ABSTRACT

Light-Emitting Diodes (LEDs) are expected to present an electrical-to-optical efficiency over 75% within the next two decades, so that they would constitute the main lighting technology. The current efficiencies of LEDs mainly suffer from the poor ionization of dopants in large bandgap semiconductors, such as Gallium Nitride or Aluminum Nitride. We present here another doping technique, based on the modulation of the Polarization of the material, and use time-dependent simulations to watch the system reach its equilibrium state. We will try to determine why some designs do not work, while others do. Finally, we will compare the performance of impurity-based and Polarization-based LEDs, in terms of efficiency and light-emission.

# ABRÉGÉ

Les Diodes Électro-Luminescentes (LEDs) ont vocation à présenter des efficacités de conversion electrique-optique au-dela de 75%, ce qui feraient d'elles le principale technologie. d'éclairage. Les efficacités que les LEDs présentent aujourd'hui souffrent essentiellement de la faible ionisation des dopants dans les semiconducteurs a large bande interdite, comme le Nitrure de Gallium ou d'Aluminium. Nous présentons ici une methode alternative de dopage, basée sur la modulation de la Polarisation du matériau, ainsi que des simulation temporelles pour voir l'évolution du système vers l'équilibre. Nous allons essayer de déterminer pourquoi certains designs ne marchent pas, là où d'autres marchent. Finalement, nous comparerons les performances de deux LEDs, l'une dopée par dopants, l'autre dopée par polarisation.

# TABLE OF CONTENTS

## LIST OF FIGURES

# CHAPTER 1
# Introduction

Since prehistoric times, there have been huge benefits to finding a substitute to sunlight when it was no longer available : light was indeed an essential tool for survival. In more recent times, artificial lighting allowed men to push back the limits of the day, being therefore still efficient during the night, and to introduce light in closed spaces, resulting in a productivity as high indoors as outdoors.

The history of lighting is made of a succession of breakthroughs in efficiency, as displayed in Figure 1–1 [9]. The first source of light was fire, and presented a power conversion of less than 0.025% and a very low illumination capacity. It was only thousands of years later, during the 19th and the beginning of 20th century, that fire was replaced with gas lamps in urban areas. These gas lamps were less than 1% efficient, and were rapidly replaced by the incandescent lamps, derived from Edison's light bulb. This incandescent lighting dominated most of the 20th century, with efficiencies ranging from 1 to 5% and illumination capacities one or two order of magnitude greater than their predecessors. Fluorescent lamps arrived on the lighting market around 1950 with both much greater conversion efficiencies (around 25%) and luminous efficacies, one order of magnitude greater than the incandescent bulbs, but presented on the other hand several ecological challenges, mainly because of the mercury they contain. Finally, the end of the 20th century saw the rise of what is called solid-state lighting, based on light-emitting diodes (LEDs), whose capacities

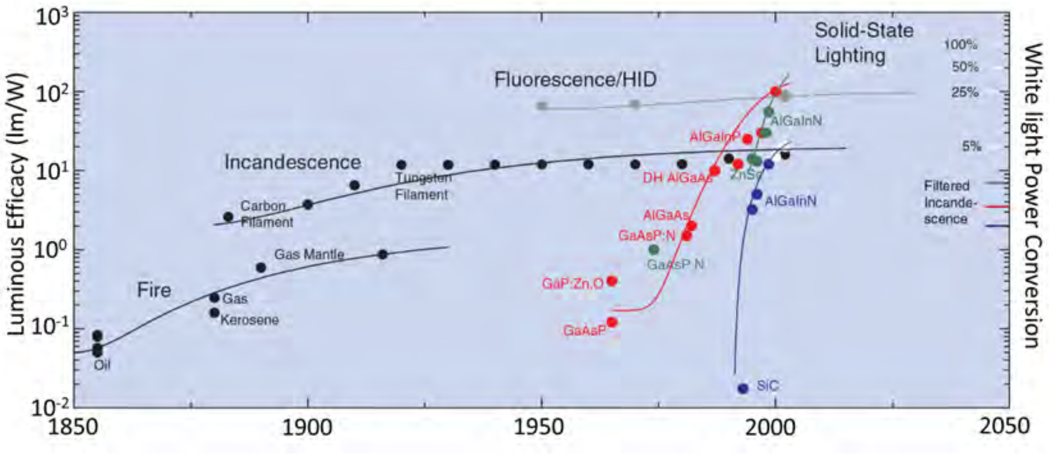


Figure 1–1: Time evolution of luminous efficacy and white light power conversion [1]

improved very quickly over a few decades, to present today the same efficiency as the current fluorescent lamps. However, as research goes, these LEDs are expected to exhibit greater and greater efficiencies [10, 11], and to reach by 2020 an efficiency over 50%, and to go eventually beyond the 75%. If such an efficiency is reached, the quest for artificial lighting will perhaps have come to an end.

We will review at first the basic principles of light emission in light-emitting diodes (LEDs), the expected evolution of the lighting market, and the reasons of this evolution. We will demonstrate that this evolution is only possible via the III-Nitride semiconductors, which are currently facing challenges regarding their efficiency. The technique of polarization-induced doping, which is the topic of the thesis, is an alternative to impurity-based doping, and could constitute a solution to these challenges. We will present the fundamentals of polarization-induced doping, and with the help of our simulations, we will investigate why it does not work in some configurations [12, 13, 14]. Eventually, we will try to compare the performance of impurity-doped and polarization-doped LEDs, to see that polarization-doping does improve the efficiency, but not through the process that we expected.

# CHAPTER 2
# Solid-State Lighting

## 2.1 LEDs and their efficiency

### 2.1.1 A solid-state emitter of light

A Light-Emitting Diode is a solid-state device made of a semiconducting crystal, or a juxtaposition of crystals. The periodic arrangement of the crystal atoms results in a continuum of electronic states, organized in bands, which are separated by gaps: we speak of a band structure [15]. In the case of a conductor, the Fermi level is inside the last band, which means that all the states in the band are not filled: the empty states can accept electrons and therefore allow their circulation. In the case of an insulator, the Fermi level is inside a large bandgap: the band below the Fermi level, called the valence band, is entirely filled and the band above, called the conduction band, is entirely empty; since the bandgap is large, it is very difficult to excite electrons from the valence to the conduction band. The valence electrons cannot consequently move, since all the states are filled, resulting in no motion of carriers. For a semiconductor, the Fermi-level is also inside the bandgap[1] , which is much smaller than that of an insulator: it is therefore possible to excite electrons

---

[1] We ignore here the case of degenerate semiconductors.

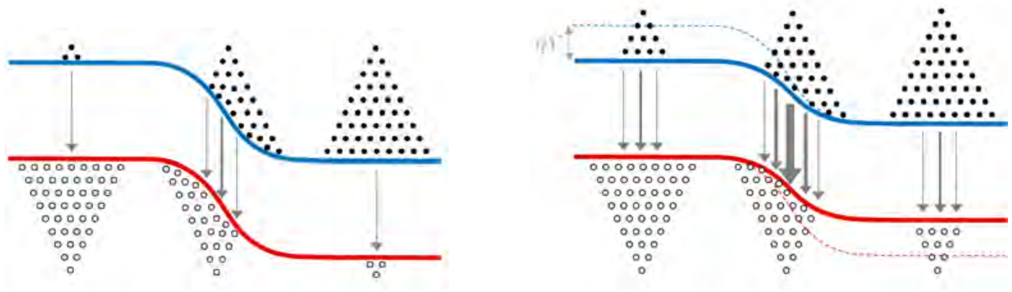
into the higher band; in that case there are empty states in the bands, the carriers can move and the crystal can carry current.

The only known way to emit light in a semiconductor is to force a conduction electron to go back to a lower-energy state in the valence band, and to release its extra-energy under the form of a photon of energy $h\nu = E_{gap}$ [16]. In order to observe this phenomenon, called radiative recombination, one must provide both electrons to the conduction band and holes to the valence band: indeed, if no state is empty in the valence band, no recombination can occur. The more numerous the carriers in both bands, the more numerous the emitted photons, the more intense the light beam. However, it is very difficult to induce in a same location high concentrations of both electrons and holes, since under equilibrium conditions, the carrier concentrations $n$ and $p$ satisfy the following relation [17]

$$n\,p = n_i^2 \tag{2.1}$$

where $n_i$ is the intrinsic carrier concentration. This relation implies that a large quantity of electrons can only coexist with a small amount of holes, and inversely.

The trick to obtain high quantities of both carriers in the same location is to juxtapose a high-electron-concentration region to a high-hole-concentration region, so that the region around the interface presents a large number of electrons and holes, due to their diffusion from one side to the other: this region is called the active region. The juxtaposition of the n-region and the p-region forms what is famously known as a pn-junction, which is at the basis of most electronic devices used today. When no voltage is applied to the junction, the active region, initially rich in electrons



(a) (b)

Figure 2–1: Emission of light in the pn-junction with (a) and without(b) bias

and holes, witnesses their recombination, so that it is eventually poor in both type of carriers: this is why it is also called the depletion region. The ionized donors and acceptors, having lost their carriers, create a barrier of potential, preventing the electrons to reach the p-side and the holes to reach the n-side, which is illustrated in Figure 2–1a. However, if a positive voltage is applied to the junction, the potential barrier can be reduced and the carriers can diffuse to the other side, as described on Figure 2–1b: the active region is then rich in both electrons and holes, which tend to recombine radiatively, resulting in the emission of light.

In order to increase again the recombination of carriers, quantum wells are often introduced in the active region in order to confine the carriers in a smaller space, and therefore increase the recombination probability, enhancing in this way the emission of photons. In that case, the energy of the emitted photons is characterized by the bandgap of the quantum wells, and not by the bandgap of the bulk semiconductor anymore. Putting several quantum wells in the same region is a method to increase further the light emission. In that case, it is even possible to change the composition of each quantum well, so that each has a different bandgap, i.e. a different wavelength of emission: consequently, it is possible to emit different wavelengths from the same device, and even to emit a white light beam from the same device [18].

LED-lamps present several interesting features, comparing to their incandescent and fluorescent equivalents:

- the light is monochromatic: except for the case of multi-quantum wells discussed above, the emitted light present a frequency theoretically equaled to $\nu_0 = \frac{E_{gap}}{h}$. However, because of different processes, the LED spectrum is not as spiky as it should theoretically be, but is more generally a Gaussian centered around $\nu_0$;
- the electrical-to-optical power conversion is remarkably high, and is even expected to reach values beyond 75%;
- the size of the active region can be as small as a few nanometers; the LED-chip can therefore be small as well. Such small sizes are particularly interesting for all kind of display applications;
- they exhibit a long lifetime comparing to other sources of light.

### 2.1.2 Applications

The story of the LED technology is interesting, because it illustrates well the fact that an objectively better technology is not bound to impose itself on the market. Indeed, LEDs have been known for a long time to be capable of high efficiencies, and to constitute a better source of light than incandescent and fluorescent lamps. However, it's only more than 40 years after their discovery that we can find LED-lamps in stores. We are going here to review the different applications of LEDs in time, to see that the first niche-applications were essential to drive the research investments.

Historically, the first application of LEDs took advantage of the long lifespan of the diodes: LEDs were used in alarm indicators and power signaling, followed 15 years later by traffic lights [19]. In the same way, the LEDs were used for backlighting applications, i.e. constituted the source of light necessary to LCD-based screens, such as TV or computer screens, in replacement of neon lamps.

The past decade saw the rise of the LED-share in the automotive lamps: first tail lights and break lights, because red LEDs were the only ones to be efficient enough, but white headlights are now beginning to shift to LEDs. Today, 50% of the tail and break lights of new cars are made with LEDs. Indeed, their small size allow various designs, their efficiency reduces the energy consumption, while their long lifespan prevents the need for frequent replacement.

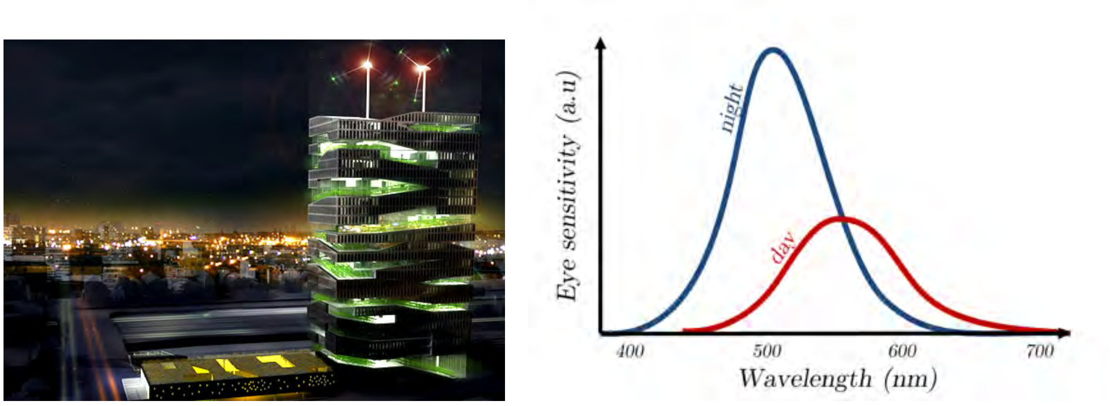
Today, because of their high efficiencies, LEDs are beginning to be used for general lighting, i.e. the illumination of spaces. As we said earlier, these LEDs have the potential to be ultra-efficient, with an efficiency over 75%, which makes them the ideal candidate to replace every light bulb in every space. However, the LED-lamps that are currently sold in our stores are unfortunately only 15% efficient, mainly because their generation of white light involve light frequency down-conversion through phosphors [20], and generation of parasitic heat. However, research efforts are focusing on improving this efficiency, so that 50%-efficient LEDs are expected to enter the market by 2020 [9].

As said above, no LED-lamp would be available today without the first applications. Because of the important investments required, no company or government would have been able to carry by themselves the research cost associated with the direct shift of lighting technology. Public and private companies increased step by step the efficiencies and, as they were making profits, gradually invested in research. For instance, the automotive companies gave a particular boost in investments in the 90s for red LEDs. And it is only in the late 90s that US, Chinese and Japanese governments decided to create public research programs for LED-lighting.

This is the situation that we are in today. However, the story of LEDs is not over. If energy savings are definitely a sufficient factor for the shifting towards LED-lighting, the best is yet to come: indeed, LEDs make it possible to modulate the spectrum of the emitted beam of light, by combining LEDs of different wavelengths [21]. If at least three LEDs are required (the classic R-G-B system) to produce a large color gamut, the more, the better. For instance, a six-wavelength LED-based system allow a fairly effective modulation of the spectrum in the visible range. Why is the spectrum modulation so interesting? We will discuss here three examples.

From the apparition of life 3.6 billion years ago, evolution has trained living organisms to adapt to their environment, and more particularly to the Sun. As soon as life got out of water, 475 million years ago, the Sun got more and more influence on the living organisms, either plants or animals. The fact is that the sunlight spectrum changes throughout the day, which means plants and animals have adapted to this daily variation.

One first application could be the use of LEDs in plant farming. Artificial lighting is already used in developed countries to grow plants for fooding applications. However, this artificial lighting is not coordinated with the sunlight cycle, resulting in a poorly efficient farming. The use of LEDs to fit as closely as possible the sunlight



(a) (b)

Figure 2–2: Artificial farming project (a) and human visual response during day and night (b)

spectrum and its daily evolution could make it possible to grow better and to grow faster. Implementing indoor farming with LEDs could make it possible to reduce transportation costs and might help tackle the hunger crisis in some regions of the globe.

Plants are not the only ones that could benefit from a spectrum-modulated light: mammals, and more particularly humans might enjoy it too. It has been shown that poor-lighting quality could have very concrete effects in mammal's health, one of the worst being cancer development. Spectrum-modulated lighting, fitting the daily evolution of the Sun spectrum, could therefore result in very positive effects in terms of human health, well-being and productivity. For instance, light therapy is already widespread in developed countries to treat chronic depression. Safety applications are also particularly interesting: early-morning-like light could be used to diminish drivers fatigue during the night, resulting in an assumed reduction in road accidents. Moreover, since the peak sensitivity of the human visual system also evolves as a function of the day (555nm during daytime and 505nm during nighttime), as displayed in Figure 2–2b, the headlights' wavelength could be adjusted to fit the optimal response of the visual system, resulting once again in a reduction in the number of road accidents.

The last application of this spectrum-modulation is definitely a leisure one: the enhancement of our visual experiences. It is known that what defines a visual experience is not only the physical characteristics of the object seen, but also the illumination conditions. A typical example (which already exists) would be the dimming of light to create a romantic atmosphere. The extension of this concept is straightforward: the manipulation of the spectrum could drastically change our visual experiences, and therefore our way to see world.

If all these applications are just concepts and ideas now, research progresses by giant leaps, and these ideas could become reality within 10 to 20 years.

### 2.1.3 The droop in efficiency

Ideally, the injection of an electron in the conduction band and a hole in the valence band will result in the emission of a photon in the active region. We can then define the efficiency $\eta$ as the ratio of optical power exiting the device over the electrical power injected into the LED, and decompose this efficiency as a product of sub-efficiencies [20], which reflect the different losses occurring from the injection of carriers to the emission of photons:

$$\eta = \eta_{volt} \times \eta_{inj} \times \eta_{rad} \times \eta_{extr} \tag{2.2}$$

$\eta_{volt}$ is the ratio of the emitted-photon energy over the electron energy supplied by the battery: it is therefore equaled to ${}^{h\nu}/_{qV}$ The electrical potential losses essentially come from ohmic losses at the contacts or at the heterojunction interfaces.

$\eta_{inj}$ quantifies the number of carriers injected into the device that actually reach the active region. Most losses come from leakage currents, such as the semiconductor surface current, the defect-related current and the electron overflow, which will be explained in further detail later.

$\eta_{rad}$ quantifies the number of carriers injected in the active region that recombine radiatively to emit a photon. Non-radiative recombinations such as SRH- and Auger-recombinations are the limitations to this efficiency, and they will also be explained more precisely later.

$\eta_{extr}$ is the ratio of the number of photons exiting the device over the number of photons emitted in the active region. Indeed, many of these photons are trapped into the device, and therefore not emitted into free space, due to total internal reflection[2] and other optical losses (absorption,...).

For more convenience, these efficiencies are presented under the form of two more commonly used quantities:

- the internal quantum efficiency $\eta_{IQE} = \eta_{inj} \times \eta_{rad}$, defined as the number of electrons injected in the LED that actually recombine to emit a photon;

---

[2] The index of refraction of several semiconductors is high (2 for GaN.

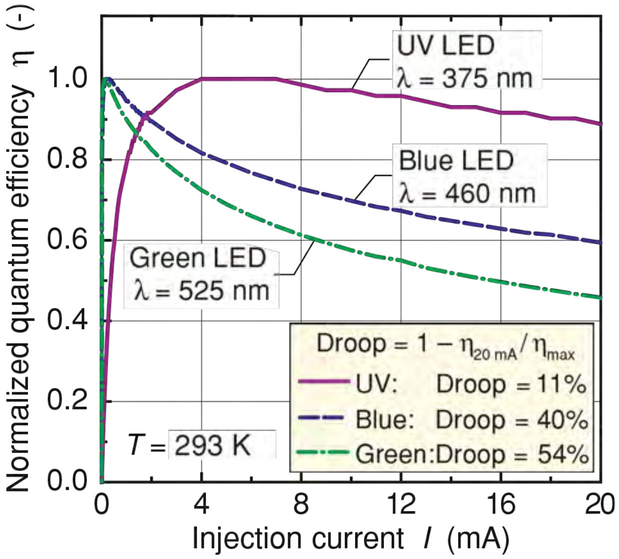


Figure 2–3: Droop in efficiency for different wavelengths [2]

- the external quantum efficiency $\eta_{EQE} = \eta_{inj} \times \eta_{rad} \times \eta_{extr}$, defined as the number of electrons injected into the LED that will result in the emission of a photon exitting the device.

As we have seen in the previous section, the application of LEDs to lighting is interesting because they are in theory much more efficient than incandescent and fluorescent lamps. However, a phenomenon called the droop in efficiency [22, 23, 24], has been observed: as the current going through the LED increases, the total efficiency $\eta$ of the device drops drastically. This droop, presented in Figure 2–3, has been the subject of intensive research for the past two decades, since it constitutes today the main obstacle to switch to LED-based lighting. Since light extraction has reached a good efficiency, $\eta_{extr}$ cannot be improved much more: most efforts concentrate now on improving $\eta_{IQE}$.

To improve $\eta_{IQE}$, we need to know what the causes of the losses are, and what scenarios the carriers can go through. When injected in the active region of the device, the carriers can follow different fates [17], presented below.

- The radiative recombination

  This process involves a conduction electron and a valence hole near the same location: in this case, the electron tends to fill the lower-energy empty state, releasing the surplus of energy under the form of a photon of frequency $\nu = {}^{E_{gap}}/_{h}$. If we assumed that this extra-energy was released via phonons, whose energies are about 50-100 meV, the conduction-to-valence transition could only happen through intermediary states, which do not exist in the bandgap: this extra-energy cannot be released by phonons. In LEDs, the radiative recombination is the process to be enhanced while reducing the other ones. Since this process involves two carriers, one in the conduction band and one in the valence band, its rate is proportional to $n \times p$.

- The Schockley Read Hall recombination

  When undesirable impurities or defects are present in the crystal, the carriers might get caught by them, and it is therefore very difficult to release them back in the bands: we speak of trap states, whose energies are generally situated deep inside the bandgap. These states also act as intermediaries in valence-conduction transitions. If they can help generate holes and electrons, they also offer an alternative to the radiative recombination: instead of a direct conduction to valence transition, the electrons first go to the trap state, releasing its energy under the form of a phonon (or photon[3] ), and then go to the valence band, once again releasing the extra-energy under the form of a phonon (or photon). This process requires only one carrier at each step, therefore its rate is proportional to the carrier concentrations $n$ or $p$.

- The Auger recombination

  In this process, a conduction electron recombines with a hole, but instead of generating a photon, it transfers its energy to another conduction electron, exciting it to a state of energy $E_C + E_{gap}$. This excited electron will tend to relax to the bottom of the conduction band, and the release of the extra-energy is more likely to involve phonons than photons, because a lot of states are present in the conduction band. This process requires two conduction electrons and one hole, so that its rate is proportional to $n^2p$. Similarly, the same process involving two valence holes and a conduction electron can happen, exciting one of the hole to a state of energy $E_V - E_{gap}$, which will then relax back to the top of the valence band through phonons. In that case, the rate is proportional to $p^2n$.

- The electron overflow

  Contrary to the previous processes, which affects $\eta_{rad}$, this one affects $\eta_{inj}$. This process corresponds to the flow of conduction electrons over the active region, instead of their recombination with holes. Experimentally, this electron overflow

---

[3] This photon has an energy $E_C - E_{Trap}$, not corresponding to the bandgap: it is undesirable in most cases.

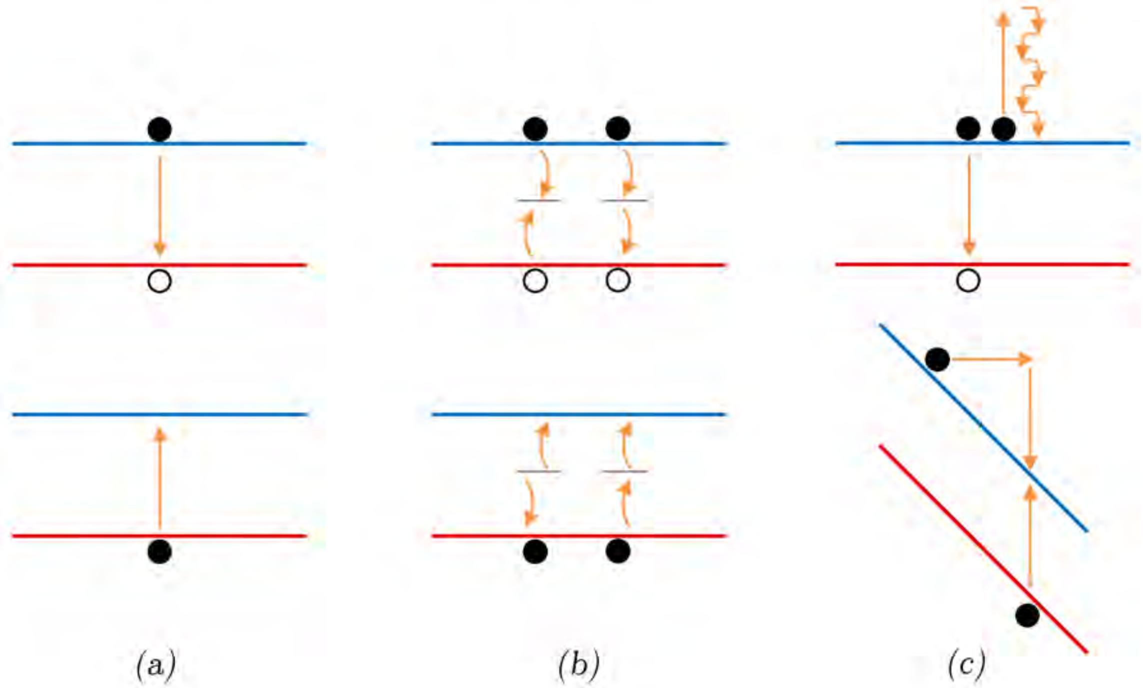


Figure 2–4: (a) Radiative recombination (top) and photogeneration (bottom) (b) SRH-recombination (top) and generation (bottom) (c) Auger recombination (top) and impact ionization (bottom) [3]

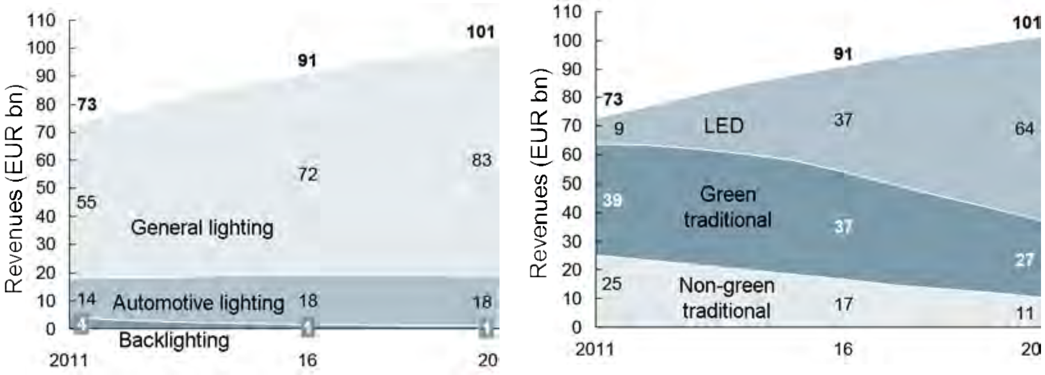
has shown a rate proportional to $n^3$ and $n^4$. Similarly, we could define a hole overflow, but it is generally lower than the electronic one, because holes are less mobile, and they already suffer from a poor injection in the device.

