Politecnico di Torino

Laurea Specialistica in Ingegneria Fisica
(Physics Engineering Master of Science Course)

Master Thesis

TCAD Models for Tunneling Processes in Narrow-Gap Semiconductors

Marco Mandurrino

December 2013
# Contents

**Abstract** iii

**Acknowledgments** iv

1 **The InfraRed Photon Detection** 1
   1.1 Why using a Narrow-gap Semiconductor .......................... 1
   1.2 Review on some Physical Aspects in HgCdTe .................... 6
      1.2.1 Introduction ........................................ 6
      1.2.2 Crystal Structure and its Growth Processes .............. 8
      1.2.3 Band Structure ...................................... 10
      1.2.4 Optical and Electric Properties .......................... 17
         A – Mobility and Intrinsic Concentration ........................ 18
         B – Carrier Lifetime and GR Processes ........................ 19
         C – Photon Absorption Properties ............................ 21
         D – Doping, Impurities and Defects ........................... 24
   1.3 Fundamental Concepts concerning IR Photodetectors .......... 24
      1.3.1 Theory of Photon Detection ............................. 25
      1.3.2 Figures of Merit and Performance Limitations ............ 39
      1.3.3 The State of the Art about HgCdTe IR Photodetectors .... 43

References .................................................. 45

2 **Towards the Modeling of the Tunneling Process** 47
   2.1 The Semi-classical Approximation of the Quantum Mechanics .. 48
   2.2 The Mechanism of Tunneling in Semiconductors ............... 50
      2.2.1 Trap-Assisted Tunneling ................................ 54
      2.2.2 Band-to-Band Tunneling ................................ 56
      2.2.3 Phonon-Assisted Tunneling ............................... 58
   2.3 Tunneling Simulations ...................................... 59
      2.3.1 Brief Introduction to the Physics-based Numerical Simulation of Electronic Devices .................................. 59
      2.3.2 Kane Model ............................................. 61
      2.3.3 Hurkx Model ............................................ 62
      2.3.4 Schenk Model .......................................... 63
      2.3.5 Other models ........................................... 65
CONTENTS

2.4 Overview on the Tunneling Simulating Tools in Synopsys® TCAD Sentaurus Device ........................................ 65
2.5 Actual HgCdTe-based Devices Modeling .................................................. 68
References ........................................................................................................... 71

3 TCAD Simulation of an HgCdTe-based IR Photodetector ............................. 73
3.1 Details about the Drift-Diffusion Transport Model ................................. 76
  3.1.1 Boundary Conditions .................................................................. 78
  3.1.2 Discretization Procedure ............................................................. 80
3.2 A Mono-dimensional Preliminary Study ................................................. 83
  3.2.1 Energy trap $E_t$ Dependence ...................................................... 83
  3.2.2 Temperature Dependence ............................................................. 85
3.3 Device Project and its Equilibrium Features ....................................... 90
References ........................................................................................................... 100

4 Results, Comments and Perspectives ......................................................... 101
  4.1 Tunneling Simulations in Dark Conditions ....................................... 101
    4.1.1 $E_t$ and $T$ Parametric Analysis ............................................. 102
    4.1.2 Trap-Assisted Tunneling Simulations ....................................... 105
    4.1.3 Band-to-Band Tunneling Simulations ....................................... 108
  4.2 A new model for BTBT in $p$-$i$-$n$ structures .................................... 111
  4.3 Tunneling Simulations under Illumination ......................................... 115
  4.4 Conclusions ......................................................................................... 118
References ........................................................................................................... 121

Appendix A Hg$_{0.7054}$Cd$_{0.2946}$Te Simulation Physical Parameters ........ 122
Appendix B C++ Routine for the 1D Kane BTBT Model .............................. 128
Appendix C C++ Routine for the 2D Kane BTBT Model .............................. 131
Appendix D C++ Routine for the 2D $p$-$i$-$n$ BTBT Model ........................ 135
Tunneling processes probably are among the most important generation-recombination mechanisms negatively affecting the behavior of photodiodes, especially in dark conditions. The tunneling effect - according to which a quantum particle can travel through classically forbidden regions - cannot be completely eliminated from a technological standpoint, since it is related to the probabilistic nature of the electron wavefunction. In semiconductor devices this implies that, in those regions where the band bending occurs, for instance due to an external electric field or to a junction built-in potential, the electrons can travel from the valence band of one side of the junction towards the conduction band of the other side, giving rise to an unwanted dark current flowing through the device also without the presence of photons. Such a mechanism can occur both directly (in this case it is named Band-to-Band Tunneling, BTBT) or with the contribution of some impurity levels located within the forbidden gap (the so-called Trap-Assisted Tunneling, TAT). Moreover, for the IR detection applications narrow-gap semiconductors are needed. Unfortunately this requirement produces an increase of current, given that within the diode depletion region a narrower gap means a shorter tunneling path and thus a higher tunneling probability. Also temperature and bias voltage affect this probability, making the device design crucial. For these reasons TCAD simulation is an essential tool, both for structures and their performance optimization, as well as for limiting the influence of noise physical mechanisms.

In the present work we analyze the use of narrow-gap semiconductors in IR photodetectors, giving a complete description of one of the most widely appreciated materials in literature for infrared detection, HgCdTe, a ternary alloy that allows to vary the bandgap width through the modification of the so-called molar fraction (the abundance ratio of CdTe over HgTe). Then the principles of photodetection and the theoretical implications of tunneling in the IR framework are treated. After an insight about the most used models for BTBT and TAT mechanisms, we finally explore the capabilities of the drift-diffusion modeling suite Synopsys® TCAD Sentaurus about tunneling simulation by studying their effects on the reverse I-V characteristics. A first simple 1D HgCdTe-based pn diode in dark conditions was analyzed in order to demonstrate the efficacy of the implemented models in terms of physical and analytical parameter response. Then a case-study represented by a 2D HgCdTe-based p-i-n structure was tested more deep in detail, both in dark conditions and under illumination. Besides the comparison between all available models, a recently published new formulation for the BTBT generation rate - explicitly conceived for p-i-n diodes - was introduced and compared with such models.

In conclusion, this thesis aims to provide the reader both the basic and advanced tools to understand the tunneling mechanisms, the state of the art concerning the techniques for its simulation and the effectiveness of the simulation itself in the IR detection field of research, giving also some ideas to improve the theoretical models for specific structures design.
I would like to thank my tutors, Professor Michele Goano and Dr. Francesco Bertazzi, for their helpfulness, as well as the “Dipartimento di Elettronica e Telecomunicazioni – DET” at Politecnico di Torino and his Director, Professor Giovanni Ghione, for making available the license software of Synopsys® TCAD Sentaurus (ver. H-2013.03) and the server Urano, essential to simulations. A very special thanks also goes to Dr. Marco Vallone – who has patiently watched my draft thesis and my codes – for his useful replies to my work. I dedicate this thesis to my family and to Aurora.
«Nature, as we understand it today, behaves in such a way that it is fundamentally impossible to make a precise prediction of exactly what will happen in a given experiment»

Richard P. Feynman — “Six Easy Pieces” (1964)
Before analyzing the main physical aspects about tunneling and its possible implementations in TCAD simulators or in theoretical transport models, we introduce now the fundamental concepts concerning the operating principles of narrow gap infrared photodetectors and their physical and optoelectronic properties. From our argumentation will naturally arise the need of deepen one of the most serious performance-limiting feature in this kind of semiconductor devices: the tunneling mechanism.

So, the aim of this first chapter is to move, starting from the basic theory, towards the study of the narrow-gap semiconductors employed in the infrared detection and then to insert them into the tunneling simulation context. To do this, we will examine firstly the scientific and historical reasons for this choice of materials in our particular application field. Secondly we will focus on one of the best narrow-gap semiconductor for IR detection and we will analyze also the fundamental working principles of IR photodetectors, both on the quantum and the engineering point of view.

After discussing the main limiting effects on detectors, especially in the case of HgCdTe-based devices, we will end this section browsing the state of the art in literature about this kind of electronic systems. Hence, this chapter collocates at the beginning of the theoretical dissertation on central subject of the present work: TCAD models for tunneling processes in narrow-gap semiconductors, which will be developed in the next chapter. At the end of the whole introduction we will be able to understand and analyze a concrete study of a device physics-based numerical simulation.

1.1 Why using a Narrow-gap Semiconductor

One of the most diffused applications of the narrow-gap semiconductors is without doubts its employment as an active material for the light detection in the IR range (ideally 0.7 \( \mu \text{m} \lesssim \lambda \lesssim 1000 \ \mu\text{m} \)). Since the photo-excitation – the physical phenomenon at the origin of the transduction from light to a current signal – is a energy-dependent process, the electrons involved are promoted into the conduction band thanks to this energy, that comes from incident photons.