It should be noted that SRH, radiative, and Auger recombinations present inverse generation processes, known as RG-center generation, photogeneration (the principle at the basis of solar panels) and impact ionization (present in avalanche devices).

The important point about these processes is the rates at which they occur: determining the dominant process undermining the radiative recombination is essential to improve the radiative efficiency, and therefore the overall efficiency of the device.

The droop in efficiency presented in Figure 2–3 occurs in a high injection regime, where a lot of electrons and holes are injected in the bands: therefore, the SRH recombination in not the main reason of the droop in efficiency, since its rate proportional to $n$ is negligible comparing to the other processes and their rates above $n^3$. A long debate took place to determine which of Auger recombination and electron overflow was the cause of the droop in efficiency. A recent study [25] seems to have proved that it is actually the Auger recombination that is the main reason of this droop.

If this droop is tackled, and the efficiencies can reach 50% or even 75%, LEDs could become much more efficient than other traditional sources of light, such as incandescent or fluorescent lamps. This is the reason why the lighting market is bound to shift to LED-technology.



(a) (b)

Figure 2–5: Evolution of the lighting market in terms of revenues by sub-markets (a) and by lighting technology (b) [4]

## 2.2 A lighting market bound to shift to LEDs

### 2.2.1 The expected mid-term evolution of the market

In 2012, 16% of the electrical energy produced worldwide was used for lighting applications, which corresponds to one fifteenth of the energy used over world [19]. Light consumption constitutes therefore a huge market, which is usually presented as the regroupment of three sub-markets [26]: the general lighting, corresponding to the illumination of spaces, the automotive lighting, corresponding to lighting in cars (headlights, taillights and interior lights), and the backlighting, mainly corresponding to display applications.

Figure 2–5 shows the expected evolution of the global lighting market between the years 2011 and 2020. It is expected [27] to grow annually by 5% between 2011 and 2016 and by 3% between 2016 and 2020, to generate revenues over 100 billion euros by 2020. The general lighting is expected to grow from 75% of the total lighting market to 80% by 2020, growing annually by 6% before 2016 and 3% after 2016. The

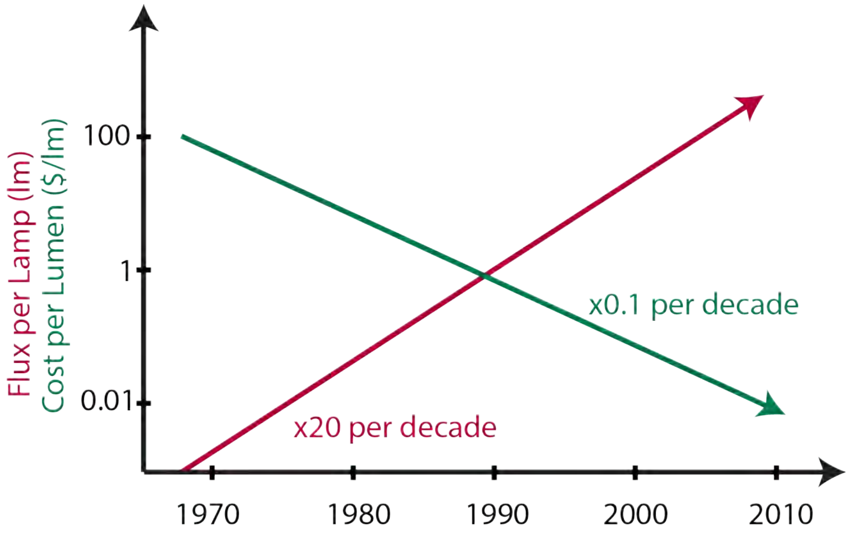


Figure 2–6: Haitz's law

automotive market is expected to saturate around a fifth of the total market, and to reach global revenues of 18 billion euros. On the contrary, the backlighting market is expected to shrink to 1% of the total market, revenues falling from 4 billion euros in 2011 to 1 billion euros in 2020. This trend does not correspond to a reduction of sales, but rather to a substantial decrease of LED prices.

As shown on Figure 2–5b , the LED share in the lighting market is expected [26, 27] to increase drastically, by 33% per year before 2016 and 15% per year after 2016, to reach a 65% share of the global market, becoming the leading technology in the market. As a result of government policies, incandescent and halogen lamps (corresponding to non-green traditional) are expected to shrink from 35% to 12%, and fluorescent lighting will follow a similar but softer trend, losing 25% of its 2011 share to LEDs.

This spectacular penetration of the market by LEDs is linked to the expected evolution of the LED capacities, stated in an equivalent of transistors' Moore's law, called the Haitz's law, displayed on Figure 2–6. Haitz predicted in 2000 that, every decade, the cost of light per lumen would decrease by a factor 10× and the amount of light exiting the LED would increase by a factor 20×. If this cost reduction tends to diminish the market revenues, as it is the case for the backlighting market, another effect is an increase in the demand, resulting in higher revenues.

LEDs are therefore expected to invade the lighting market within a short time, market which is itself booming. How can we explain these two trends?

### 2.2.2 Reasons of this shifting

Numerous reasons lead to an expected growth of the lighting market, and more specifically to the share of the LEDs in this market.

First of all, why is the lighting market expected to increase so much by 2020? One answer lies with the undeniably increasing demand resulting from the growing world population, expected to rise from 6.9 billion in 2010 to 7.7 billions in 2020 [26]. These additional 800 millions inhabitants will require new resources, among which food, energy and housing, and therefore lighting. Asia, and more particularly China, are driving this population growth, and are consequently expected to dominate the future lighting market: China's general lighting market is expected to double between 2011 and 2020 and its share to go from 18% to 23%, while Asia is also expected to double its market size,its share rising from 40% to 50%. The future lighting market will definitely be centered on Asia.

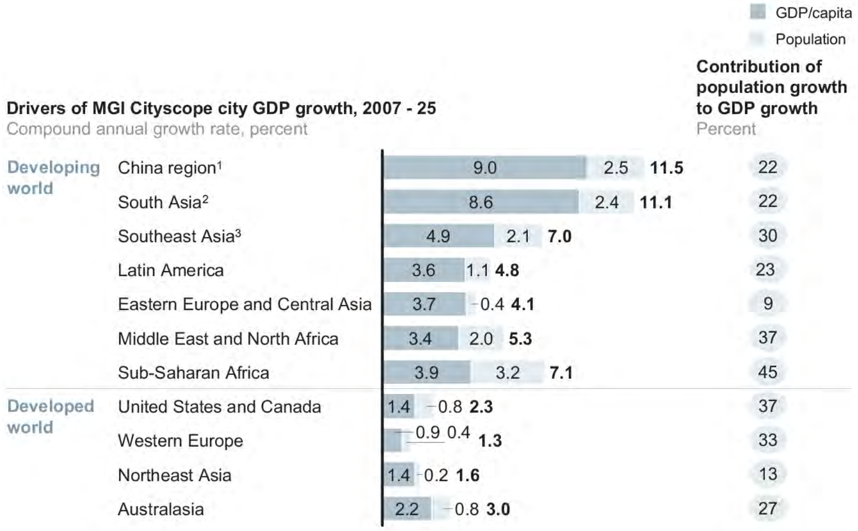


Figure 2–7: Growth rate of population and per-capita gdp for different regions of the world [5]

Another answer is the growing urbanization [26]. For the past decades, the rural exodus has intensified, resulting in new and bigger cities. The growing population will also obviously contribute to this need for new cities. Since cities require more lighting than the rural areas they replace, a strong increase in the lighting demand is expected. The growth in demand will essentially concern developing countries, in Asia but also in Latin America: by 2025, 140 newly-built cities (100 in China alone!) are expected to join the 600 richest cities. These considerations are in agreement with the observed trend that construction efforts are proportional to the GDP, which is also expected to increase from 3% to 4% each year worldwide.

The second feature of the expected evolution of the lighting market was the rapid increase of the LED-share. How can this be explained? Most important of all, LEDs promise tremendous energy savings, due to their projected ultra-efficiency. The past decades have shed light on the scarcity of resources and global warming crisis that we are facing, and reducing our energy consumption has become inevitable for the public opinions. LEDs therefore constitute an efficient way of saving energy, within a horizon of less than a decade: switching to LED-lighting would result in a reduction of 10% of our electricity consumption and 4% of our total energy consumption [9]. Moreover, a recent study showed that shifting from incandescent and fluorescent to LED is not only a good way of saving energy, but is also one of the rare that will also generate profit [26], instead of costing money to the tax payer, as is shown in Figure 2–8. LEDs would generate profits around 140$ per ton of $CO_2$ saved, where solar panels (yet much more efficient in terms of $CO_2$ abatement) would cost around 80$ per ton of $CO_2$ saved. Another study [28] shows that an 10% increase in LED-efficiency would result in an \$65B/year increase in world GDP, which is explained in Appendix C. It is therefore one of the easiest investments both in terms of revenues and time delays.

To summarize, the growing population and urbanization are expected to drive the increase in the demand for lighting, while ecological considerations will cause the penetration of LEDs in this market.

### 2.2.3 Government policies and public opinions

It is undeniable that shifting to LED-lighting would result in great results for our society. Unfortunately, as said previously, the promises are not enough for the change to happen. The shifting requires both massive investments to improve the efficiency and governmental policies to induce a change in society.

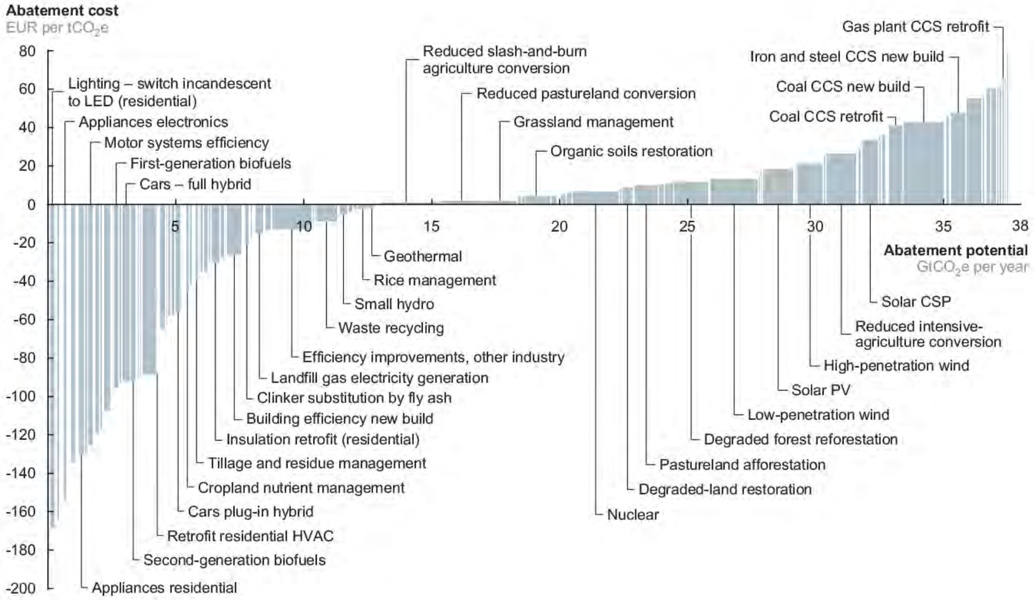


Figure 2–8: Abatement costs and reduction of $CO_2$ emissions for different ecological policies [5]

The first example of such policies is the incandescent ban decided by several government [27, 29], whose goal, as its name suggests, is to ban any incandescent lamp from residences, offices and public spaces. Brazil and Venezuela were actually the first to implement the ban in 2005. Europe started it in 2009, alongside with Australia, Switzerland and Japan: all kinds of incandescent lamps were banned in 2012 and a new progressive ban on halogen lamps was decided in 2014, so that none of them would be used by the end of 2016. Similarly, the United States have banned in 2014 all kinds of incandescent lamps. More importantly, China has also implemented such a ban, in order to switch to LED-lighting, on a 5-year phase-out plan, starting in 2012 to finish in 2016. As we have said, Asia, driven by China, is expected to increase its share in the global lighting market, reaching 50% by 2020: it is therefore essential that Asian governments take a strong position towards LED-lighting for the switching to happen.

A good example of how public opinion induces government policies is the Fukushima catastrophe that happened in Japan at the beginning of 2011 [27, 30]. On March 2011, $11^{th}$ a tsunami, generated by an earthquake of magnitude 9, originated 70km away from the Japan coasts and stroke the Fukushima nuclear reactor 50 minutes later. Even though the reactors were shut down, the water flooded the emergency power generator, resulting in a heating of the reactors, followed by the meltdown of three of them, driving Japan into its worst crisis since the second World War.

The catastrophe had a great impact on public opinions, especially in European countries, forcing governments to review their nuclear policy, as shown in Figure 2–10 if not completely abandoning nuclear power. For instance, Germany decided to quit nuclear power by 2020[4] , while France[5] decided to reduce its nuclear share to 50% by 2025. Japan, despite the catastrophe, chose in early 2014 to maintain its nuclear power capacities, in spite of the public opinion. The phase-out and reduction policies decided by governments will undeniably result in a increased energy cost and perhaps shortage of energy production. Reducing our energy consumption is once again a necessity, and it is in this way that LED-lighting can meet the new energy demands of societies.

---

[4] German nuclear share in the energy production was 18% in 2011.

[5] French nuclear share in the energy production was 75%.

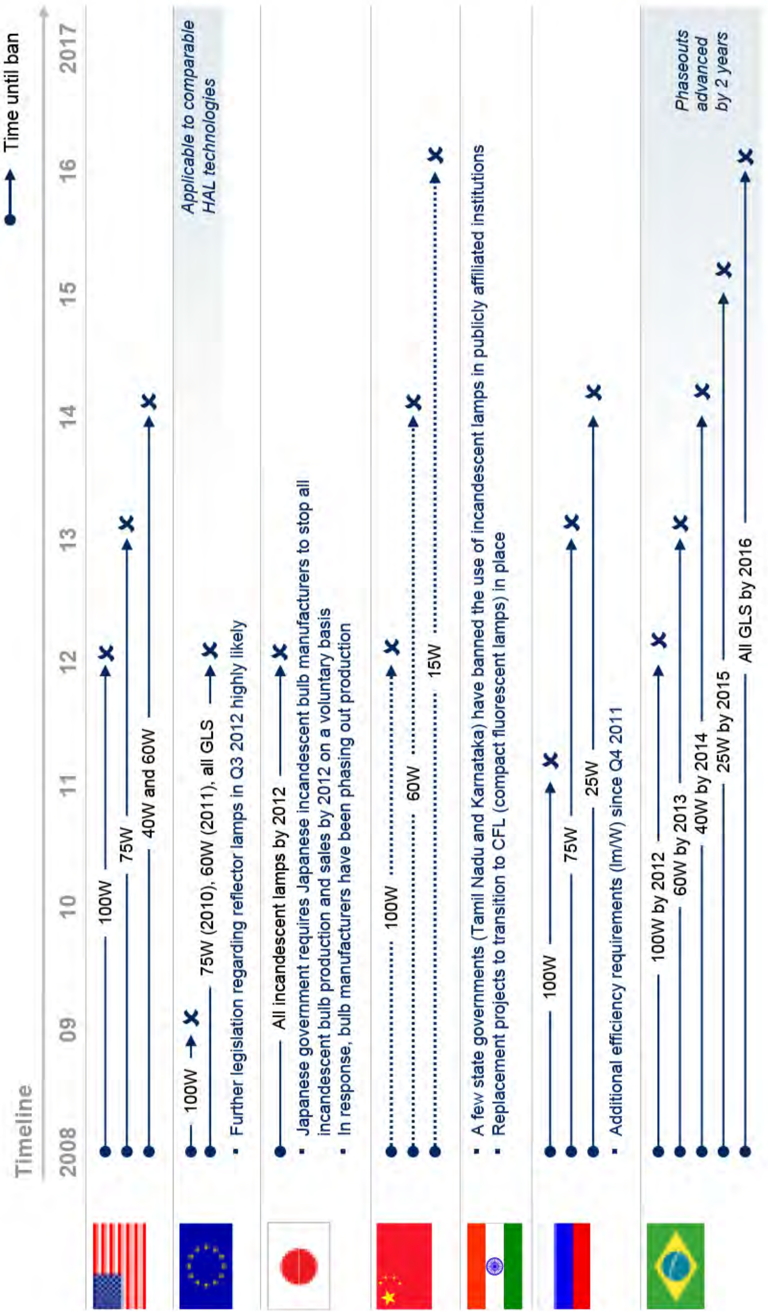
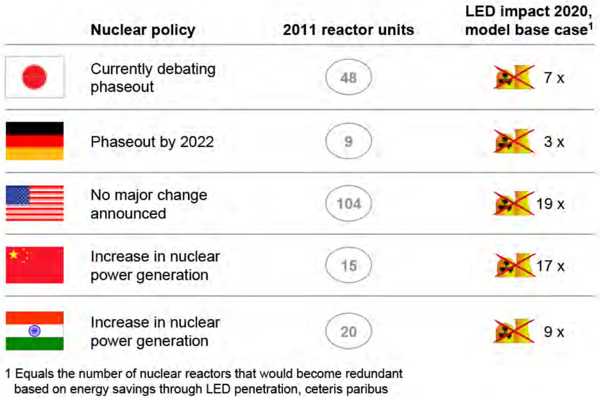


Figure 2–9: Incandescent ban of several countries [4]



| | Nuclear policy | 2011 reactor units | LED impact 2020, model base case[1] |
|---|---|---|---|
| | Currently debating phaseout | 48 | 7 x |
| | Phaseout by 2022 | 9 | 3 x |
| | No major change announced | 104 | 19 x |
| | Increase in nuclear power generation | 15 | 17 x |
| | Increase in nuclear power generation | 20 | 9 x |

1 Equals the number of nuclear reactors that would become redundant based on energy savings through LED penetration, ceteris paribus

Figure 2–10: Evolution of the nuclear policies after the Fukushima accident [4]

## 2.3 The III-Nitride family

### 2.3.1 Their necessity for white light

As seen in the previous sections, LED-based lighting is expected to invade the lighting market. However, we have seen that the LED-spectrum is known for its spiky features, where white light is generally characterized by a smoothly varying and wide spectrum. The ultimate white light is the Sun spectrum, since it is the light that living organisms have adapted to. How can approach this spectrum with LEDs? There are currently two known ways:

- the combination of Red, Green and Blue LEDs: it is the technique currently used in most screens; each pixel is made of 3 sub-pixels, 1 green, 1 blue and 1 red, whose intensities are regulated in order to produce a given color [6] . If the 3 sub-pixels emit at the same high intensity, the pixel can produce a light ranging from black to white

---

[6] Not all colors are reachable, and the gamut of colors is defined as the colors the screen is able to produce.

passing through grey. This technique is not limited to 3 sub-pixels, and the color gamut can be greatly improved by increasing the number of LEDs constituting the pixel. For instance, a 6-LED pixel would allow for a more interesting gamut.

- the combination of a blue LED and a yellow phosphor: as it can be seen on the color space in Figure 2–11, coating a blue LED with a yellow phosphor result in an emitted light color on the blue/yellow axis, where we can find colors close to white. However this technique is less efficient, since it involves a down-conversion from blue to yellow through phosphors, associated with an energy loss over 50%.

If reproducing exactly the sunlight spectrum seems difficult, it can yet be approached, and the two known ways to create white light both require blue LEDs. A blue LED needs to have a bandgap greater than $2.7eV$, and the most commonly used semiconductors corresponding to these bandgaps are GaN and AlN, belonging to the famous III-Nitride family

### 2.3.2 A III-V semiconductor family

The III-Nitride family consists of binary compounds between atoms of column III in the periodic table (Boron, Aluminum, Gallium, Indium,...) and Nitrogen atoms, which belong to column V. In our case, we only focus on AlN, GaN and InN, because all three of them crystallize in a very specific structure, called the wurtzite structure.

The wurtzite lattice correspond to the interpenetration of two close-packed hexagonal lattices, one corresponding to the column-III atom, the other corresponding to the Nitrogen atom . Three essential features of this structure for its application in polarization-doping are:

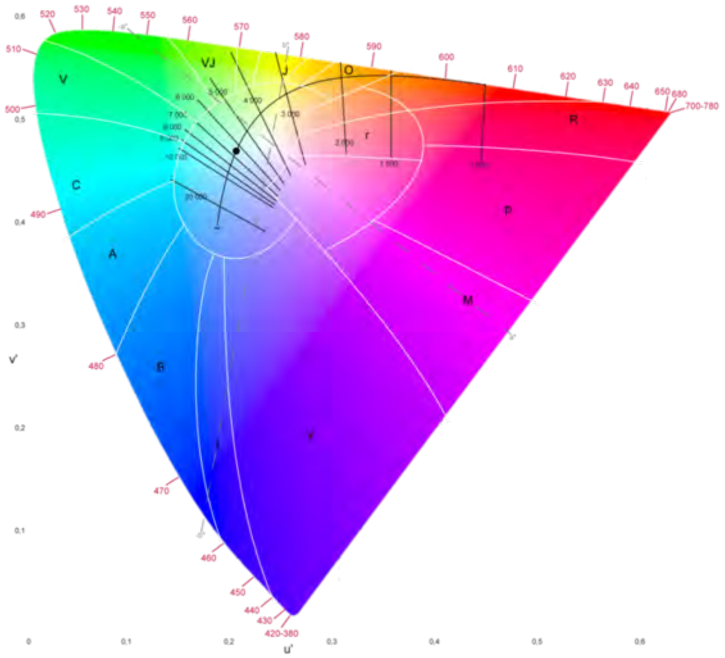
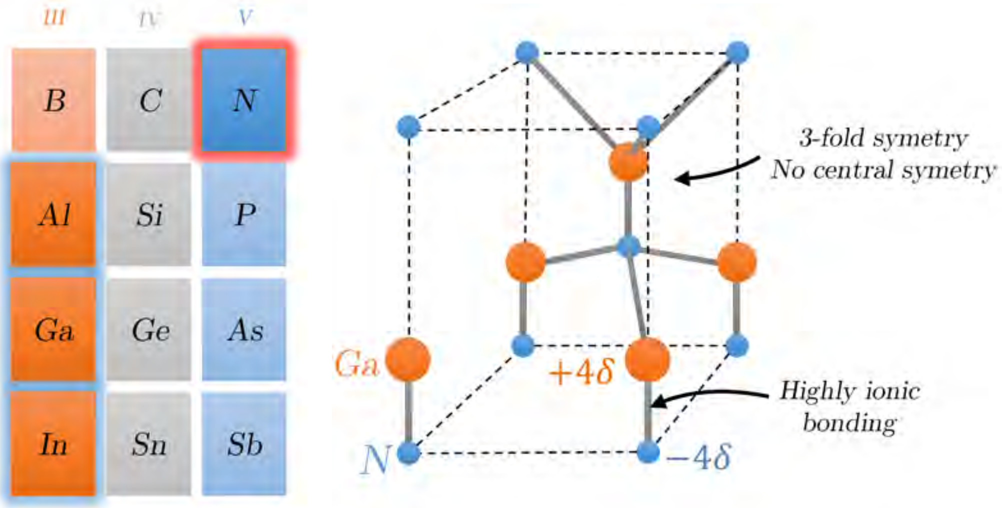


Figure 2–11: Color space: producing white light with LEDs is possible via (1) the combination of Red Green and Blue (2) the combination of Blue and Yellow.