If today applications of this technology are spreading in various areas of our life, at the very beginning, i.e. in the XIX century, infrared detection was relegated to the only
thermometric field. Three important inventions connoted this era, from the Herschel’s thermometer (1800) to the Langley’s bolometer (1881), passing through the Nobili’s thermopile (1829), thanks to the thermoelectric effect (i.e. the electrical potential variation as a function of the temperature) just discovered by Seebeck (1821). Only after the 1873, when Smith discovered the photoconductive effect (i.e. when the conductance of a material is enhanced under lighting conditions), the scientists’ efforts began moving towards the use of light as an ingredient for electronics, but the true first photoconductor for the infrared light was developed later, in 1917, by Case. It was made of thallium and sulfur, and in the following years many other materials were tested to improve the capability of IR light detection, through the lead sulfide devices firstly (especially during the World War I) and then with thallium sulfide. This first generation of devices were built using a single cooled element of these polycrystalline lead salt materials and they were characterized by some technological difficulties in the fabrication process that limited their reproducibility. During the World War II, when the Quantum Mechanics theories was almost entirely developed, the season of semiconductors (especially PbSe and PbTe) starts, reaching up to the present day through a continuous research of the best material under the performance standpoint. With the discovery of transistors by Shockley and his co-workers, in the 1947, and the contemporary introduction of controlled impurities techniques in Ge (as Cu, Zn, Au or – some decades later – also Hg) the age of extrinsic photodetectors begins, based on the extrinsic absorption (see Figure 1.1 - A). This process involves occupied states in the forbidden region of the main lattice constituent, due to the presence of some dopants. So an incident photon can undergo a photon-electron scattering, yielding totally or partially its energy $h\nu$ to the electron trapped in the impurity level, which can be now promoted into the conduction band.

![Figure 1.1. Schematic description of fundamental excitation processes in semiconductors: A – extrinsic absorption (between an impurity center, i.e. a trap, and the conduction band); B – intrinsic absorption; C – free carrier absorption. Adapted from [1].](image)

After the end of the World War II in 1945, and especially during the Cold War, the interest in IR detection – driven mainly by telecommunications and military demands – has focused mainly on the two atmospheric windows: 3-5 $\mu$m (the Mid-Wavelength Infra Red – MWIR) and 8-14 $\mu$m, where the latter is often called LWIR region (i.e. Long Wavelength
Infra Red). This band and the following VLWIR (the “V” stands for Very), from 14 to 30\,\mu m, was the usual functioning domain of the Ge-based extrinsic photodetectors. But the trend to detect the IR light with more and more longer wavelengths without the cooling requirements of these detectors (this is generally needed in detecting $\lambda > 3\,\mu m$), which in turn means loss of quantum efficiency, gave rise to the second generation of infrared photodetectors. This new kind of devices, by contrast, exploits the \textit{intrinsic absorption} (see Figure 1.1 - B) to get rid of all the technological and physical problems of controllability linked to the impurities introduction. Such a process was possible only through the huge advances in the field of narrow-gap semiconductors, that meet contemporarily the need for long-wavelength detection and intrinsic photo-stimulation. We deal with this

\textbf{Figure 1.2.} Transmission of the atmosphere for a 2 km horizontal path at sea level containing 17 mm of precipitate water (from [1]). The two windows, 3-5\,\mu m and 8-14\,\mu m, are pretty evident because of their high transmission.

category of semiconductors when the energy gap is $E_g \leq 0.5\,\text{eV}$ or if the wavelength at which the IR absorption decays by 3 dB – i.e. the cut-off wavelength – is over 2\,\mu m. The first materials used were semiconductor alloys made by elements of the III-V family as in the case of InSb, a zincblende crystal with $E_g$ between 0.17\,eV (at 300 K) and 0.23\,eV (at 80 K) depending on the temperature as follows [2]:

$$E_g(T) = 0.24 - 6 \cdot 10^{-4} \frac{T^2}{T + 500}. \tag{1.1}$$

In the late 1950s also the alloys of IV-VI and II-VI were discovered and then introduced in the technological framework, providing a large control on the detectable frequency range: once the intrinsic photo-stimulation is allowed if $h\nu/\lambda > E_g$, we are able to reach larger $\lambda$ choosing a narrower energy gap by changing the material. Between the 1950 and the 1960 also some ternary alloys were discovered, like the InAs$_{1-x}$Sb$_x$ (III-V) and Pb$_{1-x}$Sn$_x$Te (IV-VI). The latter was progressively abandoned due to its high dielectric constant and thermal expansion coefficient.

In Figure 1.3 it is shown the present possibility of tuning the energy gap and, thus, the spectral range of IR detection through the modern technologies. Each point (the filled squares) in the plane $E_g$–lattice constant represents a particular state corresponding to a binary alloy. Some of these points are linked through a line. These compounds are ternary alloys, obtained through growth processes that properly combine the abundance of one of the involved elements with respect to the others, in the fundamental cell. A particular case is represented by Si and Ge: the line linking them is the ensemble of all the possible intermediate configurations of the type Si$_{1-x}$Ge$_x$, where $x$ is the so-called \textit{molar fraction}.
1.1 Why using a Narrow-gap Semiconductor

Figure 1.3. Representation of the $E_g$ versus the lattice constant in several binary and ternary semiconductors (from [3]). The points (filled squares) represent the binary compounds, while the lines are all the possible ternary configurations between the two situations at the extremals. In the scheme are shown also Si and Ge.

It is worth noting the presence in the scheme of Figure 1.3 of an interesting compound, denoted with the long almost vertical line which their extremals are represented by the alloys HgTe and CdTe. Mixing them, in 1959, Lawson and his co-workers demonstrated for the first time the producibility of a new semiconductor: the $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (II-VI) [4]. They described the wide and unprecedented tuning possibility of this material from the HgTe ($x = 0$), that is ideally gapless and has an absorption edge at $\lambda = 38 \mu$m, up to the CdTe ($x = 1$), where this limit shifts at $\lambda = 0.8 \mu$m.

Nowadays HgCdTe is one of the most studied and employed semiconductors in the field of IR detection (outranked only by Si and GaAs), even if it is connoted by several technological problems by its production standpoint. This is mainly due to the high vapor pressure of Hg. Nonetheless there are many advantages in the use of this and other kinds of narrow-gap semiconductors, as their high optical absorption coefficient and high electron mobility.

As already said, intrinsic detectors need less stringent cooling requirements then the extrinsic ones: this fact is due to the competition between thermal transition processes and the optic ones, especially near room temperature, because of the presence of impurity traps that allow energetic interband hopping processes. So, thanks to the invention and progresses of new lithography standards, after the 1960 first BLIP (Background-LImited Performance) linear array photodetectors operating in the LWIR domain with a single cryogenic stage started to be produced. The following decade, with the introduction of CCDs (Boyle and Smith), the device architecture became more complex and at the same time also more compact, through a new kind of component integration and sophisticated read-out schemes which combines detection and processing all together. This innovation imposes much greater attention in the device design, not only by the tunneling standpoint but also, for instance, with regard to the electromagnetic interferences within the whole electronic system, meant as photon-IC and sensor-IC interaction.
1.1 Why using a Narrow-gap Semiconductor

Figure 1.4. Comparison between (normalized) detectivity $D^*$ curves of different compound photo-voltaic (PV), photo-conductive (PC), photo-emissive (PE) and photo-electromagnetic (PEM) devices, operating at 300 K (from [1]). All detectors have a cut-off frequency of 1 kHz, except for the thermopile, thermocouple, thermistor bolometer, Golay cell and pyroelectric detector (10 Hz). Dashed curves are calculated for the theoretical background-limited case.

Another important issue to be accounted for in this kind of detector design is the signal interface with the other systems. In other words a high-impedance is a good requirement in the interfacing with field-effect transistors (FETs). Seen in this terms, they were promising candidates for IR detection: PtSi Schottky barriers, InSb and HgCdTe photodiodes or PbSe, PbS and extrinsic Si photoconductors (PC).

The invention of SPRITE (Signal PRocessing In The Element) devices, which exploit the Time Delay and Integration (TDI) technology in a single elongated detector element, opened the way also in the use of the HgCdTe-based PC detectors. But through the 1980s, instead, the need of low-power dissipation and high-impedance tended to prefer the PV detector configuration, yielding a new generation of HgCdTe devices, made by 2D arrays, that developed until now. Only recently the research created a modern version of them, leading to a higher operating temperature (as in HOT devices) or by incorporating, for example, structures for multi-color detection. And, even if alternative alloys and systems are under investigation (mainly through Quantum Well Infrared Photodetectors – QWIPs, or quantum dots), HgCdTe is still the most widely used variable-gap semiconductor in infrared PV sensors, with a business of several billions of dollars.