(a) (b)

Figure 2–12: The periodic table (a) and the wurtzite lattice (b)

- its 3-fold symmetry in the $x - y$ plane, resulting in considerations only along the c-axis, perpendicular to the base;
- its lack of inversion symmetry: the unit-cell and its upside-down equivalent are not identical;
- its highly ionic bondings: because of the strong difference in their electronegativities, any bonding between a III-atom and a N-atom will result in a positive charge $+\delta$ located on the III-atom and a negative charge $-\delta$ located on the N-atom. Since any atom in the wurtzite structure is involved in 4 direct bonds, any N-atom carries a charge $-4\delta$ and any III-atom carries a charge $+4\delta$.

The lack of inversion symmetry and the ionic bondings are necessary to observe what we will define later as spontaneous polarization.

Another important characteristic of III-Nitride semiconductors, in terms of application, is their direct bandgap, which means that the maximum of the valence band and the minimum of the conduction band occur at the same point in $k$-space. On the contrary, $Si$, $GaP$, $AlP$ and $AlAs$ present indirect bandgaps, which means that any conduction-valence transition will require the emission/absorption of a phonon to change $k$-points [7] . The necessary presence of this phonon greatly reduces the probability of the transition, so that all optoelectronic devices based on direct-bandgap materials are much more efficient than their indirect-bandgap equivalents.

---

[7] A photon cannot carry a sufficiently big moment to ensure the conservation of momentum during the transition.

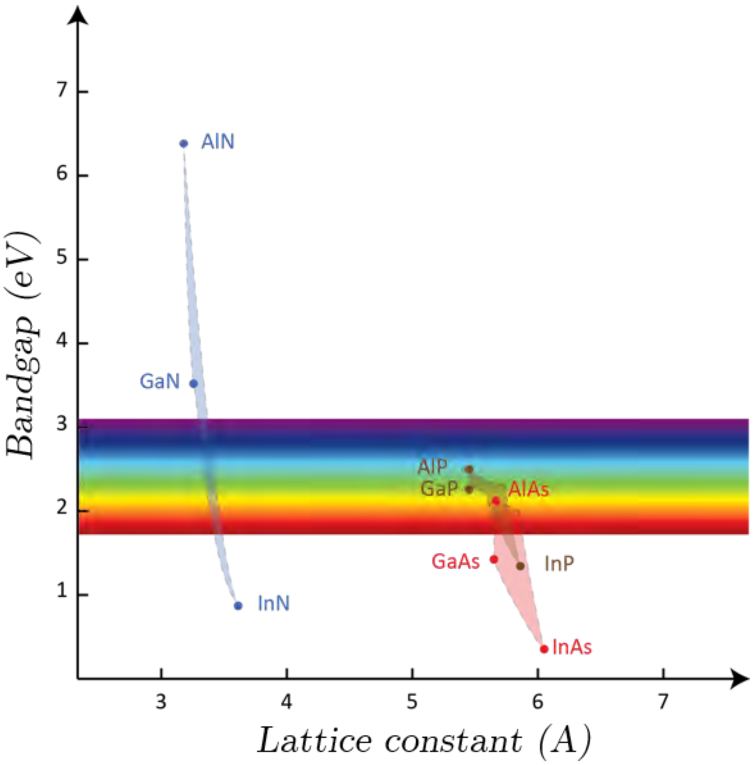


Figure 2–13: Bandgaps of III-V compounds

Eventually, the last major advantage of the III-Nitride semiconductors concern the wide range of bandgaps reachable by alloying them. It is therefore theoretically possible to reach any point in the blue triangle of Figure 2–13 by creating an alloy $Al_xIn_yGa_{1-x-y}N$, but the lack of precise experimental measurements in this family makes it difficult to choose $x$ and $y$ to reach precisely the wanted bandgap and lattice constant. This large range of bandgaps results in the possible emission of photons from mid-infrared to deep-UV, covering the whole visible window.

### 2.3.3 Other general applications

We can find the III-Nitride family in various semiconductor devices, such as:

- Field Effect Transistors: because they exhibit a large dielectric strength, AlGaN and GaN FETs are used today in all systems requiring high-voltage operations and power electronics.

- High Electron Mobility Transistors: a highly n-doped layer of AlGaN is put next to a intrinsic GaN layer, resulting in the presence of a 2D electron gas at the interface between the layers for two reasons: a higher electron affinity in GaN and the presence of a positive charge at the interface, due to a change in polarization between the materials[8] . This 2DEG exhibits an extremely good mobility, since there is no impurity in this region, and its conductivity can be modulated by the application of a gate voltage, as for FETs.

---

[8] This phenomena will be explained in Section 3.2, and is at the basis of polarization-induced doping.

- Chemical sensors: the principle is the same as that of the HEMT, i.e. a 2DEG is created between a layer of AlGaN and GaN. When a polar molecule approaches the surface of the sensor, the electrostatic potential due to this molecule acts just like a gate voltage, and therefore modify the conductivity of the 2DEG. This change of conductivity can be easily measured, and one can identify the presence of a molecule. Molecule-specific receptors are added to the surface, so that only the specific polar molecule can disturb the 2DEG.

- LEDs, lasers and light detectors: as we have seen previously, the III-Nitride family is characterized by its ability to reach wide bandgaps, which corresponds to the emission of high frequency light beams. It is therefore possible to emit blue light or UV radiations. LEDs are used to create blue pixels in screen, light detectors can be designed to detect UVs, but the most interesting aspect of these optoelectronic devices is the use of blue/UV lasers with detectors: since the photons exhibit a high frequency, it is possible to transfer data via optical fibers at a high speed, and the global communications systems (internet, telecoms,...) today requires these optical fibers. Another example of high speed transfer of data is the well known Blu-Ray system, used to store a lot of information (25GB per layer) on a single disk and to process the data at a high speed.

### 2.3.4 The difficult p-doping

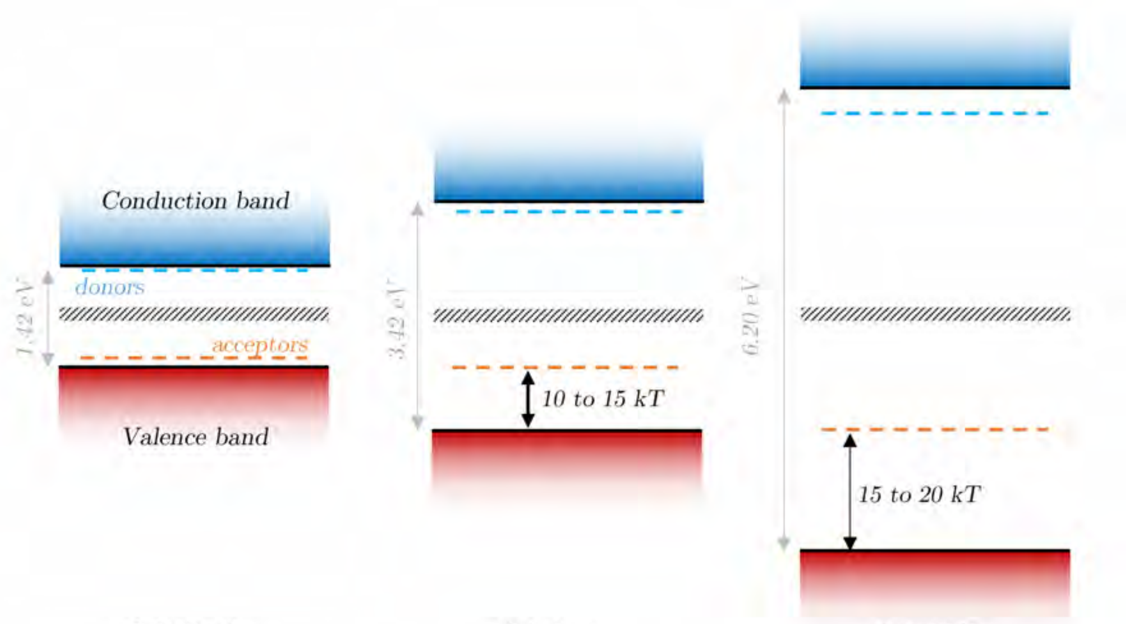
As we have seen in Section 2.1.1, LEDs (and any optoelectronic devices) rely on conduction-valence transitions, requiring therefore a lot of electrons in the conduction band and a lot of holes in the valence band.

In order to provide electrons to the conduction band, *donors* are introduced in the lattice. These donors need to have more valence electrons than the atoms they replace, so that these extra-electrons can be easily released into the conduction band, when enough thermal energy is present in the crystal. In the band diagram, the energy of donors is represented by a dashed line under the conduction band; the closer to the conduction band, the exponentially easier the ionization.

Similarly, *acceptors* are capable of releasing holes in the valence band, or equivalently of catching the electrons from the valence band. An acceptor present a lower number of valence electrons than the atom it replaces, being in this way capable of accepting electrons in its valence empty states. Once again, the closer to the valence band the acceptor states are, the exponentially easier the ionization.

If finding donors close to the conduction band was not a problem, looking for acceptors of small ionization energies turned out to be much more difficult for III-Nitrides. Indeed, the ionization of dopants is essentially thermal, which means the crystal needs to release a bit of its thermal energy to ionized the dopants. However, the thermal energy the crystal is able to provide is generally considered to be around a few $k_BT$. In the case of GaN and AlN, the ionization energy of the best acceptors (Mg) is respectively more than 10 $k_BT$ and 15 $k_BT$, i.e. much more than the few $k_BT$ available. The ionization of these acceptors is consequently very low, and experimental measurements have shown that it is less than 1%.

Therefore, in III-Nitride semiconductors, it is not an issue to provide electrons to the conduction band, but it is one to provide holes to the valence band, since the acceptor-based p-doping is extremely inefficient. The 1% acceptor-ionization



(a) GaAs (b) GaN (c) AlN

Figure 2–14: Ionization energy of acceptors for different semiconductors

would require to increase the amount of acceptors by a factor 100, which implies the introduction of 100 times as many impurities, with a strong effect on the electron and hole mobilities and a strong increase in non-radiative recombination, such as SRH recombination: the device will exhibit a poor efficiency. This is where polarization-induced doping (p-doping in our case, but it can be n-doping as well) steps in as an alternative to dopants to provide the necessary carriers to the device.

# CHAPTER 3
# Polarization and polarization-induced doping

## 3.1 The nature of polarization

Polarization (also known as polarization density) refers to the density of electric dipole moments in a given material. If we look at two particles carrying opposite charges $\pm q$ placed at position $r_1$ and $r_2$, we can define the electric dipole moment.

$$\mathbf{d} = q\left(\mathbf{r}_2 - \mathbf{r}_1\right) \tag{3.1}$$

In the case of a crystal containing two or more charged particles in each unit cell, we can define the polarization as

$$\mathbf{P} = \frac{\langle \mathbf{d} \rangle}{V} \tag{3.2}$$

$$\mathbf{P} = \frac{\sum_{unit\ cell} q_i \mathbf{r}_i}{V_{unit\ cell}} \tag{3.3}$$

$$\mathbf{P} = \frac{q_+ \mathbf{R}_+ - q_- \mathbf{R}_-}{V_{unit\ cell}} \tag{3.4}$$

$$\tag{3.5}$$

where $q_{+\,(-)}$ corresponds to the sum of the positive (negative) charges, and $\mathbf{R}_{+\,(-)}$ refers to the geometric center of the positive (negative) charges. For any complexity of the unit cell, we can always go back to a simpler 2-particle electric dipole, resulting from a charge $q_+$ at $\mathbf{R}_+$ and a charge $q_-$ at $\mathbf{R}_-$.

To obtain a non-zero polarization, we must have charged particles, whose geometric centers are separated in space. There are different ways to obtain a physical separation between the centers of positive and negative charges.

If we consider any crystal, ionic or not, and an electric field is applied on it, the positive nuclei will try to follow the direction of the field, while the negative electronic cloud will tend to move in the opposite direction, resulting in a separation between the center of the nuclei and the center of the electrons, as shown in Figure 3–1 (a,b). The polarization is called the *electric polarization* and is proportional to the electric field:

$$\mathbf{P}_{el} = \chi \mathbf{E} \tag{3.6}$$

where $\chi$ is called the electric susceptibility. In that case,everything happens as if the material was feeling a corrected electric field $\epsilon \mathbf{E}$ where $\epsilon = 1 + \chi$.

In the case of ionic or partially-ionic crystals, the constitution energy of the crystal often reaches a minimum for symmetric structures. However, in some structures like wurtzite, the lattice lacks a degree of symmetry. In that case, the centers of positive and negative charges may not coincide (Figure 3–1 (c,d)), resulting in a non-vanishing polarization at equilibrium. It is called the *spontaneous polarization*, and is only determined by the ionicity of the bonds between the atoms, the position of the centers of negative and positive charges and the volume of the unit cell.

The last case of polarization discussed here is what is called the *piezoelectric polarization.* If one applies a strain on a crystal, the lattice will distort, so that the atoms in the unit cell are displaced from their equilibrium positions. If there is an

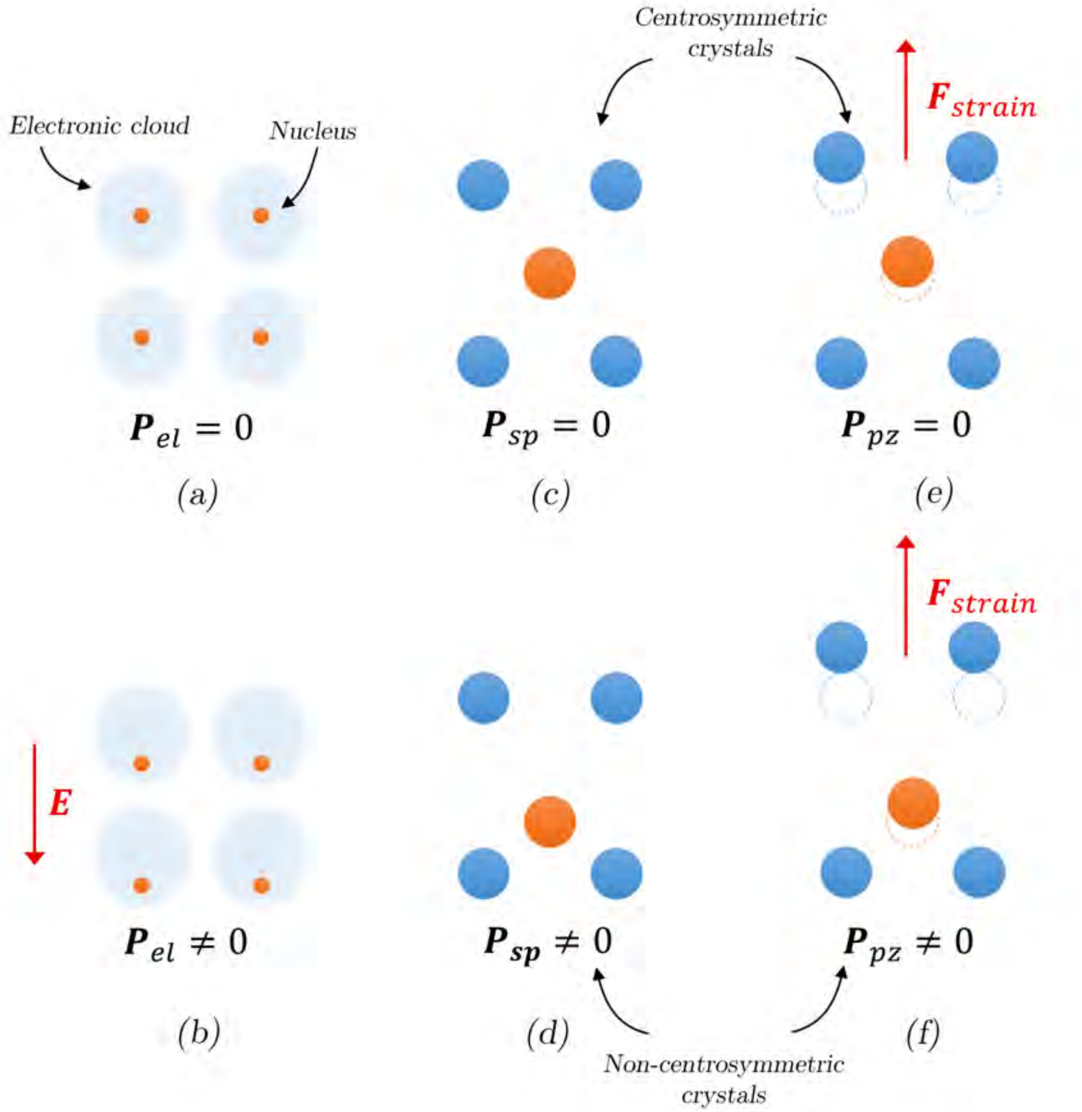


Figure 3–1: The different types of polarization: electric polarization (a,b), spontaneous polarization (c,d), piezoelectric polarization (e,f)

inversion symmetry, as explained in Figure 3–1 (e,f), whatever strain is applied, the centers of positive and negative charge will still coincide: no polarization is observed. But if the ionic crystal lacks inversion symmetry, then the positions of the center of positive and negative charges will change, and there will be a correction to the polarization of the crystal: this correction is the piezoelectric polarization.

The two essential conditions to observe spontaneous polarization (and therefore piezoelectric polarization as well) are the following:

- (partially) ionic bondings between the atoms;
- an absence of inversion symmetry.

## 3.2 The effects of polarization

To understand the effects of polarization on the charge carriers, let's look at the situation of a crystal exhibiting spontaneous polarization. As previously said, we can simplify a unit cell just to the centers of positive and negative charges. In this picture each positive charge is placed between two negative charges, except for the one on the left boundary: we can therefore deduce that the absence of the negative charge on the left is equivalent to a positive surface charge at the left-boundary of the crystal. In the same way, the absence of the positive charge on the right on the crystal is equivalent to a negative surface charge at the right-boundary of the crystal. These boundary-charges, resulting from the polarization, will be referred to as the *polarization-charges*. This intuitive reasoning is supported by the Maxwell-Gauss equation

$$\nabla \cdot \mathbf{D} = \rho \tag{3.7}$$

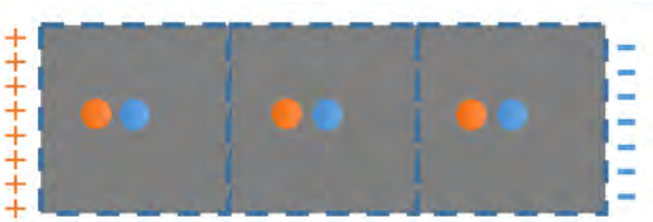


Figure 3–2: The polarization-charges

By performing the classical integration over a cylinder of area $A$ and hight $h$ centered on the boundary, and forcing $h \to 0$, we find

$$\lim_{h\to 0}\int_{-h/2}^{h/2} \nabla\cdot\mathbf{D} = A\ \big(\mathbf{D}(0^-) - \mathbf{D}(0^+)\big) = A\,\sigma \tag{3.8}$$

since $\mathbf{D} = \epsilon\mathbf{E} + \mathbf{P}_{sp} + \mathbf{P}_{piezo}$, we can define the polarization charge as

$$\sigma_{Pol} = -\Delta\mathbf{P}_{sp} - \Delta\mathbf{P}_{piezo} \tag{3.9}$$

This charge $\sigma_{Pol}$ is not a fictional charge and is at the origin of a very real electric field, whose action will result in attracting carriers of opposite signs to compensate this polarization charges: we decided to call these carriers the *compensating charges*. If the materials on the left and right of the crystal are identical, then the two surface polarization charges are identical but of opposite signs.

Usually, these surface polarization charges are so big that they cannot be screened completely by carriers, resulting in a non-zero electric field in the crystal [31]. This can be the cause of reductions in efficiency, as it is the case in some quantum well devices: for instance, when an InGaN quantum well is placed in the active region of a GaN device, the difference of spontaneous polarization between the two materials results in strong surface charges at the interfaces of the quantum well. These polarization charges are not compensated, resulting in a strong electric field inside the quantum well, which spatially separates the electrons from the holes, resulting both in a reduced recombination and a smaller bandgap, corresponding to a shift towards the red for the emitted wavelength [32].

## 3.3 The polarization-induced doping

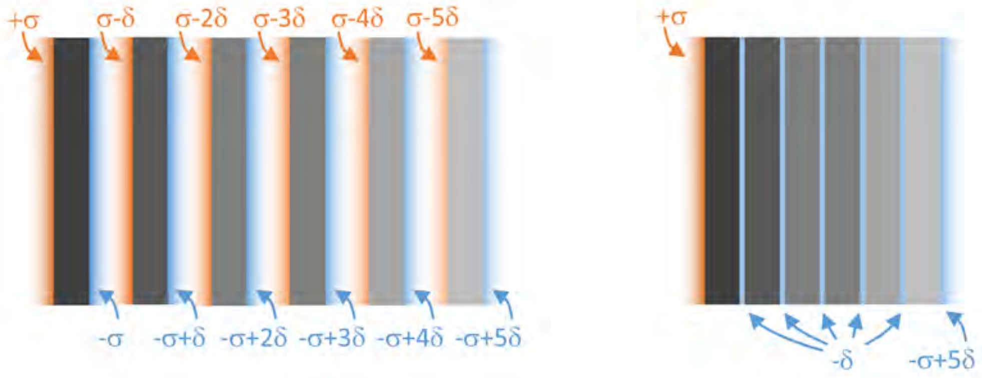
How can we go from this inconvenient surface polarization-charge to the desired bulk-doping? Let's imagine the case of a family of crystals which can grow in the same lattice structure and exhibit different polarizations, which result in different surface polarization-charges. Let's assume for simplicity that the layer $i$ exhibit polarization charges $\pm\sigma_i = \sigma - (i-1)\delta$, as described on Figure 3–3. When the different layers are grown one onto the other, the surface polarization charge on the right of the layer $i$ adds up with the surface polarization charge on the left of layer $i+1$, leaving a total surface polarization-charge $-\sigma_{i-1} + \sigma_i = -\delta$ at the interface between the layers. There is therefore a $-\delta$ charge at each of the $n$ interfaces, except at the first one, which carries a $+\sigma$ charge, and at the last one, which carries a $-\sigma + (n-1)\delta$. It should be noted that the sum of all polarization-charges equals zero.

If we follow a mathematical approach, using the Maxwell-Gauss equation, we find

$$\nabla(\epsilon\mathbf{E}) = \rho - \nabla\left(\mathbf{P}_{sp} + \mathbf{P}_{piezo}\right) \tag{3.10}$$

and we can therefore define the polarization charge as

$$\rho_{Pol} = -\nabla\left(\mathbf{P}_{sp} + \mathbf{P}_{piezo}\right) \tag{3.11}$$



(a) before superposition (b) after superposition

Figure 3–3: Grading of a material to induce a 3D polarization-charge

This polarization charge creates an electric field through the Maxwell-Gauss equation, which will attract mobile charges of opposite sign, resulting in the same situation as the impurity-based doping: a mobile charge linked to a polarization charge in one case, linked to an ionized dopant in the other case.

It should be highlighted here that the surface polarization charges located at the vacuum-semiconductor interfaces are much more important than the bulk polarization charge. On Figure 3–3, this is equivalent to saying that $\delta \ll \sigma$. We will see later that these surface charges at the boundaries of the crystal can cause major issues, and that there is a smart way to avoid these. Ideally, we would like to avoid these strong surface charges.

## 3.4 The use of the III-N family

As we have seen in the preceding section, we need to find materials which:

- are semiconducting;
- crystallize in a polar structure;

- crystallize in the same structure;
- exhibit different polarizations.

Aluminum- , Gallium- and Indium-Nitride satisfy every one of these four conditions, and any binary or ternary alloys too.

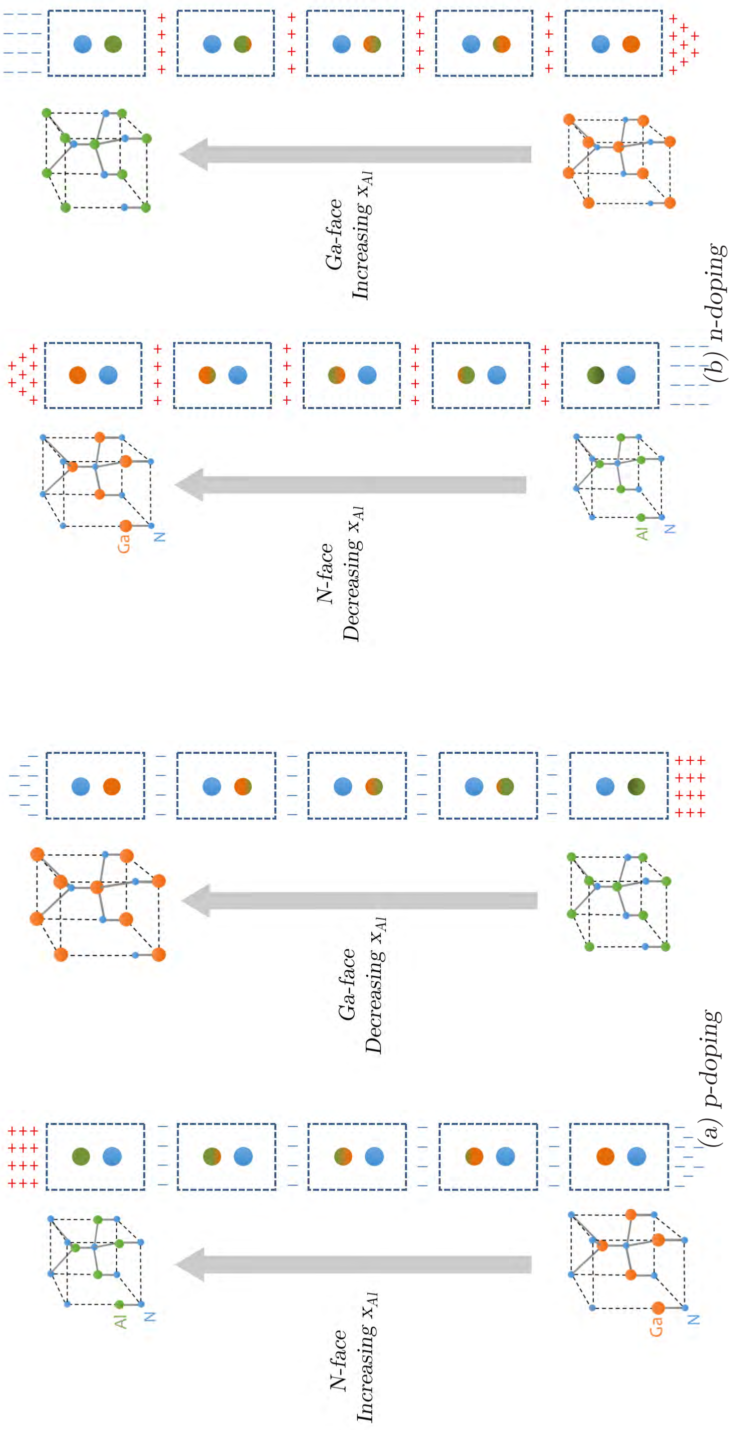
The idea to create the gradient of polarization necessary to the doping is to act on the spontaneous polarization. The spontaneous polarization of AlN is bigger than that of GaN [1] , which is bigger than that of InN. Consequently, a binary alloy $Al_xGa_{1-x}N$ should allow us to reach any spontaneous polarization between $\mathbf{P}_{sp}\left(GaN\right)$ and $\mathbf{P}_{sp}\left(AlN\right)$, by choosing the previously calculated value of $x$. It should be noted that the relationship between x and $\mathbf{P}_{sp}\left(Al_xGa_{1-x}N\right)$ is not linear, but a quadratic approximation turns out to be a good fit [33, 34, 35, 36].

There are two ways to create a type of doping. For instance, to induce n-doping, we need to create a positive polarization charge, similar to the positive charge of the ionized donors, so that it attracts electrons and constitute what is referred to as a 3D electron-slab[2] . This positive polarization charge can result either from the N-face growth from AlN to GaN, or from the Ga-face growth from GaN to AlN. However, these two ways differ on one important aspect: the location of the strong polarization charge at the boundaries. Since these boundary-charges are important, it would be

---

[1] The Al-N bonding is less ionic than the Ga-N bonding, so that the moment $\mathbf{d}$ is weaker in AlN than in GaN; but since the spontaneous polarization is defined as $\frac{\langle \mathbf{d} \rangle}{V}$, and the volume of the unit cell of AlN is less than that of GaN, the spontaneous polarization is stronger in AlN than in GaN.

[2] This name is derived from the 2D electron-gas used in HEMTs.



(a) *p-doping*

(b) *n-doping*

Figure 3–4: Growth configurations

more interesting to put the positive-one on the p-side and the negative one on the n-side. If not, there could be two depletion regions at the boundaries hindering the performance of the device.

The same reasoning applies for p-doping. To induce a 3D hole-slab, we need to create a negative polarization charge, which will attract holes. Two ways can be used: either growing N-face from GaN to AlN or growing Al-face from AlN to GaN. Once again, these two methods are not really equivalent, since the signs of the strong boundary-charges are different.

### 3.5 Calculation of the grading profiles

Inducing a smooth variation of the Aluminum content $x$ results in a variation of $\mathbf{P}_{sp}$, which creates the polarization charge. However, the variation of $x$ also causes a variation of the lattice constant $a$, which has for consequence the existence of a strain on the upper layers of the crystal[3] . This strain causes the apparition of a piezoelectric polarization, which can be either in the same or the opposite direction of the spontaneous polarization (depending on which of AlN or GaN is the substrate): this strain is obviously a side-effect that we would like to minimize, since this strain is the cause of dislocations and other defects that reduce the efficiency of the device.

---

[3] The lattice constant of the whole crystal is imposed by the lattice constant of the substrate, GaN in most cases.

Is it possible to eliminate this side-effect? We think it is. Ternary alloys $Al_xIn_yGa_{1-x-y}N$ can exhibit different values of spontaneous polarization with the same lattice constant $a$, i.e. without any strain if one is grown onto the other. However, in this thesis, only binary alloys are considered.

To reach a given value of doping, one needs to design the profile of the Aluminum content $x$ that will result in the doping. This calculation is explained in references [35, 36, 33, 34] and has been implemented in our model. As said before, the relationship between $x$ and $\mathbf{P}_{sp}\,(Al_xGa_{1-x}N)$ is not linear, but almost, so that a linear variation of $x$ results in quasi-constant polarization charge. Figure 3–5 exhibits the mean value of the polarization charge for a variation of the Aluminum content of $\Delta x$ over a distance $d$. The scale is logarithmic and the charge can be either positive or negative, depending on the direction of the grading.

We can see that the reachable polarization charges are similar for the three different binary compounds $Al_xGa_{1-x}N$, $Al_xIn_{1-x}N$, $In_xGa_{1-x}N$, mostly between $10^{13}cm^{-3}$ and $10^{20}cm^{-3}$. One interesting point on Figure 3–5 is that the visible lines on diagram (b) are less linear and more quadratic than on diagrams (a) and (c): this is due to the fact that the lattice mismatch is bigger between AlN and InN than it is for AlN/GaN and GaN/InN; this bigger lattice mismatch results in a bigger strain, and the piezoelectric polarization is a quadratic function of the strain; we can conclude that the piezoelectric polarization plays a bigger role for the $Al_xIn_{1-x}N$ alloys than for the two others. This being said, it is arguable that the extremes polarization charges on the figure can actually be reached, since:

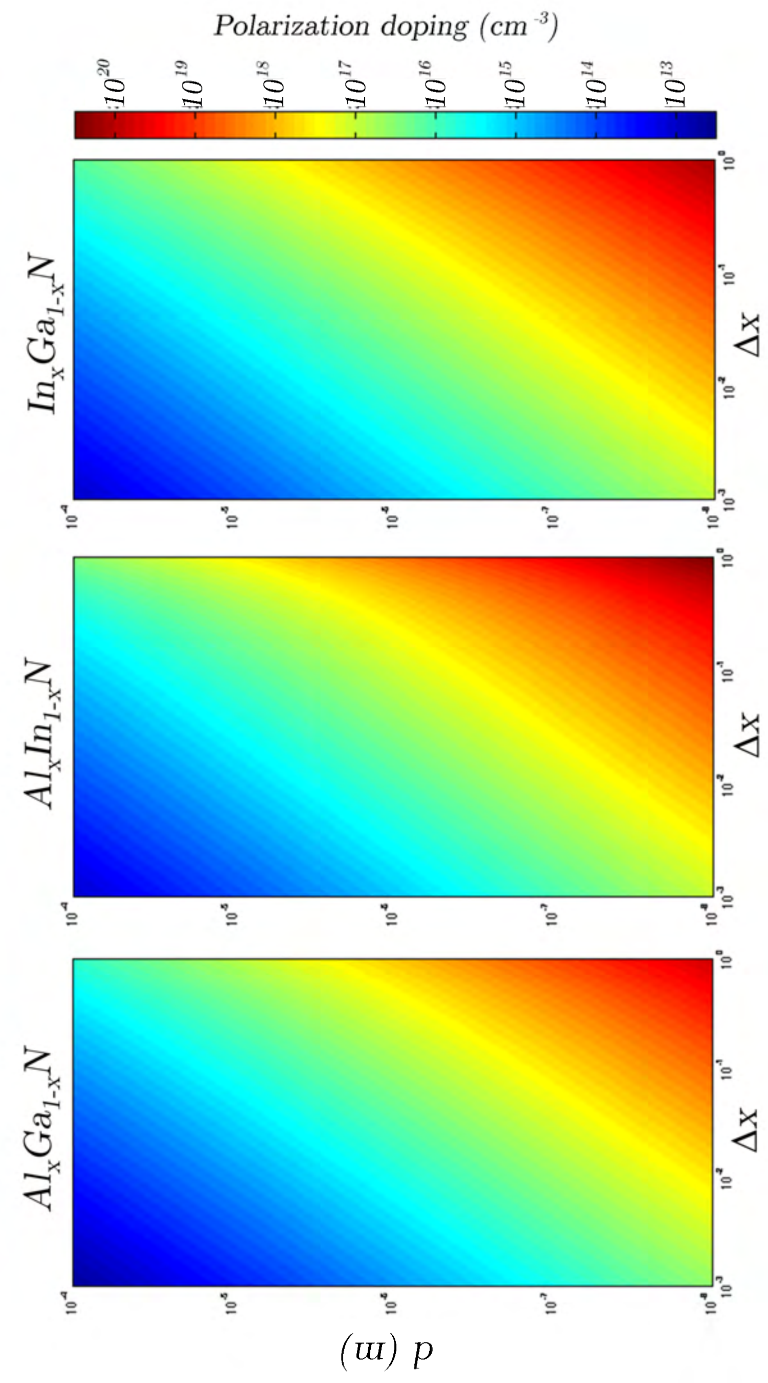


Figure 3–5: Average polarization charge obtained if the graded material corresponds to an increase of the Aluminum content by $\Delta x_{Al}$ over a distance $d$

- to reach a polarization charge of $10^{19}m^{-3}$, one would need to vary $x$ by $\Delta x = 0.001$ over a distance of $d = 100\mu m$, which would require a precision on $x$ for each layer that seems hardly achievable.
- to reach a polarization charge of $10^{26}m^{-3}$, one would need to vary $x$ by between 0 and 1 over a distance of $d = 10nm$, which would result in a considerable strain, probably causing important defects if not breaking the crystal.

Anyhow, any polarization charge between $10^{23}m^{-3}$ and $10^{25}m^{-3}$ should be easily reached, which is suitable for nanodevices of length of the order of $100nm$.

### 3.6 Comparison between the dopings

If one wants to compare the polarization-induced doping and the impurity-based doping, the main features would be the following:

- The behavior at zero temperature

  For a donor to provide a electron to the conduction band (or an acceptor to provide a hole to the valence band), the dopant needs to be ionized, and the probability of ionization follows an Arrhenius law

$$P_{ionization}(T) = A\ exp\left(-\frac{E_A}{k_B T}\right) \tag{3.12}$$

  It follows that at 0K temperature, the dopants are not ionized, since

$$\lim_{T \to 0} P_{ionization}(T) = 0 \tag{3.13}$$

and we observe what is commonly called a "freeze out" of carriers at $T = 0K$. However, for polarization-doped semiconductors, once the mobile carriers have compensated the polarization charge, they are still free to move, even at $0K$. *Polarization-doping exhibits no freeze-out at low temperatures.*

- The presence of impurities

  Introducing dopants creates impurities in the lattice, on which the carriers will scatter, reducing their ability to move, i.e. their mobility. On the contrary, no impurity is introduced in polarization-induced doping, so that *better carrier mobilities are expected.* However, the lattice mismatch between the different layers cause a strain, which, if too big, can cause dislocations and other lattice defects.

- The origin of the compensating charge

  But the major difference between the two kinds of doping is about the origin of the mobile charge. In impurity-based doping, the introduced dopant is charge-neutral: we provide the dopant to be ionized, i.e. the ionized dopant AND the charge that will be released in the band. In the initial situation, the system is charge neutral locally. On the contrary, in polarization-induced doping, we are creating a charge by modulating the polarization, and this charge needs to be compensated, i.e. to attract mobile charges of opposite sign. The initial situation is very different than that of impurity-based doping: *the semiconductor is charge-neutral overall, but definitely not neutral locally.*

To have the polarization-induced doping working, the mobile charges need to compensate the polarization charges. The question tackled in this thesis is to determine the origin of this compensating charge, necessary to reach equilibrium.

# CHAPTER 4
# Methodology: Drift-Diffusion simulations

## 4.1 The Drift-Diffusion Model

The drift-diffusion model is a widely-used model for the simulation of the motions of electrons and holes in semiconductor devices. It consists of considering the two processes deplacing the carriers : drift and diffusion.

Drift is defined as the motion of charged particles placed in an electric field **E**. If q defines the magnitude of the electronic charge, the electrons will be subjected to a $-q\mathbf{E}$ force driving them to higher-potential regions, whereas the holes will experience a $+q\mathbf{E}$ force conducting them in lower-potential regions. During their motion, the carriers experience a friction force, due to the collision events, reducing thus their speed, in the same way the friction force exerted by the air on a plane reduces its speed. The major phenomena affecting the ability of the carriers to move freely are lattice scattering, impurity (ionized or not) and defects scattering, carrier-carrier scattering and piezoelectric scattering [17, 37].

In the Drift-Diffusion model, each of these collision events are taken into account in the mobility $\mu_n$ and $\mu_p$ of the carriers, defined as the proportionality constant between the carrier velocities and the applied electric field

$$\mathbf{v}_{n,p} = \mu_{n,p}\mathbf{E} \tag{4.1}$$

We can therefore define the respective drift currents for electrons and holes as

$$\mathbf{J}_{n,drift} = -\left(-qn\mathbf{v}_n\right) = qn\mu_n\mathbf{E} \tag{4.2}$$

$$\mathbf{J}_{p,drift} = qn\mathbf{v}_p = qn\mu_p\mathbf{E} \tag{4.3}$$

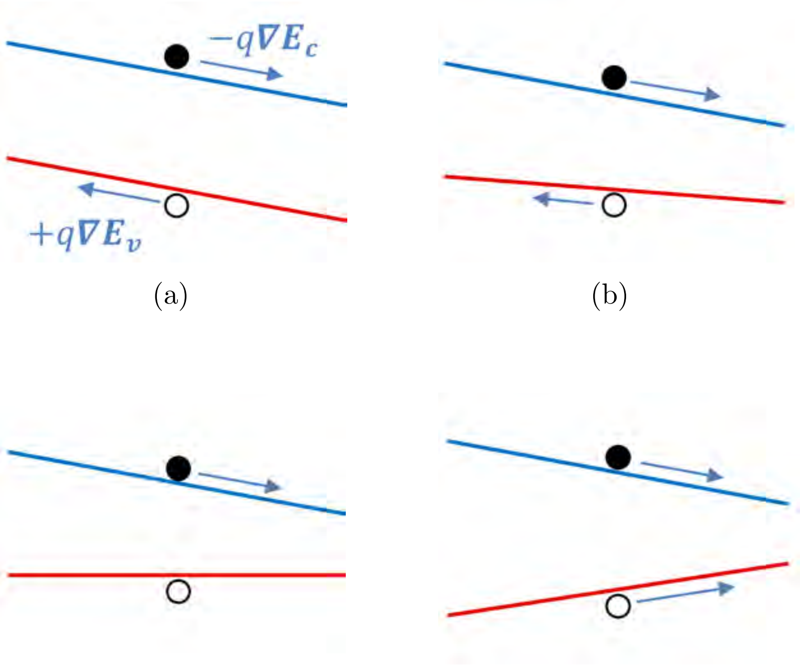
These two equations correspond to the case of a homogeneous semiconductor, where the size of the bandgap is a constant independent of position. The conduction and valence band are then parallel, with a slope equaled to the electric field.

However, in the case of a non-homogeneous semiconductor , it is possible to induce a variation of the bandgap, hence creating different slopes for the valence and conduction bands, resulting in different drift-behaviors for the electrons and the holes. Everything happens as if the electrons were subjected to a so-called quasi-electric field [38] equaled to $\nabla E_C$ and the holes to another quasi-electric field $\nabla E_V$. It is therefore possible to tune independently the magnitude of the drift for each type of carriers. As shown in Figure 4–1c it is also possible to cancel the effect on the real electric field for one type of carriers, by inducing a zero quasi-electric field, corresponding to a flat valence or conduction band. It is even possible to force the electrons and holes, despite their opposite charge, to move in the same direction, by introducing bands with slopes of different signs.

In that case, the equations for the drift currents become

$$\mathbf{J}_{n,drift} = -\left(-qn\mathbf{v}_n\right) = qn\mu_n\nabla E_C \tag{4.4}$$

$$\mathbf{J}_{p,drift} = qn\mathbf{v}_p = qn\mu_p\nabla E_V \tag{4.5}$$



(a) (b) (c) (d)

Figure 4–1: Action of the quasi-electric fields on the carriers: (a) both carriers experience forces of same amplitude but opposite directions (b) the hole experience a lower force of opposite direction (c) no force is exerced on the hole (d) electrons and holes experience forces in the same direction [6].

The second phenomenon describing the motion of carriers is the diffusion. Diffusion is a process where particles, charged or uncharged, tend to spread to reach an homogeneous concentration. It is due to the Brownian motion, to which carriers are subjected at non-zero temperature, due to collisions between these particles and the atoms and molecules of the fluid (or solid).

The process resulting in an homogeneous concentration can be understood with the help of Figure 4–2. In this experiment, the initial condition is characterized by a gradient of concentration, in a volume that we divide in 4 sub-volumes. Because of the randomness of the thermal motion, for each volume, the flux of particle exiting to the left is equaled to that to the right, resulting in an overall gain for the volumes of the first half and an overall loss for the volumes of the second half. As time tends to infinity, more and more particles migrate from the right to the left, leading a the same amount of particles in each volume. It can be shown that the flux of a particle $\alpha$ due to diffusion is

$$\mathbf{F}_{\alpha,diff} = -D_\alpha \nabla c_\alpha \tag{4.6}$$

where $c_\alpha$ is the concentration of the particles and $D_\alpha$ is known as the diffusion coefficient. This equation leads to the following expression for the current densities corresponding to the diffusion of electrons and holes

$$\mathbf{J}_{n,diff} = -\left(-qD_n \nabla n\right) = qD_n \nabla n \tag{4.7}$$

$$\mathbf{J}_{p,diff} = -qD_p \nabla p \tag{4.8}$$

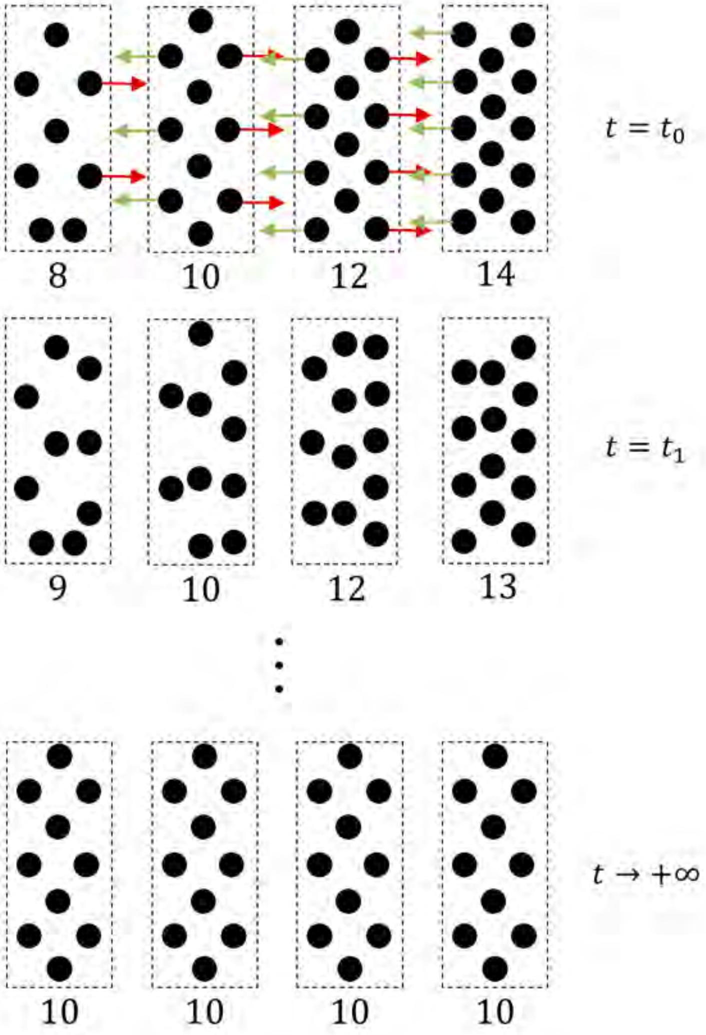


Figure 4–2: Diffusion acting on a set of particles [7]

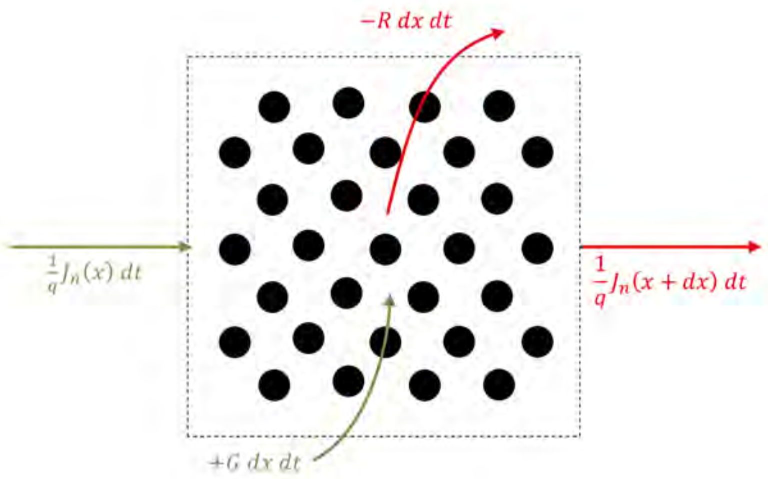


Figure 4–3: Illustration of the continuity equation

Consequently, the electron and hole currents, when submitted to both drift and diffusion, are characterized by the following equations

$$\mathbf{J}_n = \mathbf{J}_{n,drift} + \mathbf{J}_{n,diff} = qn\mu_n \nabla E_C + qD_n \nabla n \tag{4.9}$$

$$\mathbf{J}_p = \mathbf{J}_{p,drift} + \mathbf{J}_{p,diff} = qn\mu_p \nabla E_V - qD_p \nabla p \tag{4.10}$$

Drift and diffusion are the processes which govern the motion of carriers, but do not model the variation of the quantity of these carriers. For the model to be accurate, we must take into account the processes that modify the number of carriers, which have been presented in Section 2.1.3: the SRH, radiative and Auger recombinations, and their generation equivalents. Electron overflow being not in itself a generation or recombination process, it is not included the continuity equation, but it is still present in the model, as we will observe later.

If we look at the 1D problem in a segment of length $dx$, presented in Figure 4–3, we find that:

$$\begin{aligned}
(n(t+dt)-n(t))\ dx = -\frac{1}{q}\left(J_n(x)-J_n(x+dx)\right)\ dt + (G-R)\ dt\,dx \qquad (4.11)\\
\frac{\partial n}{\partial t}\,dx\,dt = \frac{1}{q}\frac{\partial J_n}{\partial x}\,dt\,dx + (G-R)\ dt\,dx \qquad (4.12)\\
\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + (G-R) \qquad (4.13)
\end{aligned}$$

Going from 1D to 3D gives us the continuity equations:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla\mathbf{J}_n + (G-R) \qquad (4.14)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla\mathbf{J}_p + (G-R) \qquad (4.15)$$

## 4.2 The time-independent algorithm

The goal of our algorithm is to determine the carrier concentrations and the potential profile, for a given doping distribution and applied voltage. First, the Poisson's equation is solved for an equilibrium situation, where no voltage is applied. Then, a loop involving both Poisson's equation and the equations of the state[1] is used to determine the carriers profiles when a voltage is applied.. The iterative method we use here is described in Figure 4–4.

---

[1] The equations of the state refers to the current equations (4.9 and 4.10) and the continuity equations (4.14 and 4.15).