The main today competitors of this semiconductor are: Schottky barriers on Si, SiGe heterojunctions, AlGaAs multiple QWs, GaInSb strain layer superlattices, high temperature superconductors and thermo-detectors as pyroelectric detectors and Si bolometers. Those have advantages by the fabrication point of view but, with respect to the perfor-
mannances, the HgCdTe remains always the best choice. For a complete description in Table 1.1 the most important characteristics of modern infrared photon detectors are reported, divided by type, physical mechanism (the transition column), output signal and specified with some relevant material examples.

<table>
<thead>
<tr>
<th>Type</th>
<th>Transition</th>
<th>Electrical Output</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>intrinsic</td>
<td>interband</td>
<td>photoconductive</td>
<td>PbS, PbSe, InSb, HgCdTe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>photovoltaic</td>
<td>InSb, InAs, PbTe, HgCdTe, PbSnTe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>capacitance, PEM</td>
<td>InSb, HgCdTe</td>
</tr>
<tr>
<td>extrinsic</td>
<td>from impurity to the CB</td>
<td>photoconductive</td>
<td>Si:In, Si:Ga, Si:As, Ge:Cu, Ge:Hg</td>
</tr>
<tr>
<td>free carriers</td>
<td>intraband</td>
<td>photoemissive</td>
<td>Schottky barriers GaAs/CsO, PtSi, PtSi, IrSi</td>
</tr>
<tr>
<td></td>
<td></td>
<td>photoconductive</td>
<td>InSb electron bolometer</td>
</tr>
<tr>
<td>quantum wells</td>
<td>to/from quantized levels</td>
<td>photoconductive</td>
<td>HgTe/CdTe, GaAs/AlGaAs, InGaAs/AlGaAs, InAs/InGaSb</td>
</tr>
<tr>
<td></td>
<td></td>
<td>photovoltaic</td>
<td>InSb/InAsSb, InAs/InAsSb</td>
</tr>
</tbody>
</table>

Table 1.1. Comparison of infrared detectors (adapted from [1]).

1.2 Review on some Physical Aspects in HgCdTe

1.2.1 Introduction

Since Hg$_{1-x}$Cd$_x$Te is probably the best material in the field of LWIR photodetection, in the last 30 years researchers focused their attention on the behavior of this ternary alloy. Although it has very good properties concerning electronic transport – like high quantum efficiency, responsivity and specific detectivity [5] – mercury cadmium telluride typically shows the evidence of some deviation from semiconductor theory, especially in dark conditions. This fact puts the basis of a field of research tending to find the physical reasons of such a phenomena and then to build a general theory for this kind of devices. Nonetheless HgCdTe became the most competitor in this framework, especially in military or astronomical applications, thanks to its direct bandgap, tunable by varying the chemical abundance of Hg and Cd and, thus, reflecting in the cut-off wavelength tuning. The first fact, i.e. the direct bandgap, yields a high absorption coefficient in the IR domain (this is the main reason for its just mentioned high quantum efficiency). Other figure of merit is the pretty high operating temperature in LWIR applications, higher than extrinsic Si. For these reasons it has been possible up to now to produce low-cost and high-performances HgCdTe photodetectors, also in terms of wafer surface, composition and doping stability.