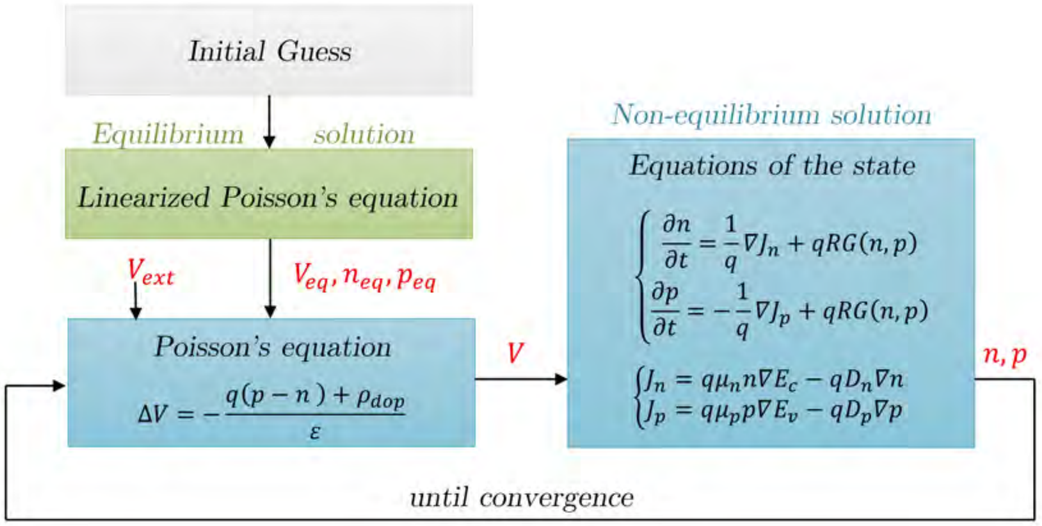


Figure 4–4: The equilibrium algorithm

Once initial conditions have been chosen for the potential $V_0$ and the carrier concentrations $n_0$ and $p_0$, we update the value of the potential corresponding to $n_0$ and $p_0$ by solving Poisson's equation.

### 4.2.1 First step: solving the equilibrium Poisson's Equation

Before beginning the loop, we must solve the Poisson's equation for an equilibrium situation, where no voltage is applied. In that case, the Fermi-levels can be taken equaled to zero, and the carrier concentrations satisfy

$$\begin{cases} n = n_i\, exp\left(\frac{-qE_i}{k_BT}\right) \\ p = n_i\, exp\left(\frac{qE_i}{k_BT}\right) \end{cases} \tag{4.16}$$

giving [2]

$$\begin{cases} n = n_i\, exp\left(\dfrac{-q(E_{i0}-V)}{k_BT}\right) \\ p = n_i\, exp\left(\dfrac{q(E_{i0}-V)}{k_BT}\right) \end{cases} \tag{4.17}$$

The Poisson's equation become then

$$\Delta V = -\frac{q(p-n)+\rho_{dop}}{\epsilon} = -\frac{q}{\epsilon}\left(n_i\, exp\left(\frac{q(E_{i0}-V)}{k_BT}\right) - n_i\, exp\left(\frac{-q(E_{i0}-V)}{k_BT}\right)\right) - \frac{\rho_{dop}}{\epsilon} \tag{4.18}$$

If we consider small corrections $V_{new} = V + \delta V$ to the potential, we can then linearize the equation 4.18 :

$$\begin{aligned} &\Delta V_{new} = \Delta V + \Delta\delta V = \\ &-\frac{q}{\epsilon}\left(n_i\, exp\left(\frac{q(E_{i0}-V-\delta V)}{k_BT}\right) - n_i\, exp\left(\frac{-q(E_{i0}-V-\delta V)}{k_BT}\right)\right) - \frac{\rho_{dop}}{\epsilon} \end{aligned} \tag{4.19}$$

---

[2] We have to take into account the quantity $E_{i0} = \chi + \frac{E_{gap}}{2}$ where $\chi$ is the electron affinity and $a$ a constant smartly chosen. This quantity takes into account the variation of the middle of the bandgap in graded structures. In a homogeneous semiconductor, there is no such variation, and we can fix this quantity to be zero.

$$\Delta V + \Delta\delta V = -\frac{q}{\epsilon}\left(n_i\, exp\left(\frac{q(E_{i0}-V)}{k_BT}\right)\left(1-\frac{q\delta V}{k_BT}\right) - n_i\, exp\left(\frac{-q(E_{i0}-V)}{k_BT}\right)\left(1+\frac{q\delta V}{k_BT}\right)\right) - \frac{\rho_{dop}}{\epsilon} \tag{4.20}$$

resulting in a linear equation for $\delta V$.

$$\Delta\delta V - \frac{q}{\epsilon}\left(n_i\, exp\left(\frac{q(E_{i0}-V)}{k_BT}\right) + n_i\, exp\left(\frac{-q(E_{i0}-V)}{k_BT}\right)\right)\delta V = -\Delta V + \frac{q}{\epsilon}\left(n_i\, exp\left(\frac{q(E_{i0}-V)}{k_BT}\right) - n_i\, exp\left(\frac{-q(E_{i0}-V-dV)}{k_BT}\right)\right) - \frac{\rho_{dop}}{\epsilon} \tag{4.21}$$

Discretizing the laplacian and inverting the matrix acting on $\delta V$ will allow to compute this correction to the potential. Iterating the process until this correction becomes negligible will lead to the equilibrium solution. With this value $V_{eq}$, we can also compute the equilibrium concentration of the carriers $n_{eq}$ and $p_{eq}$ through equations 4.17.

### 4.2.2 Second step: implementing the loop for non-equlibrium situations

Once the equilibrium potential $V$ has been updated, we can compute the non-equilibrium situation, where an external voltage is applied on the device. In that case, we implement a loop where we solve successively the equations of the state and the Poisson's equation. To solve the equations of the state, we need to use the Scharfetter-Gummel approximation, described in details in reference [39], which allows us to estimate the value of a quantity $(n, p, V, E_C, E_V, ...)$ at any point in space, even if we know their values only on the nodes of our grid. In this approximation, we have

$$\mathbf{J}_n(x) = qn\mu_n \nabla E_C + qD_n \nabla n = \mathbf{J}_n(x_{i+1/2}) + \left(x - x_{i+1/2}\right) \frac{\partial \mathbf{J}_n(x)}{\partial x} \tag{4.22}$$

The solution for n is found to be

$$n(x) = n_i(1 - g(x, E_c)) + n_{i+1} g(x, E_c) \tag{4.23}$$

$x \in [x_i, x_{i+1}]$ where g is defined as

$$g(x, E_c) = \frac{1 - exp\left(-q\frac{E_{C_{i+1}} - E_{C_i}}{kT}\frac{x - x_i}{\Delta x}\right)}{1 - exp\left(-q\frac{E_{C_{i+1}} - E_{C_i}}{kT}\right)} \tag{4.24}$$

We can then compute the quantities

$$\mathbf{J}^n_{i+1/2} = \frac{qD^n_{i+1/2}}{\Delta x}\left[n_{i+1} B\left(-q\frac{E_{C_{i+1}} - E_{C_i}}{kT}\right) - n_i B\left(q\frac{E_{C_{i+1}} - E_{C_i}}{kT}\right)\right] \tag{4.25}$$

$$\mathbf{J}^n_{i-1/2} = \frac{qD^n_{i-1/2}}{\Delta x}\left[n_i B\left(-q\frac{E_{C_i} - E_{C_{i-1}}}{kT}\right) - n_{i-1} B\left(q\frac{E_{C_i} - E_{C_{i-1}}}{kT}\right)\right] \tag{4.26}$$

where $B(x) = \frac{x}{e^x - 1}$ is referred to as the Bernoulli function.

Replacing $\nabla \mathbf{J}_n = \frac{\mathbf{J}^n_{i+1/2} - \mathbf{J}^n_{i-1/2}}{\Delta x}$ in the continuity equation, we can find the following relation between $n_{i-1}$,$n_i$ and $n_{i+1}$

$$\nabla \mathbf{J}_n = a^n_i\, n_{i-1} \; + \; b^n_i\, n_i \; + \; c^n_i\, n_{i+1} \; = \; RG_i \tag{4.27}$$

$$a_i^n = \frac{D^n_{i-1/2}}{\Delta x^2} B\left(q\frac{E_{C_i} - E_{C_{i-1}}}{kT}\right) \tag{4.28}$$

$$b_i^n = -\frac{D^n_{i+1/2}}{\Delta x^2} B\left(q\frac{E_{C_{i+1}} - E_{C_i}}{kT}\right) - \frac{D^n_{i-1/2}}{\Delta x^2} B\left(-q\frac{E_{C_i} - E_{C_{i-1}}}{kT}\right) \tag{4.29}$$

$$c_i^n = \frac{D^n_{i+1/2}}{\Delta x^2} B\left(-q\frac{E_{C_{i+1}} - E_{C_i}}{kT}\right) \tag{4.30}$$

The same approach for the holes gives us the following equation

$$\nabla \mathbf{J}_p = a_i^p \, p_{i-1} \ + \ b_i^p \, p_i \ + \ c_i^p \, p_{i+1} \ = \ RG_i \tag{4.31}$$

$$a_i^p = \frac{D^p_{i-1/2}}{\Delta x^2} B\left(-q\frac{E_{V_i} - E_{V_{i-1}}}{kT}\right) \tag{4.32}$$

$$b_i^p = -\frac{D^p_{i+1/2}}{\Delta x^2} B\left(-q\frac{E_{V_{i+1}} - E_{V_i}}{kT}\right) - \frac{D^p_{i-1/2}}{\Delta x^2} B\left(q\frac{E_{V_i} - E_{V_{i-1}}}{kT}\right) \tag{4.33}$$

$$c_i^p = \frac{D^p_{i+1/2}}{\Delta x^2} B\left(q\frac{E_{V_{i+1}} - E_{V_i}}{kT}\right) \tag{4.34}$$

The equations 4.27 and 4.31 can therefore be solved by a simple matrix inversion:

$$\begin{cases} n = {M_n}^{-1} \, RG \\ p = {M_p}^{-1} \, RG \end{cases} \tag{4.35}$$

where the matrices $M_{n,p}$ are defines as.

$$M_{n,p}\left(E_C, E_V\right) = \begin{pmatrix} b_1^{n,p} & c_1^{n,p} & 0 & 0 & 0 & & & & \\ a_2^{n,p} & b_2^{n,p} & c_2^{n,p} & 0 & 0 & \ddots & & & \\ 0 & a_3^{n,p} & b_3^{n,p} & c_3^{n,p} & 0 & \ddots & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & & \\ & & & & & & & & \\ & & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & & 0 & a_{N-2}^{n,p} & b_{N-2}^{n,p} & c_{N-2}^{n,p} & 0 \\ & & & & & 0 & a_{N-1}^{n,p} & b_{N-1}^{n,p} & c_{N-1}^{n,p} \\ & & & & & & 0 & a_{N-1}^{n,p} & b_{N-1}^{n,p} \end{pmatrix} \tag{4.36}$$

Solving next the Poisson's equation is simpler, since one only need to invert the Laplacian to get the updated value of the potential.

### 4.3 Verification on an example

To verify that the algorithm described above is working, we model a device for which we already know the behavior. If the code gives us the expected behavior, we can assume that it might work well.

The example chosen here is the pn-junction, explained succinctly in Section 2.1.1. In a few words, the juxtaposition in a region rich in electrons and a region rich in holes results in the formation of a depletion region, where ionized dopants are uncompensated. A potential barrier arises because of these charges, preventing the further diffusion of the carriers to the opposite side.

When a negative voltage is applied to the device, the potential barrier is increased, and the carriers are less numerous in the depletion region than in the equilibrium situation. This situation is well depicted by our simulated device in Figures 4–5 (a,b).

When a positive voltage is applied, the potential barrier in the bands is reduced, allowing the electrons to penetrate in the p-region and the holes in the n-region. This situation is also depicted in Figures 4–5 (a,b), and our simulation matches the expected behaviour. This diffusion of majority carriers allows a current to circulate in the device, and it can be shown [17] that this current is

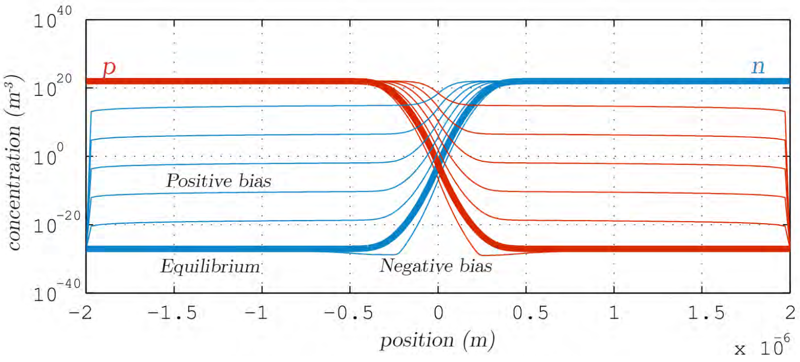
$$I = I_0 \, exp\left(\frac{qV}{k_B T}\right) \tag{4.37}$$

This linear relation between the logarithm of the simulated current and the voltage exhibits a constant of proportionality equaled to $38.9V^{-1}$ in Figure 4–5 (c), which is really close to the expected $\frac{q}{k_B T} = 38.7V^{-1}$.

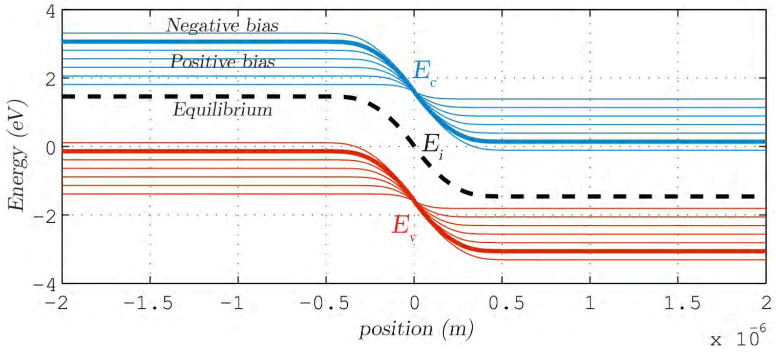
We can therefore assume that our algorithm and code are working.

### 4.4 The time dependent algorithm

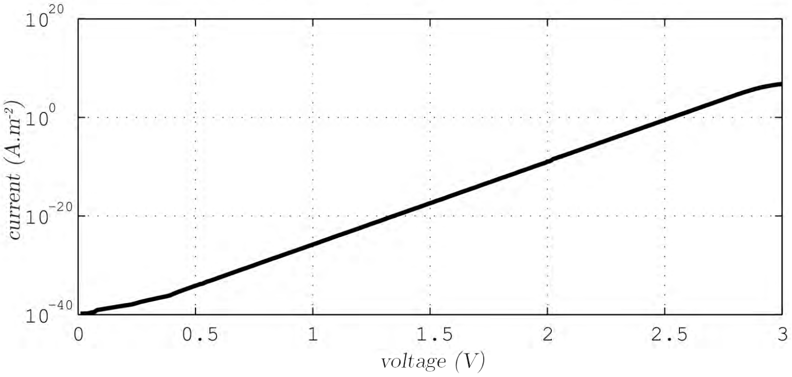
As we have seen, one question to be addressed is the origin of the electrons and holes compensating the polarization charge. To be able to determine this origin, we use a time dependent algorithm whose principle is explained in Figure 4–6. At each step of time, the potential corresponding to the given carrier densities $n(t)$ and $p(t)$ is calculated. By solving the equations of the state, we compute the values $n(t+\Delta t)$ and $p(t+\Delta t)$, using either an explicit scheme or an implicit scheme.



*(a) Electron and Hole concentrations*



*(b) Valence and Conduction bands*



*(c) I-V curve*

Figure 4–5: Simulated behavior of a pn-junction

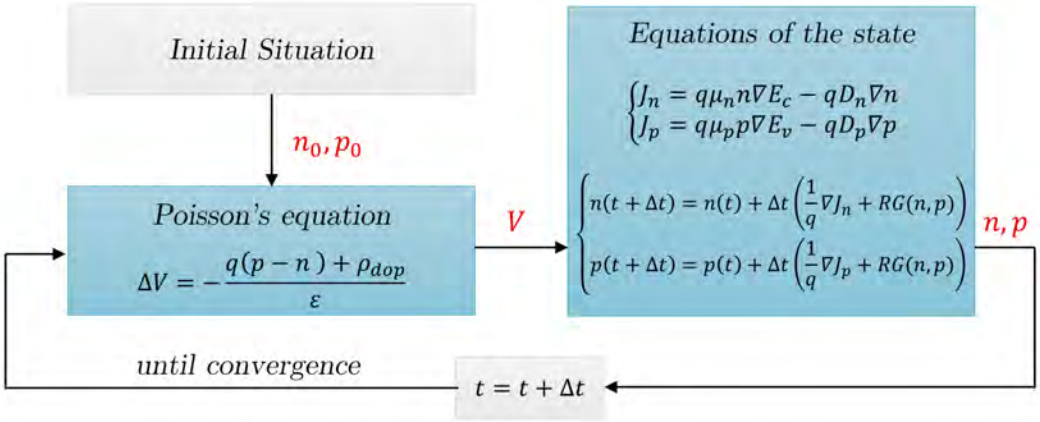


Figure 4–6: The time-dependent algorithm

For the explicit scheme, the equations of the state becomes:

$$n_{i,t+\Delta t} = n_{i,t} + \Delta t \left(a_i^n n_{i-1,t} + b_i^n n_{i,t} + c_i^n n_{i+1,t}\right) + \Delta t\; RG \tag{4.38}$$

$$p_{i,t+\Delta t} = p_{i,t} + \Delta t \left(a_i^p p_{i-1,t} + b_i^p p_{i,t} + c_i^p p_{i+1,t}\right) + \Delta t\; RG \tag{4.39}$$

which can be put under the form

$$n_{t+\Delta t} = \left(I + \Delta t M_n\right) n_t \;+\; \Delta t\; RG \tag{4.40}$$

$$p_{t+\Delta t} = \left(I + \Delta t M_p\right) p_t \;+\; \Delta t\; RG \tag{4.41}$$

However, this explicit scheme is highly unstable and requires very small time steps, which make the calculation time too long to be carried out. This is why we have chosen to use an implicit scheme, in which the equations of the state become

$$n_{i,t+\Delta t} = n_{i,t} + \Delta t \left(a_i^n n_{i-1,t+\Delta t} + b_i^n n_{i,t+\Delta t} + c_i^n n_{i+1,t+\Delta t}\right) + \Delta t\; RG \tag{4.42}$$

$$p_{i,t+\Delta t} = p_{i,t} + \Delta t \left(a_i^p p_{i-1,t+\Delta t} + b_i^p p_{i,t+\Delta t} + c_i^p p_{i+1,t+\Delta t}\right) + \Delta t\; RG \tag{4.43}$$

or

$$n_{t+\Delta t} = (I - \Delta t M_n)^{-1} (n_t \ + \ \Delta t \, RG) \tag{4.44}$$

$$p_{t+\Delta t} = (I - \Delta t M_p)^{-1} (p_t \ + \ \Delta t \, RG) \tag{4.45}$$

As we have seen earlier, the question of the origin of the compensating charges is of prime importance. To be able to determine without any doubt this origin, we define two functions $f_n$ and $f_p$ which attribute to each position $x$ a distribution function of the origin of the carriers at $x$ at time $t$. The quantity $\int_{y_1}^{y_2} f_n(x,y,t) \, \mathrm{d}x$ is therefore the probability for the electrons at position $x$ at time $t$ to come from the segment $[y_1, y_2]$. Since we are working on a discretized space, we prefer using the discretized functions $f_n$ and $f_p$ where $f_n(i,j,t)$ refers to the probability for an electron at site $i$ at time $t$ to originate from site $j$. Therefore the quantity $n_{i,t} f_n(i,j,t)$ refers to the number of electrons at site $i$ at time $t$ which were at site $j$ at time $t = 0$, and the evolution of this quantity (in the implicit scheme) is given by

$$\begin{aligned} \frac{n_{i,t+\Delta t} f_n(i,j,t+\Delta t) - n_{i,t} f_n(i,j,t)}{\Delta t} &= a_i^n n_{i-1,t+\Delta t} f_n(i-1,j,t+\Delta t) \\ + \ b_i^n n_{i-1,t+\Delta t} f_n(i,j,t+\Delta t) \ &+ \ c_i^n n_{i+1,t+\Delta t} f_n(i+1,j,t+\Delta t), \\ &\forall i,j \in [1,N]^2 \end{aligned} \tag{4.46}$$

which gives the matrix equation for a given $j$

$$\begin{pmatrix} f_n(1,j,t+\Delta t) \\ f_n(2,j,t+\Delta t) \\ \vdots \\ \\ f_n(i,j,t+\Delta t) \\ \\ \vdots \\ f_n(N-1,j,t+\Delta t) \\ f_n(N,j,t+\Delta t) \end{pmatrix} = \frac{1}{\Delta t} O_{n,j}(t)^{-1} \begin{pmatrix} f_n(1,j,t) \\ f_n(2,j,t) \\ \vdots \\ \\ f_n(i,j,t) \\ \\ \vdots \\ f_n(N-1,j,t) \\ f_n(N,j,t) \end{pmatrix} \tag{4.47}$$

where

$$O_{n,j}(t) = \begin{pmatrix} \left(\frac{1}{\Delta t} - b_1^n\right) \frac{n_{1,t}}{n_{1,t+\Delta t}} & c_1^n \frac{n_{2,t}}{n_{1,t+\Delta t}} & & & \\ a_2^n \frac{n_{1,t}}{n_{2,t+\Delta t}} & \left(\frac{1}{\Delta t} - b_2^n\right) \frac{n_{2,t}}{n_{2,t+\Delta t}} & c_2^n \frac{n_{3,t}}{n_{2,t+\Delta t}} & & \\ \ddots & \ddots & & \ddots & \\ & a_i^n \frac{n_{i-1,t}}{n_{i,t+\Delta t}} & \left(\frac{1}{\Delta t} - b_i^n\right) \frac{n_{i,t}}{n_{i,t+\Delta t}} & c_i^n \frac{n_{i+1,t}}{n_{i,t+\Delta t}} & \\ & \ddots & \ddots & & \ddots \\ & & a_{N-1}^n \frac{n_{N-2,t}}{n_{N-1,t+\Delta t}} & \left(\frac{1}{\Delta t} - b_{N-1}^n\right) \frac{n_{N-1,t}}{n_{N-1,t+\Delta t}} & c_{N-1}^n \frac{n_{N,t}}{n_{N-1,t+\Delta t}} \\ & & & a_N^n \frac{n_{N-1,t}}{n_{N,t+\Delta t}} & \left(\frac{1}{\Delta t} - b_N^n\right) \frac{n_{N,t}}{n_{N,t+\Delta t}} \end{pmatrix} \tag{4.48}$$

We notice that the matrix is independent of $j$, we can create the matrix $F_n(t) = (f_n(i,j,t))_{(i,j)\in[1,N]^2}$. Since multiplying a matrix on the left acts on its columns, we can compute with only one single matrix inversion the updated values of the origin profiles.

$$F_n(t + \Delta t) = \frac{1}{\Delta t}\, O_n(t)^{-1}\, F_n(t) \tag{4.49}$$

Similarly, we can create the matrix $O_p(t)$ by replacing all the quantities $X_n$ in $O_n(t)$ by their equivalent $X_p$ such that we have

$$F_p(t + \Delta t) = \frac{1}{\Delta t}\, O_p(t)^{-1}\, F_p(t) \tag{4.50}$$

# CHAPTER 5
# The Origin of the compensating charges

As we have seen earlier, one major difference between impurity-based doping and polarization-induced doping is that dopants are introduced with their compensating charge, while the polarization charge is not compensated when the device is created. To understand how polarization-doping works and why it does not work in some configurations, it is fundamental to know where the compensating charge comes from. We will try to simulate different configurations to explore the compensation mechanisms. In all the following examples, we try to induce p-doping, but the same reasoning should be applied to n-doping. These simulations were realized through Matlab via the algorithms described in the preceding section.

## 5.1 Design 1: growth on a reservoir of charges

The first example that comes to mind is to create the p-side of a pn-junction via a gradient of polarization. The polarization charge $\rho_{Pol} = -\nabla \mathbf{P}$ needs to be negative in order to attract the positively charged holes. In this example, we want to confirm the intuitive idea that if we put the graded part next to a reservoir of holes (a highly p-doped semiconductor, a metal,...), the holes compensating the polarization charge will come from this reservoir. In this simulation, we use a small bandgap material (comparing to GaN and AlN), for numerical reasons: the time to reach equilibrium turns out to be much smaller. It corresponds to a N-face grading from $In_{x_1}Ga_{1-x_1}N$ to $In_{x_2}Ga_{1-x_2}N$ with $x_1 < x_2$. The experiment is the following: we start from our reservoir, which also acts as a substrate, and we try to reproduce what happens during a growth by Molecular Beam Epitaxy, i.e. the aggregation of layers one onto the other. The outline is the following:

- at time $t_0$ the reservoir is introduced;
- at time $t_i$, we add the $i^{th}$ layer of material and we let the system reach equilibrium. All significant quantities, such as $n, p, E_V, E_i, E_c, V$ are recorded for different values of $t$ satisfying $t_{i-1} \leq t \leq t_i$. Once equilibrium is reached [1] , the $i+1^{th}$ layer is introduced.

The evolution of the holes and electrons is presented on Figure 5–1, while the band diagrams is presented on Figure 5–2. In both figure, the $i^{th}$ figure represents the time evolution of the system after the introduction of the $i^{th}$ layer: the lightest curve represent the earliest moments after time $t_i$ and the curves become darker as time goes by. The equilibrium situation is represented by the thickest line.

In each diagram, the reaching of equilibrium can be decomposed in three steps:

1. Drift of the carriers

   At time $t_i$, the $i^{th}$ layer is deposited, with an intrinsic concentration of electrons and holes : $n = p = n_i$. The introduction of the layer changes the location of the semiconductor-air interface, and therefore the location of the surface charge at this interface. At time $t_i$, however, there is still a strong concentration of holes at the interface between the layers $i-1$ and $i$: this strong surface charge induces a

---

[1] The deposition rate of each layer is long $(7,5\mathring{A}s^{-1})$ comparing to the time required to reach an equilibrium situation (between $10^{-12}$ and $10^{-9}s$ for each layer).

strong change of slope in the bands. This slope, corresponding to a quasi-electric field, induces a drift of the holes to the right and a drift of the electrons to the left. As the holes drift from the left to the right, the charge at the interface reduces, flattening the bands[2] , and the polarization charge in the $i^{th}$ layer is compensated.

2. Diffusion of the electrons

   The electrons introduced in the $i^{th}$ layer are in a much higher concentration than on the left. Therefore, a diffusion motion from the left to the right takes place.

3. Recombination

   Once the electrons have finished diffusing, they recombine with the holes to reach the equilibrium situation, where $n_{eq}\, p_{eq} = n_i^2$. This recombination involves the same number of electrons and holes: however, since the holes are in a much higher concentration than the electrons, and the recombination is limited by the number of available electrons, this hole concentration does not quantitatively change during the process.

Once this p-side is created, we can add the n-side to create the pn-junction. This n-side could be created via polarization-doping too, but since there is no problem of donor ionization, we just consider here an homogeneous n-side made with donors.

---

[2] The valence band is almost flat here, because $\rho_{Pol}$ is almost constant, which means that at equilibrium, $p_{eq} = -\rho_{Pol}$ is almost constant; in the Boltzmann statistics, if $p$ is independent of position, then so is $E_V$.

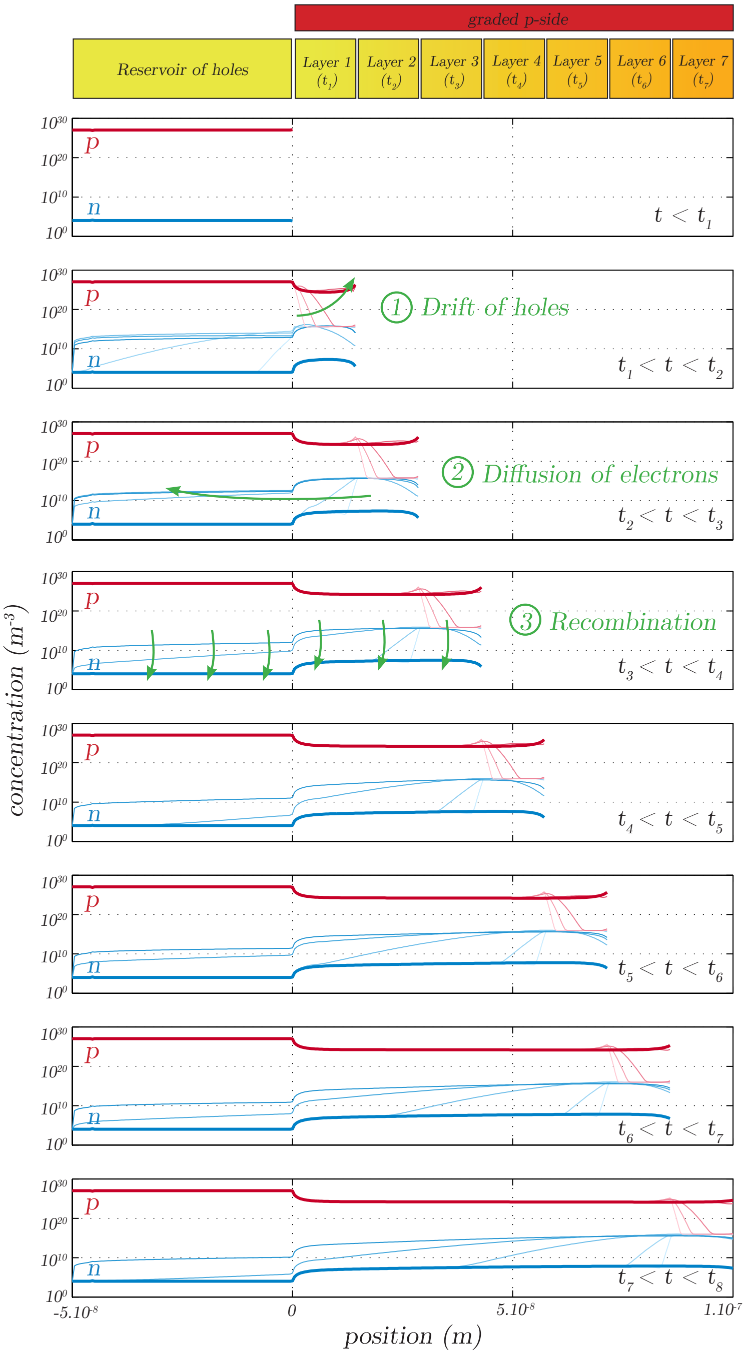


Figure 5–1: P-side of Design 1, evolution of the carrier concentrations

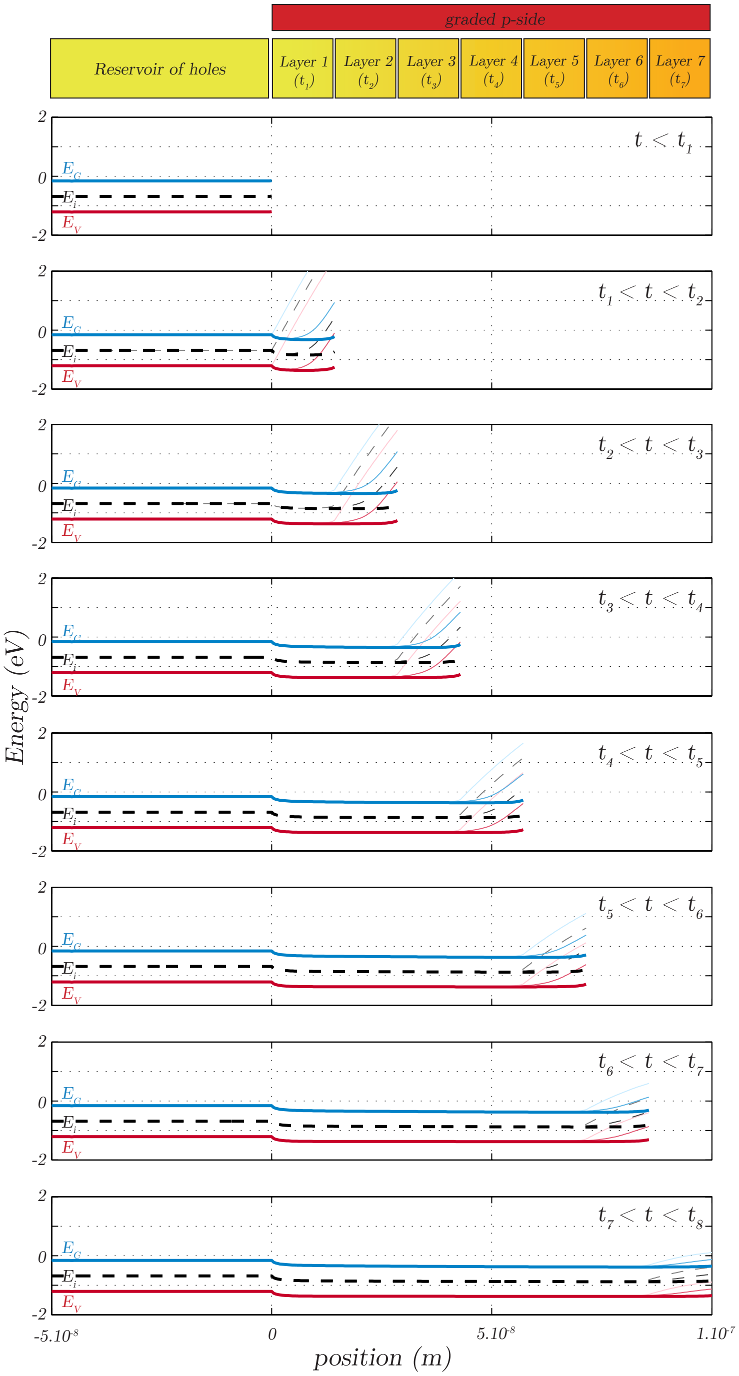


Figure 5–2: P-side of Design 1, evolution of the band structure

As the layers $(i)_{8 \leq i \leq 14}$ are introduced, the evolution of the system is characterized by:

- a diffusion of the electrons from the n-side to the p-side;
- a diffusion of the holes from the p-side to the n-side;
- the recombination of the carriers, which results in the formation of the depletion region, and therefore of the built-in potential.

It should be noted that in our case, there is a depletion region at the end of the n-side. Indeed, there is a strong negative surfacic charge resulting from the semiconductor-air interface, forcing the electrons to leave this region. This is the issue talked about previously: there is positive polarization charge on the n-side and a negative one on the p-side. This could induce some problems in terms of injection of carriers.

One might argue that it is difficult to find a good reservoir of holes for some materials. It is the case for materials like AlN and GaN, since their large bandgap would require a large work-function metal to create an ohmic p-contact, i.e. a contact that efficiently inject holes in the device, and it is almost impossible to use acceptors to create the reservoir, since their percentage of ionization is very low. A solution to this problem consists in replacing the p-contact by an n-contact and a tunnel diode: in that case, the holes are created by the tunneling of the valence electrons into the conduction band through the tunnel-diode. The band structure of this device is presented in Figure 5–5.

What happens now if the n-side is grown first, and the graded p-side is then added on top of it? Some articles show that in the case of nanowires, the compensation

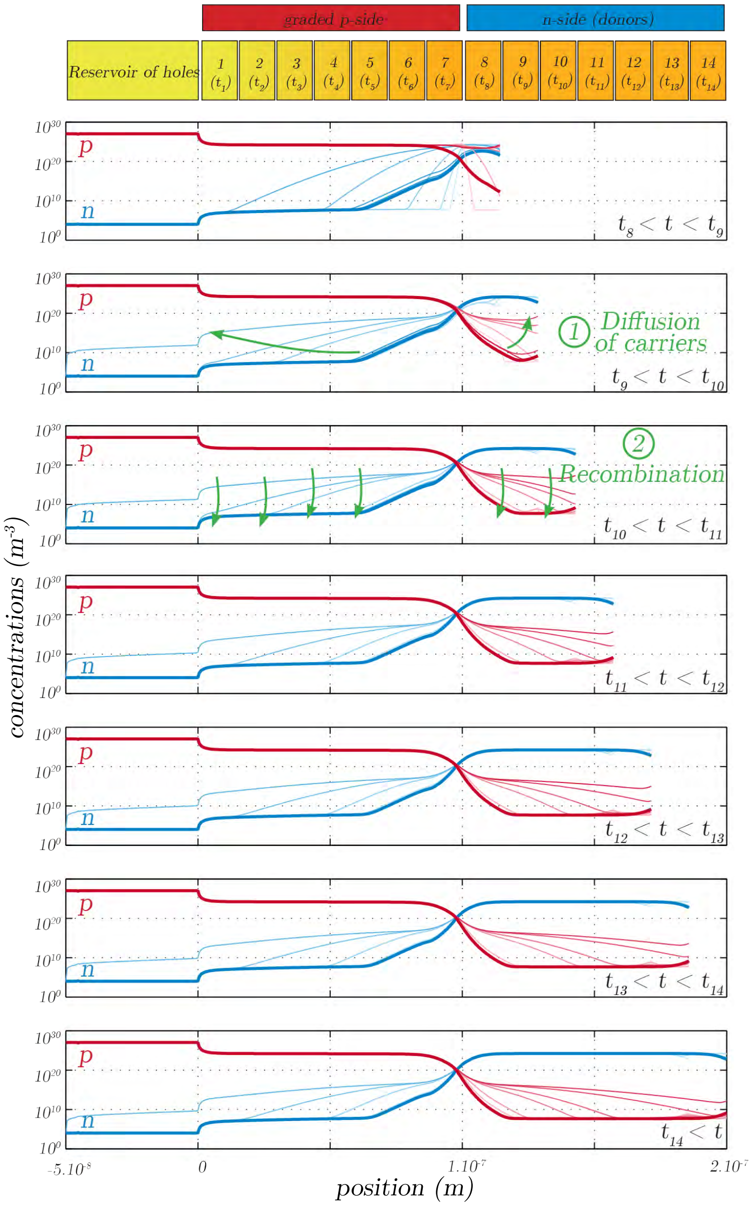


Figure 5–3: Design 1, evolution of the electron and hole concentrations

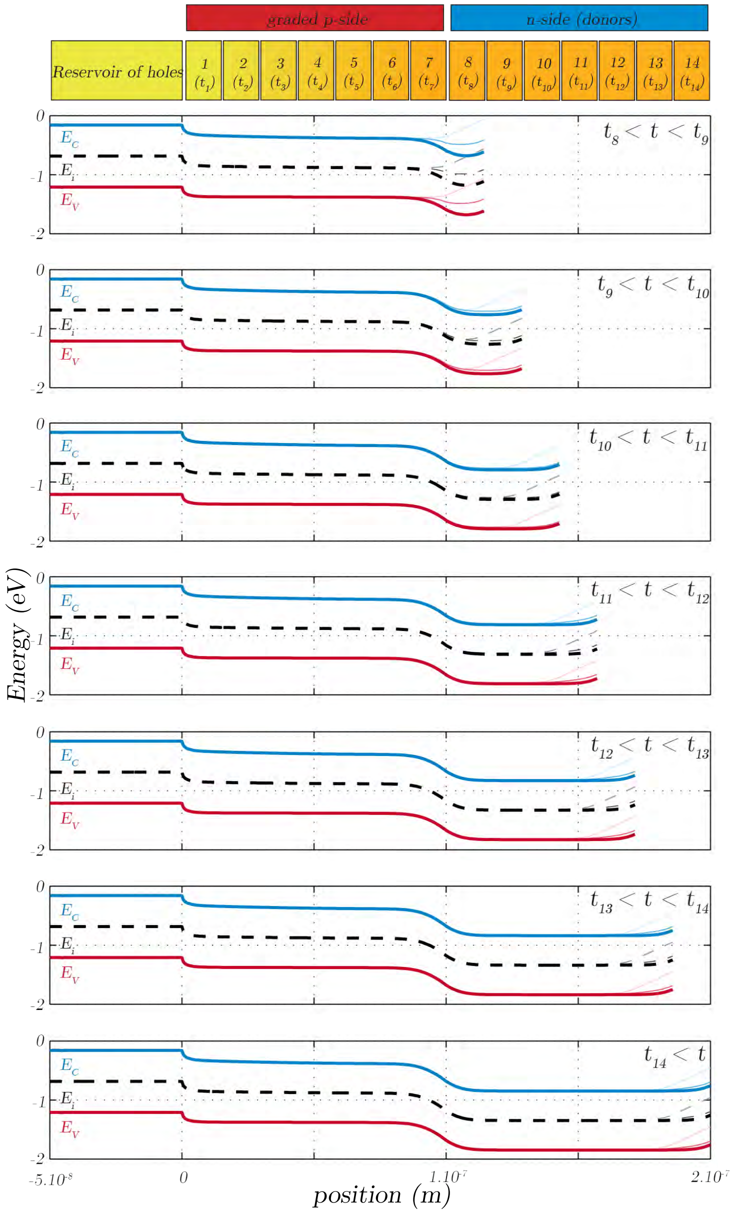


Figure 5–4: Design 1, evolution of the band structure

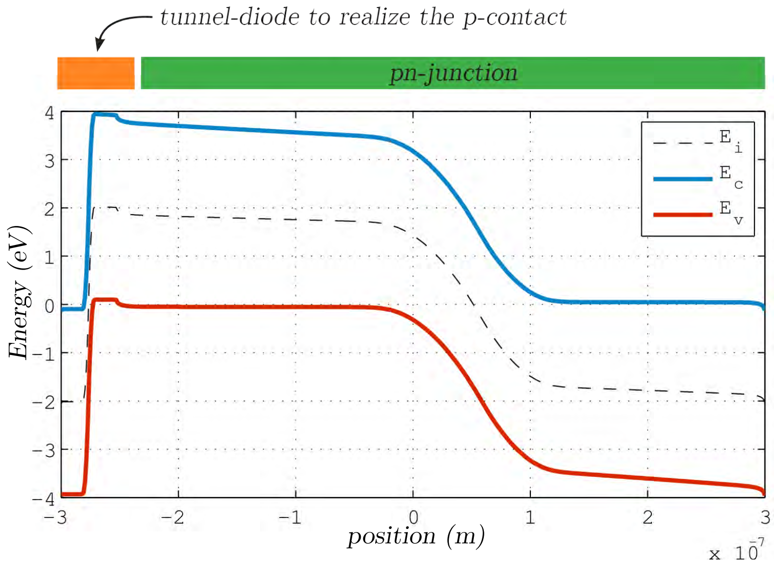


Figure 5–5: The use of a tunnel-diode to realize the p-contact

of the polarization charge occured, creating the doping. This might be because the necessary quantity of holes is quite small (because linked to the volume of the nanowire): these holes are said to come from acceptor-like states on the surface of the nanowire. However, on a bulk device, the large volume (compared to that of the nanowire) results in a large required quantity of holes, which cannot be provided by surface-states. In that case, where do they come from?

## 5.2 Design 2: growth without reservoir

If the graded p-side is grown on the n-side, then there is no holes to compensate the polarization charge. Our simulation here consists of adding the layers corresponding to the graded p-side on the n-side, in the same way as above. The curves are presented in Figures 5–6 and 5–7. The important features of this process are the following:

- the first layers introduce a negative polarization charge, which will not be compensated, since there is not enough holes in the device. As a consequence, the bands are bending downwards.
- as a new layer is introduced, the bending becomes larger than the bandgap, and empty states in the conduction band are at a lower energy that filled states in the valence band. If the distance between these states is small enough, the valence electrons will tunnel through the bandgap[3] , leaving holes in the valence band and electrons in the conduction band. These holes will drift to the top of the valence band and screen the polarization charge: the electric field and the band bending will be reduced. This process goes on until no tunneling can occur (either the tunneling distance is too big, or no conduction empty state is at a lower energy than a filled valence state).
- since as many electrons as holes are created, and since the bands are bending downwards, there is a high concentration of electrons on the extreme-right region, which might result in a n-p-n device, i.e. the juxtaposition of two diodes in opposite directions. However, the introduction of a good contact will remove these electrons, as shown in the last diagram of Figure 5–6.

However, since the relation between the electric field resulting from the uncompensated charge, the bandgap and the tunneling distance is purely trigonometric,

---

[3] The calculation details of this current is explained in Appendix B.

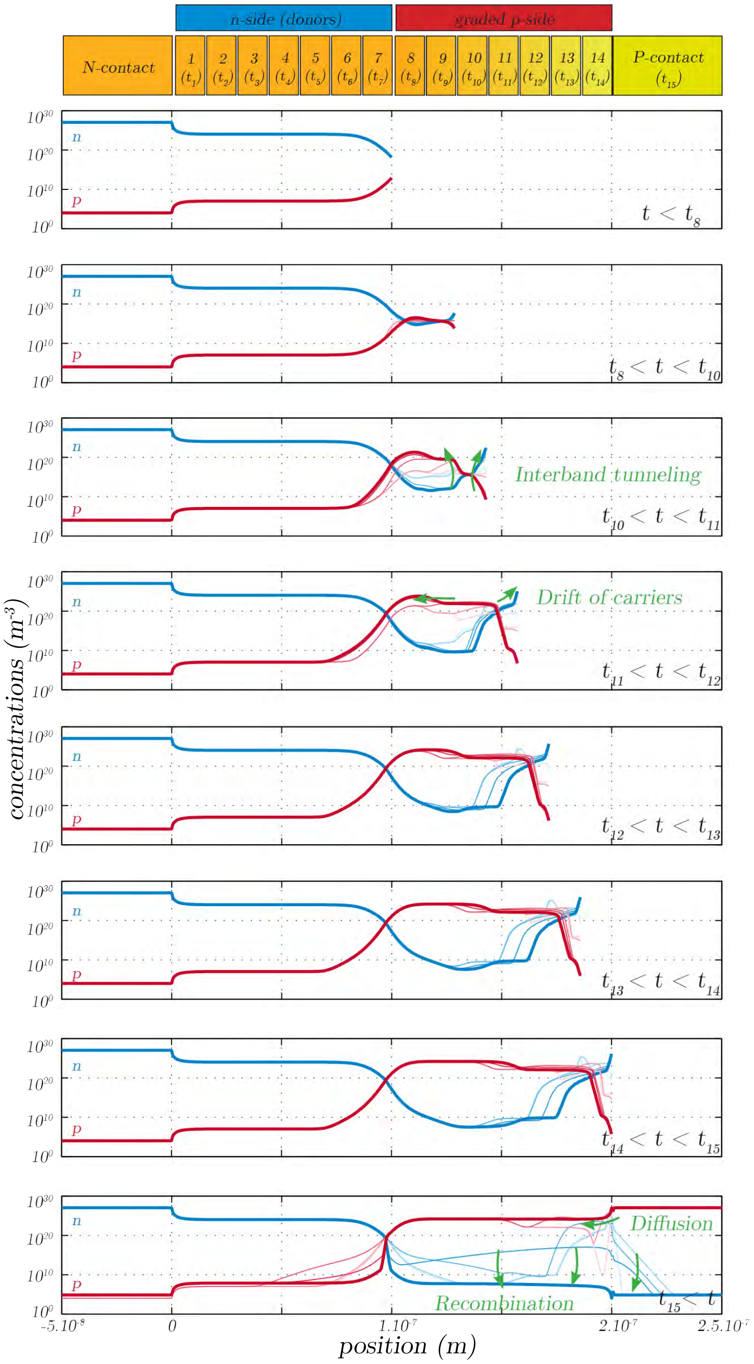


Figure 5–6: Design 2, evolution of the carrier concentrations

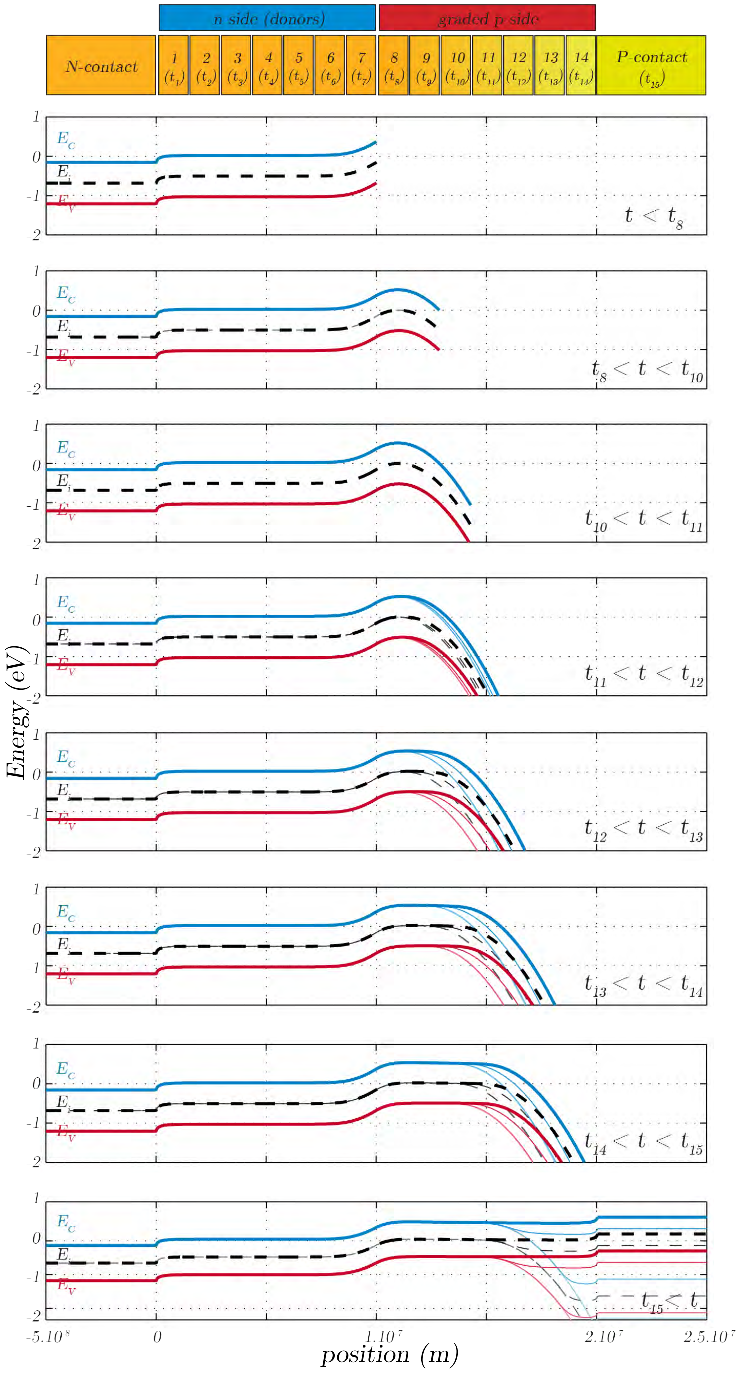


Figure 5–7: Design 2, evolution of the band structure

using wide-bandgap semiconductor might reduce or even prevent the interband tunneling. The wide-bandgap semiconductors are indeed known for their ability to sustain high electric fields. But once again, the introduction of a good p-contact (or p-reservoir) should result in the diffusion of the holes to compensate the polarization charge.

This is where our simulations do not match the experimental results presented in reference [40]. This study reported the realization of an LED, where the p-side was realized via polarization-induced doping. However, they reported that they had to introduce acceptors in the lattice for the doping to work[4] , suggesting between the lines that without acceptors, the polarization-doping was not working. We will see in the next section why acceptors can help polarization-induced doping, but in the meantime, as far as our simulations are concerned, their design should have worked without acceptors. The only reasons that could explain their result are:

- a bad p-contact, resulting in no transfer of holes to the graded p-side;
- the strong positive surface charge resulting from the AlGaN/contact interface: indeed, they grew their device N-face from GaN to AlGaN, which is the "bad" choice, since the strong polarization charge at the semiconductor-air interface on the n-side is negative, and the one on the p-side is positive (as shown in Figure 3–4a).

---

[4] The very point of polarization-induced doping is to avoid the use of acceptors!

## 5.3 Design 3: the introduction of acceptors

The paper [40] mentioned above suggested to introduce acceptors to provide the holes to the valence band, which would then compensate the polarization charge. As said above, the very point of turning to polarization-doping is to avoid the inefficiency in the acceptor ionization, and we will try to show how this method is still highly inefficient. As explained in Appendix A, in the presence of a strong (pseudo-)electric field, the ionization of acceptors can be greatly enhanced, and it might be possible to provide a sufficient quantity of holes to the valence band, to compensate the polarization charge.

Two things should be highlighted here:

- an acceptor can only be ionized in the presence of an electric field;
- there is no overall charge reduction: a hole coming from an acceptor to compensate a polarization charge has left an ionized acceptor behind, which carry a negative charge. It is all about displacing charges.

Figures 5–8 and 5–9 represent the growth of a device, where the donor-based n-side has been grown first, and the graded p-side has been added then. Acceptors are introduced in the p-side as well, to provide the necessary holes. At times $(t_i)_{8\leq i\leq 14}$, the $i^{th}$ layer is introduced:

- the presence of a strong electric field resulting from the previous charge configuration results in a complete ionization of all the acceptors: a lot of holes are present in the $i^{th}$ layer shortly after time $t_i$. Then the electric field induces a drift of the holes to fill the depletion region of the previous layer, resulting in a reduction of the electric field. As the electric field is reduced, less and less acceptors are ionized, and holes from the valence band are going back to the acceptor sites. Finally, an equilibrium situation is reached where a depletion region at the end of the $i^{th}$ layer is created: in this region, all the acceptors are ionized and their holes moved to the left to screen the polarization charge. The diagrams on the right in Figure 5–8 represent the evolution of the percentage of ionized acceptors as a function of time.

- as for the previoux experiment, the holes diffuse to the n-side and the electrons diffuse to the p-side.
- the carriers recombine to eliminate their surplus.

It should be noted that this technique is highly inefficient, since only the acceptors in the last depletion region are ionized. Moreover, there should be a high quantity of acceptors in this depletion region to be able to provide holes to the rest of the p-side. Finally, this depletion region can constitute a significant barrier for the hole injection, and therefore only a very high quantity of acceptors can enable tunneling through this barrier to have a good p-contact.

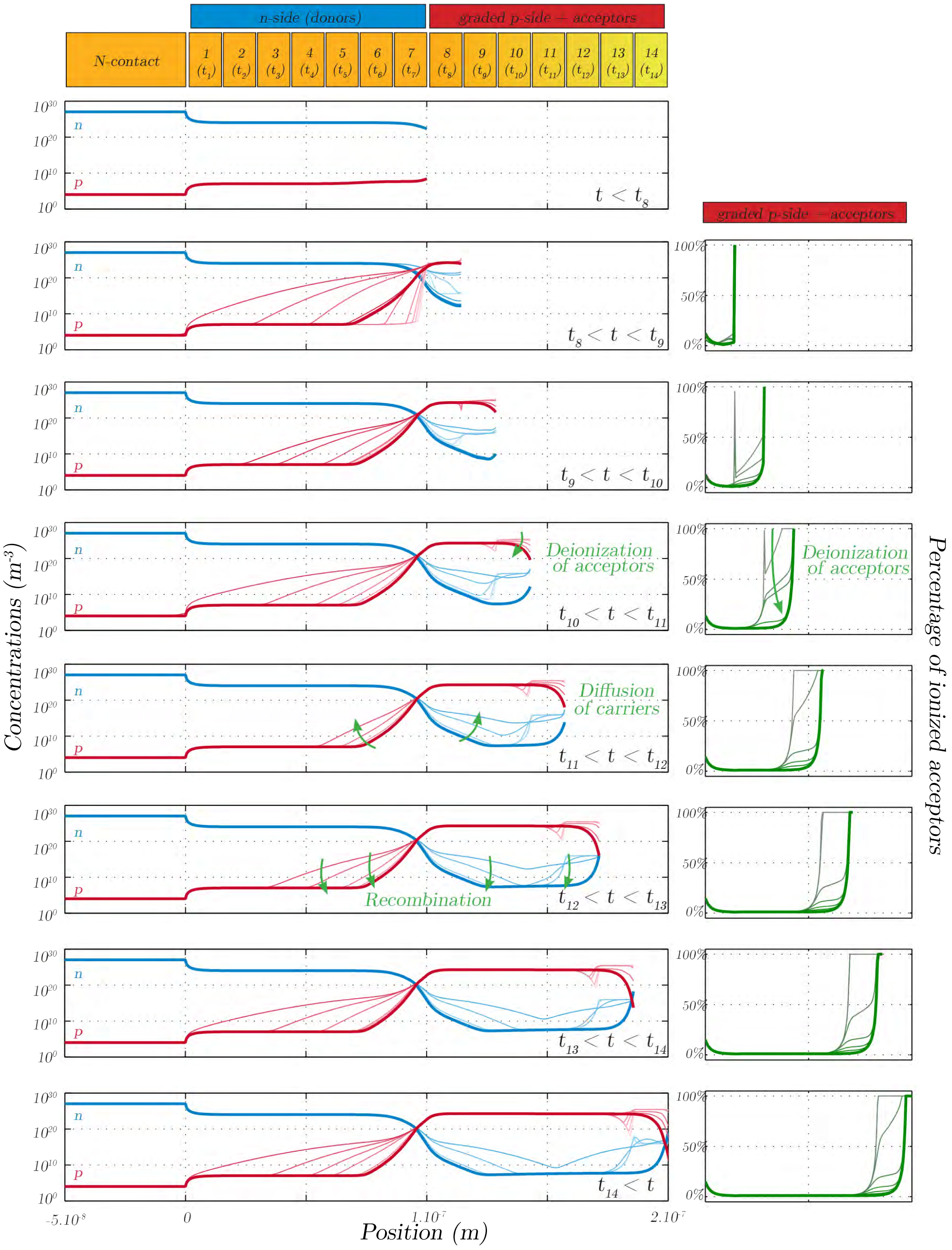


Figure 5–8: Design 3, evolution of the carrier concentrations

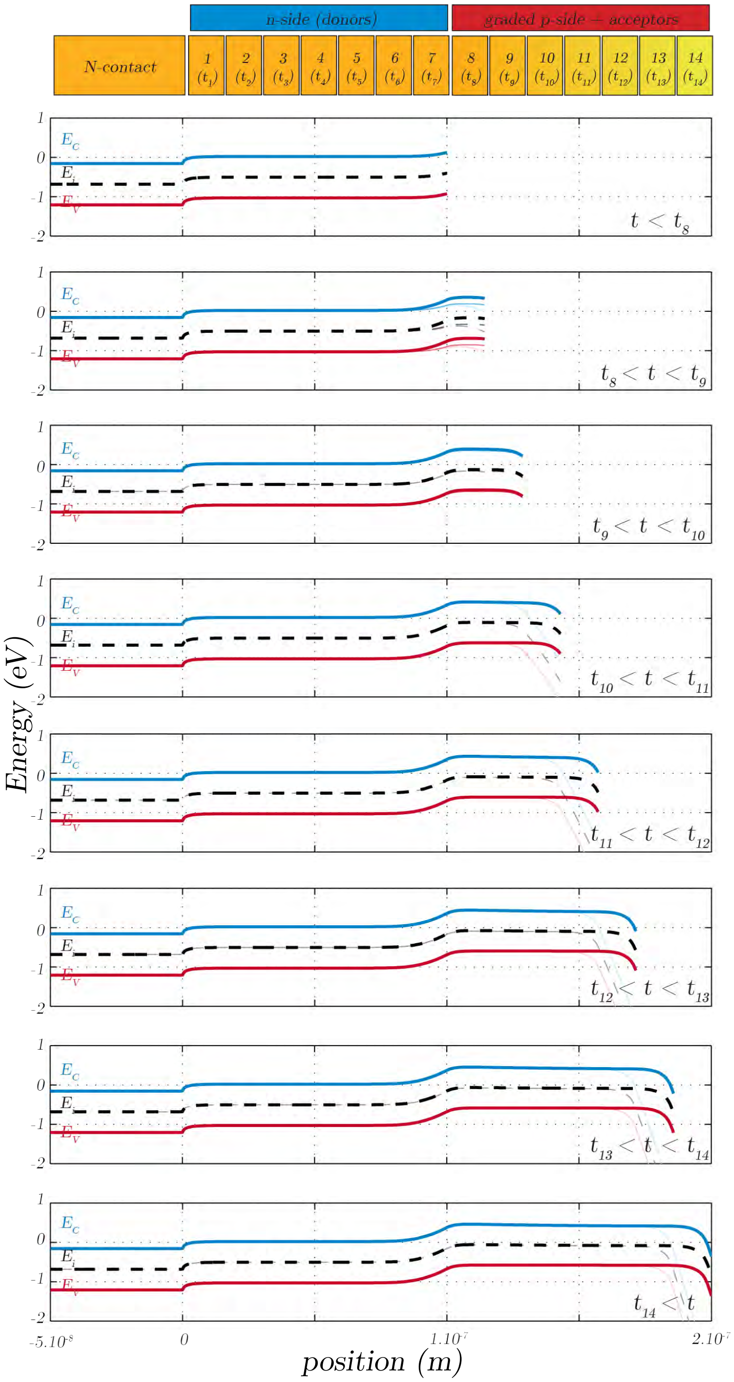


Figure 5–9: Design 3, evolution of the band structure

# CHAPTER 6
# Features of a polarization-induced LED

## 6.1 Pertinence of the problem

The idea of comparing a classic LED, made of GaN, with a polarization-induced LED, is tempting, but requires further thinking. Indeed, the two systems are not made of the same material, do not present the same composition, which makes it very difficult to predict the behavior. We will study here three devices:

- the first one is a simple p-n junction made of GaN, and its band structure is presented in Figure 6–1(a). This band structure is the one of a classic p-n junction.
- the second one is a p-n junction, whose p-side is graded [1] , and its band-structure is presented in Figure 6–1(b). In this band structure, the valence band is flat because the polarization charge is constant in this region, resulting in a constant concentration of holes. These holes obey to the Boltzmann statistics: a constant concentration is equivalent to a constant energy. The valence band is therefore flat. Because of the grading, the bandgap increases on the p-side as we move away from the active region, resulting in a slope for the conduction band, which can then act as an Electron Blocking Layer.

---

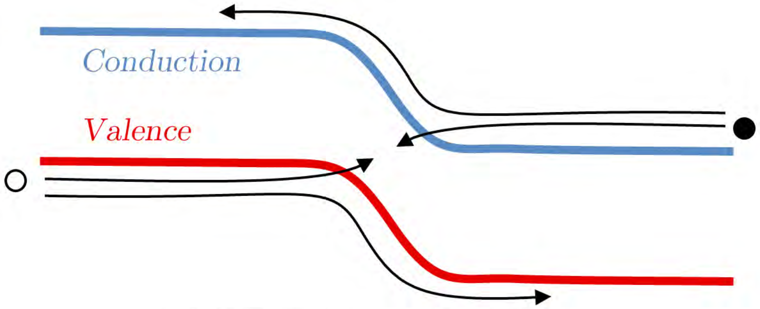
[1] A graded p-side corresponds to a grading from one material (for instance GaN) to another (AlN): if the two materials do not present the same bandgap, there is a variation of the bandgap over the p-side.

- the third one is a p-n junction, whose p- and n-sides are both graded, which results in a band structure presented in Figure 6–1(c). For the same reason as above, the valence band on the p-side and the conduction band on the n-side are flat, resulting in an Electron Blocking Layer on the p-side and a Hole Blocking Layer on the n-side.

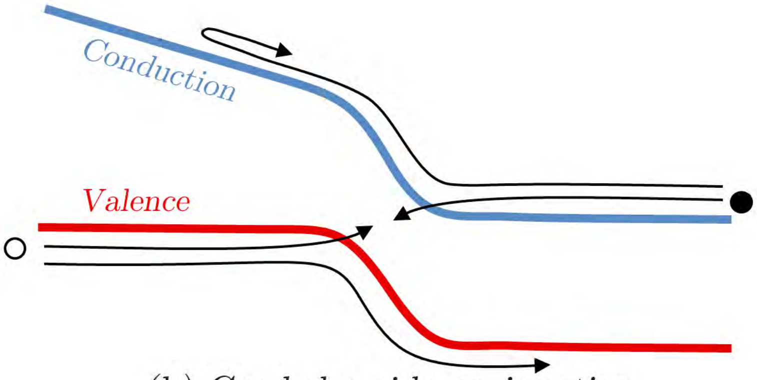
## 6.2 Comparison of the perfomance

In Figure 6–2 are presented the I-V curves of the three devices. As we can see, The three devices present a region where the current is an exponential function of the applied voltage, as is expected for any kind of p-n junction. However, at high injection currents, we can see different behaviours:

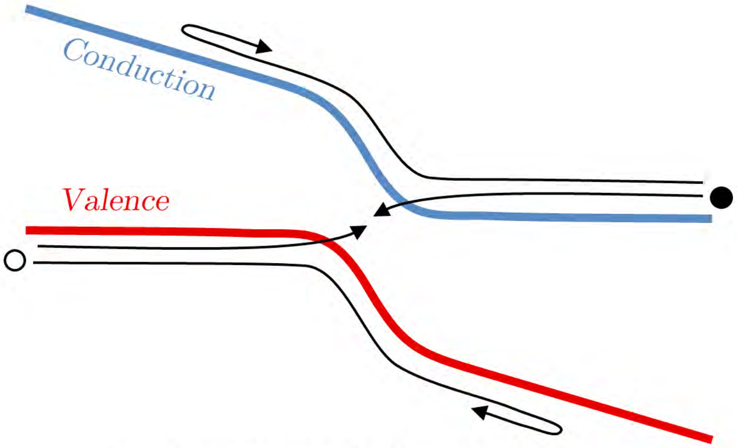
- first, the graded device exhibits a current that follows the behaviour of the GaN-device, but the current is lower by a factor 3. How can this be explained? As explained in reference [17], a p-n junction is all about the diffusion of minority carriers to the other side, i.e. diffusion of holes to the n-side and diffusion of electrons to the p-side. At high voltages, above 2.6V on Figure 6–2(a), the barrier is significantly reduced, the diffusion currents consequently become really important. However, the Electron Blocking Layer in the graded device prevents the diffusion of the electrons to the p-side, creating a traffic jam for the electrons. All the electrons injected have to recombine in the active region, and the recombination is the limiting factor: the electronic current is not even 1% of that of the GaN device, and most of the current is actually carried by the holes. The ratio $\frac{\mu_p}{\mu_p+\mu_n}$, equaled to 3 in our simulation, gives us the ratio between the total currents of the graded device and the GaN device.



*(a) Regular pn-junction*



*(b) Graded-p-side pn-junction*



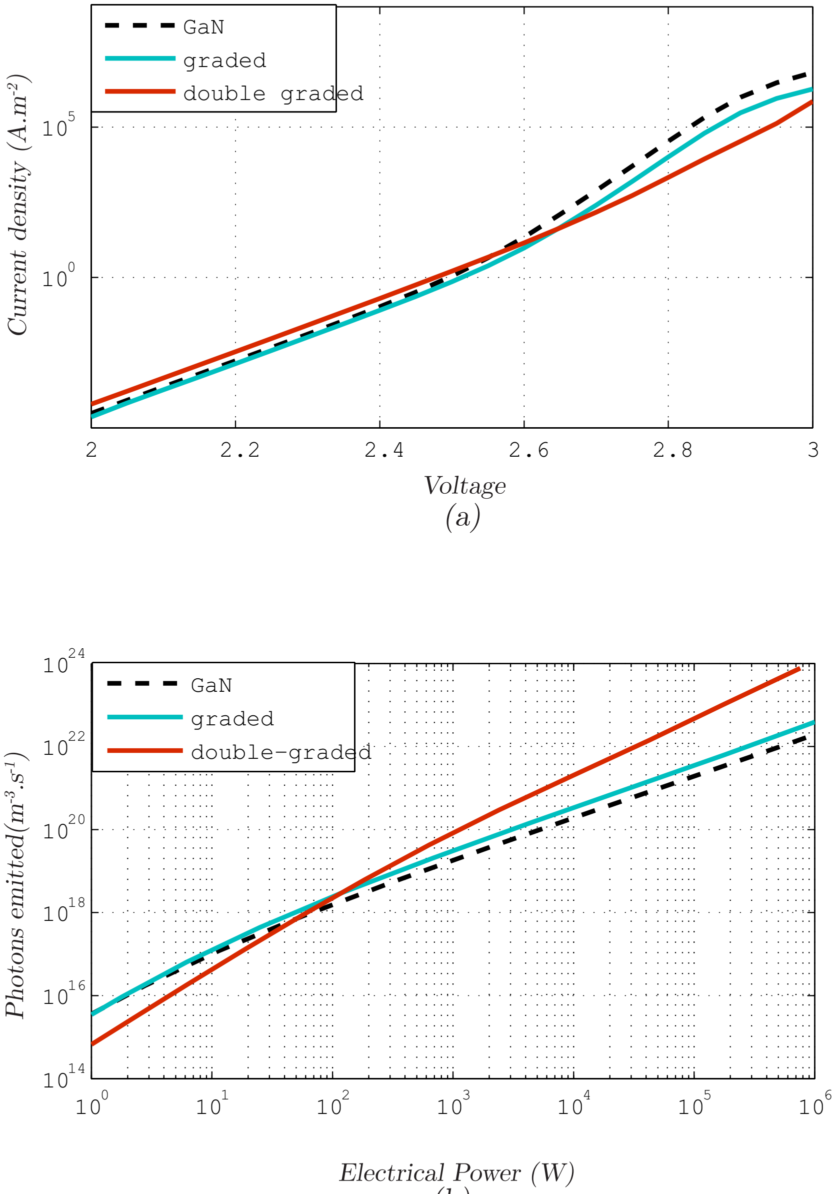
*(c) Double-graded pn-junction*

Figure 6–1: Band structures and behaviour of the carriers in the three devices

- second, the double-graded device, made of two graded sides, presents a current which does not follow the behaviour of the two previous devices, and there is this region where its current is 10 to 30 times lower than the others. The fact that its current do not follow the same trend as the two other curves can be easily explained by the total absence of diffusion, for both holes and electrons. At high voltages, the Electron Blocking Layer creates a traffic jam for the electrons and the Hole Blocking Layer creates one for the holes. The carriers injected have to recombine in the active region, and this recombination is a limiting factor.

If we have a look at the number of photons emitted as a function of the voltage, we see that the three devices present almost the same behavior. Does it mean that the three devices have the same efficiency? No it does not, because the efficiency is linked to the energy and not the voltage. Now, if we plot the number of photons emitted as a function of the current, we see that the graded device present an emission three times as important as the GaN device for the same current, and that the double-graded device present an emission 10 to 30 times greater than the two other devices. What we want to compare is actually the number of photons emitted per second as the function of the elecrical power provided to the device. The curves represented in Figure 6–2.b. This electrical power corresponds to the product $I \times V$, and the plot is presented in Figure 6–2. Since the behavior of the devices as a function of the voltage does not really change between the three devices, their behavior as a function of the power is very close to that as a function of the current:

- the graded device present an emission three time as important as the GaN device;



(a)

(b)

Figure 6–2: Comparison of the performance of the three devices (GaN device, graded p-side device and double-graded device)

- the double-graded device present an emission 10 to 30 times larger than the two other devices.

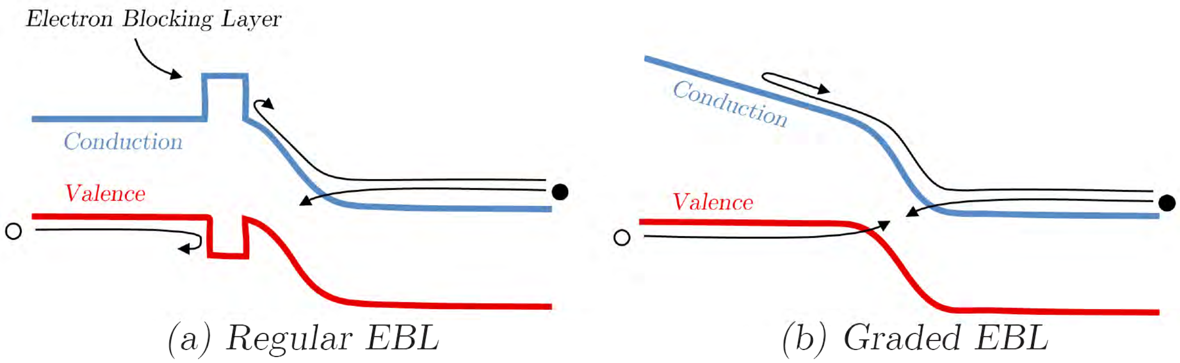
The key idea here is that the presence of the two Blocking Layers force the injected carriers to recombine; nothing ensures the radiative recombination, but at least the carriers do not escape the active region. For the same emission of light, less electrical power is required, because the carrier overflow has been completely removed: the device is simply more efficient.

One important point here is that this behavior is linked in our case to the polarization-doping, but it can easily be separated from it: the Electron and Hole Blocking Layers results from the variation of the bandgap and from the respective flat valence and conduction bands. However, if one grades a non-polar material, from GaAs to AlAs for instance, and if the impurity-based doping is constant we will have:

- a flat valence band on the p-side, with a varying bandgap, resulting in a Electron Blocking Layer;
- a flat conduction band on the n-side, with a varying bandgap, resulting in a Hole Blocking Layer;
- the same behavior as the double-graded device: the carriers cannot reach the opposite regions, i.e. all the current injected is undergoing recombination.

## 6.3 Graded vs regular Electron Blocking Layers

The idea of an Electron (or Hole) Blocking Layer did not arrive with this thesis. Indeed, EBL are already widely used in LEDs to prevent the direct electron overflow. Generally, this EBL is made of a thin layer of a wider bandgap semiconductor (for instance 10nm of AlN in a GaN-LED) placed on the p-side, next to the active region,



(a) Regular EBL (b) Graded EBL

Figure 6–3: Comparison of the two Electron Blocking Layers

to limit the direct escape of the electrons from the active region. However, this kind of EBL present several inefficiencies.

First of all, this EBL is made of a semiconductor where the ionization of acceptors is difficult: the hole concentration in this EBL is much lower than on the rest of the p-side. Since the carriers follow the Boltzmann statistics, the valence band in the EBL is at a lower energy than in the rest of the p-side, as described on Figure 6–3. The EBL, preventing the electrons from reaching the p-side, also prevents the holes from reaching the active region: this is the poor injection of holes.

Moreover, it has been shown that this EBL is not capable of preventing an indirect electron overflow, resulting from Auger recombination. The paper in reference [25] explains how an electron can escape the active region despite this EBL:

- an electron and a hole undergo Auger recombination, and a third carrier is excited in a state at energy $E_C + E_{gap}$.
- this carrier is stuck in another higher band and its relaxation time to the bottom of the conduction band is higher than the time it needs to diffuse over the EBL. The

electron do not undergo direct SRH-recombination to the bottom of the conduction band, but tend to diffuse to the p-side.

- because the electron is stuck in this higher band, it is in a state at higher energy than the top of the EBL, and has no difficulty passing over it, and gets to the p-side. This is an escape from the active region, in the same way as direct electron overflow.
- once the electron is in the p-side, it relaxes to the bottom of the conduction band, and there is no way it can go back into the active region. It is completely lost.

The advantage of the graded EBL is double:

- first, the valence band can be completely flat, presenting no issue of hole injection in the active region.
- second, the action of the graded EBL is not limited to its spatial extension. Even if the Auger electron penetrates in the p-side, the action of the quasi-electric field is extended to all the p-side: the electron is therefore brought back in the active region, and not completely lost in the p-side.

Consequently, graded (and extended) EBLs are expected to be more efficient than regular ones.

# CHAPTER 7
# Conclusion

Light-Emitting Diodes are solid-state devices, capable of transforming an electrical current into a beam of light, through a process called radiative recombination. However, current LEDs present a relatively low electrical-to-optical power conversion, from 10% to 50%, comparing to their possible efficiency, above 75%. This low efficiency in mainly due to the well-known efficiency droop at high current-injections.

There is no doubt that this increased efficiency will result in a shifting of the lighting market, currently based on incandescent and fluorescent lighting, towards LED-lamps. Indeed, the economical and ecological savings promised by these LED are interesting both for developing countries, where the growing population and urbanization increase the demand for lighting, and developed countries, where global warming and resource scarcity is very present in the minds. **80% of the market is expected to have shifted to LEDs by 2020**.

The main barrier to this market-shifting is the current LEDs efficiency. III-Nitride LEDs are necessary to generate white light, and their efficiency is not good enough today, mainly because of the poorly-efficient ionization of acceptors in these materials. Therefore, another doping technique might turn out to improve the efficiency.

Polarization-induced doping is such a technique. **Inducing a continuous variation of the polarization results in the presence of charges** in the device, which will attract carriers of opposite sign and thus create the doping. The III-Nitride semiconductors are good candidates for this structure, since they exhibit different polarizations and crystallize in the same lattice structure, the wurtzite lattice. However, one major difference with impurity-based doping is that during the growth of the crystal, the polarization charge is not compensated, i.e. carriers must come from somewhere to realize the doping. Past studies were unclear about the origin of this compensating-charge, and even assumed that polarization p-doping could only work with acceptors.

With a simple time-dependent model, we have shown that **we do not necessarily need acceptors to have a working polarization-induced p-doping**, contrary to what seems to be thought: growing the p-side on a reservoir of holes should work as well. The use of acceptors also makes it work, but is very inefficient, since only the acceptors in the depletion region are ionized, and provide the holes to the rest of the p-side. If no acceptor is introduced, interband tunneling can provide the necessary holes, but the large bandgap of the materials considered makes it unlikely to happen.

**Polarization-induced doping reduces the SRH recombination, since it requires no acceptors, but has no impact on Auger recombination**, which turned out to be the main cause of the efficiency droop. But the grading also results in the presence of an Electron or Hole Blocking Layer, which **strongly limits the electron or hole overflow**, such that a large portion of the current actually recombines. And **this effect is not limited to polar materials only**: any grading can result in these Blocking Layers.

The last point to be suggested here is that **not only III-Nitride semiconductors are concerned by polarization-induced doping**: any semiconductor crystallizing in a polar lattice can be used, and there actually exists many. For instance, III-Arsenide semiconductors, so widespread in the current market, can crystallize in the wurtzite lattice when grown in nanowires.

Polarization-induced doping is therefore an alternative to impurity-based doping, but one really need to pay attention to the way of growing the device for the doping to work. This technique will possibly improve the device efficiency, but cannot tackle completely the efficiency droop, since it has no impact on one the main loss processes, the Auger recombination.

# Appendix A - Field ionization of dopants

As we have seen, deep-level dopants are very reluctant to release their mobile charges, since the probability of thermal ionization decreases exponentially with the difference in energy between the level and the corresponding band. However, it is still possible to ionize deep acceptors and deep donors, if an electric field is applied [41, 42, 43].

When an electric field is applied, the bands exhibit a bending, which allows the hole of an acceptor (or the electron of a donor) to be released in the valence band (or the conduction band) by one of the 3 following process:

- the Frenkel-Poole ionization: the slope of the valence band leads to a reduction of the escape potential

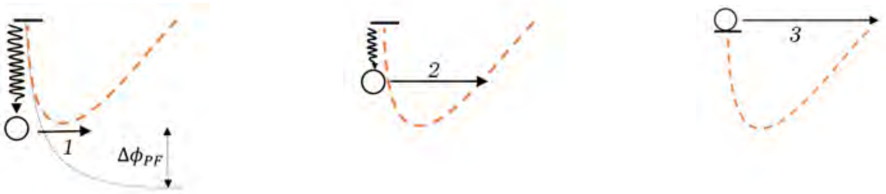
  $$\Delta\phi_{FP} = \left(\frac{q^3}{\pi\epsilon}\right)^{1/2}\sqrt{\mathbf{E}} \tag{A.1}$$

  leading in a probability of ionization increased by a factor $e^{\frac{\Delta\phi_{FP}}{k_B T}}$.

- the direct tunneling: the slope of valence band present a state filled with a valence electron at the same energy than the acceptor level, at a tunnable distance from the acceptor site: the valence electron can therefore tunnel from the band to the acceptor state, or equivalently, the hole can tunnel from the acceptor to the filled state. The probability of tunneling can be approached by the WKB approximation:

  $$T = exp\left(-2\int dx\sqrt{\frac{2m}{\hbar^2}}\left(V(x)-V_0\right)\right) \tag{A.2}$$

(a) without electric field (b) with electric field



(c) Frenkel-Poole (d) PA tunneling (e) Direct tunneling

Figure A.1: Field-ionization of deep acceptors: the band-bending due to the electric field allows the ionization of the acceptor via one of the three processes (c), (d), (e)

- the phonon assisted tunneling: an electron at a lower energy than in the previous case can tunnel into the bandgap and then get the extra energy to populate the acceptor level from the lattice vibrations (phonons). This phonon-assisted tunneling is more likely to happen at room temperature because both the barrier height and the distance to tunnel are lower than in direct tunneling, resulting in an exponentially increased tunneling probability. However, this process cannot occur at $T = 0K$ since it requires phonons.

## Appendix B - Calculation of the interband tunneling current

In our model, we need to calculate the numerical value of the interband tunneling current. References [44, 45, 46] gives the outline of the calculations, which are summarized here.

If we denote $T$ the probability for an electron to tunnel through a barrier, the WKB approximation gives us the expression A.2. With this expression, we can compute the value of the tunneling current for electrons between energies $E_A$ and $E_B$, which is the sum over the number of states in this range of energy, times their occupancy, times their group velocity, times the probability of tunneling, which gives us the following expression

$$\mathbf{J}_{tunn} = \frac{2e}{(2\pi)^3} \int_{E_A}^{E_B} T(E_x) \left(f(E) - f(E')\right) \mathbf{v}_g(E) \mathrm{d}k_x \mathrm{d}k_y \mathrm{d}k_z \tag{B.3}$$

By replacing the group velocity by its expression and assuming that almost all states of the valence band are filled and almost all states of the conduction band are empty, i.e. $(f(E) - f(E')) = 1$, we find

$$\mathbf{J}_{tunn} = \frac{2e}{(2\pi)^3} \int_{E_A}^{E_B} T(E_x) \frac{1}{\hbar} \frac{\partial E}{\partial k_x} \mathrm{d}k_x \mathrm{d}k_y \mathrm{d}k_z \tag{B.4}$$

Finally, by changing variables via $E = E_x + E_y + E_z$ and $E_\perp = E_y + E_z$, we find the following expression for the current:

$$\mathbf{J}_{tunn} = \frac{em^*}{2\pi^2\hbar^3} \int_{E_A}^{E_B} \int_0^E T(E) exp\left(-\frac{2E_\perp}{\bar{E}}\right) \mathrm{d}E_\perp \mathrm{d}E \tag{B.5}$$

where $\bar{E} = \frac{4\sqrt{2}}{3\pi} \frac{q\hbar E}{m^* E_{gap}}$ For a small grid, the energies $E_A$ and $E_B$ are really close, so that $T(E)$ can be taken constant, equaled to $T\left(\frac{E_A+E_B}{2}\right)$, and the tunneling current can be fairly approximated, after integration and at the first order, by the following expression:

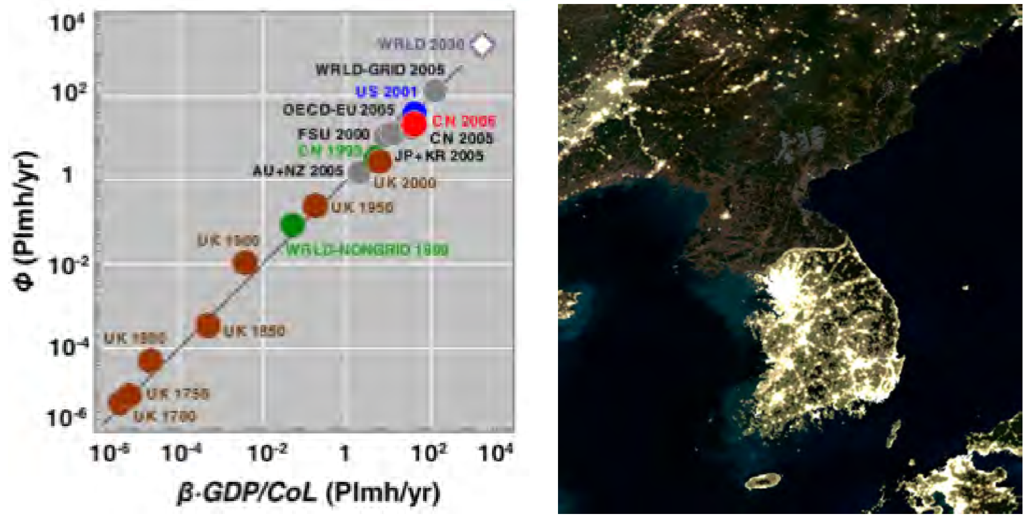
$$\mathbf{J}_{tunn} = \frac{qm^*}{2\pi^2\hbar^3} T\left(\frac{E_A + E_B}{2}\right) \bar{E}\left(E_B - E_A\right) \tag{B.6}$$

## Appendix C - Economic analysis: what is 5% of efficiency worth?

Let's imagine that we are in the case where LEDs exhibit a 95% efficiency. The question of whether or not subsidizing research to reach the missing 5% is relevant: on one hand the objective of ultra-efficiency would already have been achieved, and the financial and ecological savings would have been met; on the other hand, improving the efficiency would result in even more savings. Is the improvement worth the investment? We will review here an economic analysis carried by J. Y. Tsao [28, 47, 48, 49], showing that the increase in LED-efficiency results in an increasing wealth. A recent study [50] on the consumption of light between 1700 and today showed that there was a close relationship between the per-capita consumption of light $\psi$, the per-capita gross domestic product $gdp$ and the cost of light $coL$.

$$\psi = \beta \frac{gdp}{coL} \tag{C.7}$$

This relation is illustrated in Figure C.2a, and the proportionality constant $\beta$ is actually found to be equaled to 0.72%. Given the global 2012 GDP and the cost of light averaged over the world, this represent a global consumption of light of 440 billion dollars in 2012. This formula also indicates that the wealthier a population is, the more light it would consume, as illustrated by the Figure C.2b, showing the difference of lighting capacities between North and South Korea in 2011. Another implication is that if the price of light lowers, its consumption inversely increases.



(a) (b)

Figure C.2: Relation between consumption of light and GDP (a) and vision from space of the Korea peninsula (b) [8]

We can estimate the corresponding consumption of energy, using the relation

$$\psi = e_{\psi} \times \eta \tag{C.8}$$

where $e_{\psi}$ is the energy provided for lighting and $\eta$ is the overall efficiency of the electrical-to-optical efficiency. Consequently, the cost of light $coL$ can be rewritten as a function of the cost of energy:

$$coL = coL_{operating} + coL_{capital} = \frac{coE + \kappa\, coE}{\eta} = coE\frac{1+\kappa}{\eta} \tag{C.9}$$

where it was assumed that the capital cost of light was proportional to the operating cost of light by a factor 1/3. Combining Equations C.8 and C.9 gives us the relation between the per-capita energy used to produce light , its cost and the per-capita gdp:

$$e_{\psi} = \frac{\beta}{1+\kappa}\frac{gdp}{coE} \tag{C.10}$$

This time, the constant of proportionality is $\frac{\beta}{1+\kappa} = 0.54\%$. In the case where we consider that light is a factor of production, we can turn to a model widely used in economics, because of its simplicy and accuracy: the Cobb-Douglas model. Since light is assumed here to be a factor of production, we can write a relationship between the per-capita gdp, the per-capita consumption $\psi$ and the other factor of productions (all except labor) $\chi$:

$$gdp(\chi, \psi) = A \times \chi^{\alpha} \times \psi^{\beta} \tag{C.11}$$

Since labor is estimated at a portion of production equaled $1 - \alpha - \beta = 0.7$, we find that $\alpha + \beta$=0.3. Moreover the function of cost can be written as

$$cost(\chi, \psi) = \chi coX + \psi coL \tag{C.12}$$

We can then define the profit function as

$$\pi(\chi, \psi) = gdp(\chi, \psi) - cost(\chi, \psi) \tag{C.13}$$

and try to maximize the profit by using the Lagrangian method:

$$\begin{cases} \dfrac{\partial \pi}{\partial \chi} = 0 \\ \dfrac{\partial \pi}{\partial \psi} = 0 \end{cases} \tag{C.14}$$

$$\begin{cases} \dfrac{\partial gdp}{\partial \chi} - coX = \dfrac{\alpha}{\chi}\, gdp - coX = 0 \\ \dfrac{\partial gdp}{\partial \psi} - coL = \dfrac{\beta}{\psi}\, gdp - coL = 0 \end{cases} \tag{C.15}$$

$$\begin{cases} \chi_{opt} = \alpha \dfrac{gdp}{coX} \\ \psi_{opt} = \beta \dfrac{gdp}{coL} \end{cases} \tag{C.16}$$

This equation is in agreement with the empirical data explained in equation C.8. We can the identify $\beta = 0.0072$ and deduce that $\alpha = 0.3 - \beta = 0.2928$. We have therefore $\beta \ll \alpha$, which means that the light is a factor of production with much less impact than all others factor of production (represented by $\chi$ and labor). The optimal gdp can then be calculated:

$$gdp_{opt} = A \times \chi_{opt}^{\alpha} \times \psi_{opt}^{\beta} = A^{\frac{1}{1-\alpha-\beta}} \times \left(\frac{\alpha}{coX}\right)^{\frac{\alpha}{1-\alpha-\beta}} \times \left(\frac{\beta}{coL}\right)^{\frac{\beta}{1-\alpha-\beta}} \tag{C.17}$$

so that the gdp-independent relation between the consumption of light and the different costs can be extracted:

$$\psi = A^{\frac{1}{1-\alpha-\beta}} \times \left(\frac{\alpha}{coX}\right)^{\frac{\alpha}{1-\alpha-\beta}} \times \left(\frac{\beta}{coL}\right)^{\frac{1-\alpha}{1-\alpha-\beta}} \tag{C.18}$$

With this equation, it is clear that a reduction in the cost of light leads to an increase in the consumption light in two ways: a first direct contribution to $\psi$ via Equation C.18, contribution which increase the gdp through Equation C.11, which in turns re-increase $\psi$. For instance, an increase in efficiency from 85% to 95% would result in a growth of $65B per year of the global GDP.

This analysis shows that the main advantage of a reduced cost of light is an increase in the global GDP: it is therefore economically worth investing for a few percents more.

# References

[1] J. Y. Reproduced from Tsao, “Solid-state lighting: lamps, chips, and materials for tomorrow,” *Circuits and Devices Magazine, IEEE*, vol. 20, no. 3, pp. 28–37, 2004.

[2] J. Cho, E. F. Schubert, and J. K. Kim, “Efficiency droop in light-emitting diodes: Challenges and countermeasures,” *Laser & Photonics Reviews*, vol. 7, no. 3, pp. 408–421, 2013. Reproduced.

[3] R. F. Pierret, “Semiconductor device fundamentals,” 1996. Reproduced.

[4] T. Baumgartner, F. Wunderlich, D. Wee, and A. Jaunich, “Lighting the way: Perspectives on the global lighting market, 2nd edition,” 2012. Reproduced.

[5] T. Baumgartner, F. Wunderlich, and D. Wee, “Lighting the way: Perspectives on the global lighting market, 1st edition,” 2011. Reproduced.

[6] H. Kroemer, “Nobel lecture: Quasielectric fields and band offsets: teaching electrons new tricks,” *Reviews of modern physics*, vol. 73, no. 3, pp. 783–793, 2001. Reproduced.

[7] R. F. Pierret and G. W. Neudeck, *Advanced semiconductor fundamentals*, vol. 6. Addison-Wesley Reading, MA, 1987. Reproduced.

[8] J. Y. Tsao, J. J. Wierer Jr, L. E. Rohwer, M. E. Coltrin, M. H. Crawford, J. A. Simmons, P.-C. Hung, H. Saunders, D. S. Sizov, R. Bhat, *et al.*, “Introduction part b. ultra-efficient solid-state lighting: Likely characteristics, economic benefits, technological approaches,” *III-Nitride Based Light Emitting Diodes and Applications*, pp. 11–26, 2013. Reproduced.

[9] J. Y. Tsao, “Solid-state lighting: lamps, chips, and materials for tomorrow,” *Circuits and Devices Magazine, IEEE*, vol. 20, no. 3, pp. 28–37, 2004.

[10] M. H. Crawford, “Leds for solid-state lighting: performance challenges and recent advances,” *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 15, no. 4, pp. 1028–1040, 2009.

[11] J. J. Wierer, D. S. Sizov, A. Neumann, S. R. Brueck, and J. Y. Tsao, "Study of iii-nitride laser diodes for solid-state lighting," *CLEO: Science and Innovations*, pp. CF1M–4, 2013.

[12] C. Wood and D. Jena, "Polarization effects in semiconductors: From ab initio theory to device applications," 2007.

[13] D. Jena, S. P. Alpay, and J. V. Mantese, "Functionally graded polar heterostuctures: New materials for multifunctional devices," *Polarization Effects in Semiconductors*, pp. 307–372, 2008.

[14] Z.-H. Zhang, S. T. Tan, Z. Kyaw, W. Liu, Y. Ji, Z. Ju, X. Zhang, X. W. Sun, and H. V. Demir, "p-doping-free ingan/gan light-emitting diode driven by three-dimensional hole gas," *Applied Physics Letters*, vol. 103, no. 26, p. 263501, 2013.

[15] C. Kittel and P. McEuen, *Introduction to solid state physics, chapter 8*, vol. 8. Wiley New York, 1976.

[16] E. Rosencher, *Optoelectronics, chapter 7*. Cambridge University Press, 2002.

[17] R. F. Pierret, "Semiconductor device fundamentals," 1996.

[18] M. Yamada, Y. Narukawa, and T. Mukai, "Phosphor free high-luminous-efficiency white light-emitting diodes composed of ingan multi-quantum well," *Japanese Journal of Applied Physics*, vol. 41, no. 3A, p. L246, 2002.

[19] R. Haitz and J. Y. Tsao, "Solid-state lighting:the case10 years after and future prospects," *physica status solidi (a)*, vol. 208, no. 1, pp. 17–29, 2011.

[20] J.-I. Shim, "Active region part b. internal quantum efficiency," *III-Nitride Based Light Emitting Diodes and Applications*, pp. 153–195, 2013.

[21] E. F. Schubert and J. K. Kim, "Solid-state light sources getting smart," *Science*, vol. 308, no. 5726, pp. 1274–1278, 2005.

[22] M.-H. Kim, M. F. Schubert, Q. Dai, J. K. Kim, E. F. Schubert, J. Piprek, and Y. Park, "Origin of efficiency droop in gan-based light-emitting diodes," *Applied Physics Letters*, vol. 91, no. 18, p. 183507, 2007.

[23] J. Cho, E. F. Schubert, and J. K. Kim, "Efficiency droop in light-emitting diodes: Challenges and countermeasures," *Laser & Photonics Reviews*, vol. 7, no. 3, pp. 408–421, 2013.

[24] J. Piprek, "Efficiency droop in nitride-based light-emitting diodes," *physica status solidi (a)*, vol. 207, no. 10, pp. 2217–2225, 2010.

[25] J. Iveland, L. Martinelli, J. Peretti, J. S. Speck, and C. Weisbuch, "Direct measurement of auger electrons emitted from a semiconductor light-emitting diode under electrical injection: identification of the dominant mechanism for efficiency droop," *Physical review letters*, vol. 110, no. 17, p. 177406, 2013.

[26] T. Baumgartner, F. Wunderlich, and D. Wee, "Lighting the way: Perspectives on the global lighting market, 1st edition," 2011.

[27] T. Baumgartner, F. Wunderlich, D. Wee, and A. Jaunich, "Lighting the way: Perspectives on the global lighting market, 2nd edition," 2012.

[28] J. Y. Tsao, M. E. Coltrin, M. H. Crawford, and J. A. Simmons, "Solid-state lighting: an integrated human factors, technology, and economic perspective," *Proceedings of the IEEE*, vol. 98, no. 7, pp. 1162–1179, 2010.

[29] R. Stevenson, "The led's dark secret," *Spectrum, IEEE*, vol. 46, no. 8, pp. 26–31, 2009.

[30] B. B. Wittneben, "The impact of the fukushima nuclear accident on european energy policy," *Environmental Science & Policy*, vol. 15, no. 1, pp. 1–3, 2012.

[31] I. Batra, P. Wurfel, and B. Silverman, "Phase transition, stability, and depolarization field in ferroelectric thin films," *Physical Review B*, vol. 8, no. 7, p. 3257, 1973.

[32] Y.-H. Cho, G. Gainer, A. Fischer, J. Song, S. Keller, U. Mishra, and S. DenBaars, "s-shaped temperature-dependent emission shift and carrier dynamics in ingan/gan multiple quantum wells," *Applied physics letters*, vol. 73, no. 10, pp. 1370–1372, 1998.

[33] V. Fiorentini, F. Bernardini, and O. Ambacher, "Evidence for nonlinear macroscopic polarization in iii–v nitride alloy heterostructures," *Applied physics letters*, vol. 80, no. 7, pp. 1204–1206, 2002.

[34] V. Fiorentini, F. Bernardini, F. Della Sala, A. Di Carlo, and P. Lugli, "Effects of macroscopic polarization in iii-v nitride multiple quantum wells," *Physical Review B*, vol. 60, no. 12, p. 8849, 1999.

[35] F. Bernardini, V. Fiorentini, and D. Vanderbilt, "Spontaneous polarization and piezoelectric constants of iii-v nitrides," *Physical Review B*, vol. 56, no. 16, p. R10024, 1997.

[36] F. Bernardini, "Spontaneous and piezoelectric polarization: Basic theory vs. practical recipes," *Nitride Semiconductor Devices: Principles and Simulation*, pp. 49–68, 2007.

[37] R. F. Pierret and G. W. Neudeck, *Advanced semiconductor fundamentals*, vol. 6. Addison-Wesley Reading, MA, 1987.

[38] H. Kroemer, "Nobel lecture: Quasielectric fields and band offsets: teaching electrons new tricks," *Reviews of modern physics*, vol. 73, no. 3, pp. 783–793, 2001.

[39] D. Vasileska and S. M. Goodnick, "Computational electronics," *Synthesis Lectures on Computational Electromagnetics*, vol. 1, no. 1, pp. 1–216, 2005.

[40] J. Simon, V. Protasenko, C. Lian, H. Xing, and D. Jena, "Polarization-induced hole doping in wide–band-gap uniaxial semiconductor heterostructures," *Science*, vol. 327, no. 5961, pp. 60–64, 2010.

[41] J. G. Simmons, "Poole-frenkel effect and schottky effect in metal-insulator-metal systems," *Physical Review*, vol. 155, no. 3, p. 657, 1967.

[42] P. Kozodoy, M. Hansen, S. P. DenBaars, and U. K. Mishra, "Enhanced mg doping efficiency in al 0.2 ga 0.8 n/gan superlattices," *Applied physics letters*, vol. 74, no. 24, pp. 3681–3683, 1999.

[43] E. Schubert, W. Grieshaber, and I. Goepfert, "Enhancement of deep acceptor activation in semiconductors by superlattice doping," *Applied physics letters*, vol. 69, no. 24, pp. 3737–3739, 1996.

[44] L. Esaki, "Long journey into tunneling," *Reviews of Modern Physics*, vol. 46, no. 2, pp. 237–244, 1974.

[45] E. O. Kane, "Theory of tunneling," *Journal of Applied Physics*, vol. 32, no. 1, pp. 83–91, 1961.

[46] E. Kane, “Zener tunneling in semiconductors,” *Journal of Physics and Chemistry of Solids*, vol. 12, no. 2, pp. 181–188, 1960.

[47] J. Y. Tsao, J. J. Wierer Jr, L. E. Rohwer, M. E. Coltrin, M. H. Crawford, J. A. Simmons, P.-C. Hung, H. Saunders, D. S. Sizov, R. Bhat, *et al.*, “Introduction part b. ultra-efficient solid-state lighting: Likely characteristics, economic benefits, technological approaches,” *III-Nitride Based Light Emitting Diodes and Applications*, pp. 11–26, 2013.

[48] J. Y. Tsao, H. D. Saunders, J. R. Creighton, M. E. Coltrin, and J. A. Simmons, “Solid-state lighting: an energy-economics perspective,” *Journal of Physics D: Applied Physics*, vol. 43, no. 35, p. 354001, 2010.

[49] J. Y. Tsao, M. H. Crawford, M. E. Coltrin, A. J. Fischer, D. D. Koleske, G. S. Subramania, G. Wang, J. J. Wierer, and R. F. Karlicek, “Toward smart and ultra-efficient solid-state lighting,” *Advanced Optical Materials*, 2014.

[50] J. Y. Tsao and P. Waide, “The world’s appetite for light: empirical data and trends spanning three centuries and six continents,” *Leukos*, vol. 6, no. 4, pp. 259–281, 2010.