In 1986 Del Alamo and Swanson [6] began to investigate the quantum process of tunneling in HgCdTe, merging theories from Chynoweth and Sommers [7, 8]. Then several studies [9, 10] tried to model the behavior of a $p$-$n$ or a metal-insulator-semiconductor structure at low temperature and in reverse biased conditions, from which it comes out an evident dependence of $I$-$V$ characteristic on tunneling effect. Actually, besides other known intrinsic phenomena like generation/recombination, diffusion or Auger processes, tunneling seems to play a crucial role in the current production mechanism.
One of the most important factors affecting the tunneling in Hg$_{1-x}$Cd$_{x}$Te is the width of the energy gap. It obviously depends proportionally on some external factors like the temperature but it should be noticed that, as we will see in the following, light-holes mass value dominates tunneling between valence band and SRH centers because of the exponential dependence of the process on the $m^*$ decrease. This is a crucial aspect because in HgCdTe the conduction band and light-hole masses are approximately equal so the maximum tunnel probability occurs for midgap states: this explains the greater contribution of deep impurity compared to the shallow one in the tunnel mechanism. Therefore the energy of the traps (as their concentration or the capture cross section) is an important parameter to properly compute the tunneling rates, especially with regard to the gap edges, as we have just seen. For this reason it is even more important to extend the analysis to the full-band dispersion structure, in order to account for the overall envelope of tunneling probability. The trend of energetic bands, and particularly the width of the energy gap, can also determine the tendency of an HgCdTe-based device to detect a particular frequency range. Tending to detect the radiation from SWIR to LWIR, and at a very low temperature, the band gap is further reduced leading to the rise of tunneling effect. But we can prevent this avoiding the formation of traps near the midgap (for example by controlling the procedure - epitaxy, MOCVD or MBE - used to grow the material) or with the reduction of the electric field. The latter, unfortunately, produces other negative consequences in the $I$-$V$ characteristic and – worse – at room temperature noise is enhanced: at 100 K the dynamical resistance becomes to drop and, as the temperature increases, also shot noise and Johnson noise are more noticeable [11]. One of the responsible process of those noise-phenomena is the Auger generation/recombination: a three particle temperature-dependent mechanism that involves electrons and holes moving among bands, directly or via traps. The solution is to operate at low temperature (by using cryogenic cooling systems) and at voltage that represents a compromise between tunneling and Auger process.

More recent studies show that an optimized photodetector should consist of a lightly doped active region, with passivation and insulation of surfaces in order to prevent a further generation of carriers [12]. Several structures have been proposed but, basically, the most effective device seems to be a double junction $p^-\nu-n^+$ where $\nu$ represents the HgCdTe low-doped active region [13]: under reverse bias, the holes are removed from the $\nu$ absorber region through the $p^-\nu$ extraction junction more rapidly then they can replenished from the $n^+$ layer through the $\nu-n^+$ exclusion junction, as the bias increases. This means that the hole concentration decreases below its thermal equilibrium, thus also the electron concentration is reduced. The overall result is the drop of Auger net recombination rate in the active region

$$R_A = (C_n n + C_p p) (np - n_i^2)$$

(1.2)

(where $C_{\{n,p\}}$ are the Auger recombination coefficients) with a noise suppression mechanism in the $I$-$V$ characteristic. A very noticeable goal. This suppression means that at a certain temperature the Auger process is less prevalent with respect to traditional devices, leading to a higher operating point of this particular double-junction, up to about 145K for LWIR applications [13], and eliminating the requirement of a cooling system. Moreover the higher temperature leads, in turn, to a wider bandgap and then to a smaller tunneling rate.

Several other mechanisms are reported in literature as a source of dark current in HgCdTe-based photodiodes like surface state current, thermal diffusion of minority carriers, ohmic current and impact ionization from high energy carriers but at this moment
only tunneling and Auger process seem to be the major limiting factors of device performance. In order to better understand these phenomena we will analyze in the following paragraphs some of the most important physical features by both the solid state and electronics standpoint, namely the crystal structure of HgCdTe, its band structure and transport properties. Finally we will give a short demonstration of how defects in this material are important in the general tunneling framework.

### 1.2.2 Crystal Structure and its Growth Processes

When we refer to the Hg$_{1-x}$Cd$_x$Te we are talking about a mixed crystal made by two II-VI compounds, CdTe and HgTe, where $x$ and $(1 - x)$ represent the respective molar fractions while 6.48Å and 6.46Å are their lattice constants. Even if this parameter is precisely determined by the growth process (it is needed a tolerance of $\Delta x < 0.003$ to ensure a variation in cut-off wavelength for an LWIR HgCdTe-based photodetector less than 1 $\mu$m [14]) the occupation site of one of these two compound within the material in a given lattice position is almost random. Thanks to the visible mismatching between the CdTe and HgTe cells, they can form an HgCdTe zincblende structure, with a lattice parameter function of the molar fraction as

$$a_0(x) = 6.4614 + (8.4x + 11.68x^2 - 5.7x^3) \times 10^{-3},$$  \hspace{1cm} (1.3)

composed by two interpenetrating face-centered-cubic cells (“fcc”) displaced by a vector $(\frac{a_0}{4}, \frac{a_0}{4}, \frac{a_0}{4})$. In this arrangement each Cd$^{+}$- or Hg$^{+}$-ion occupies the sites of one lattice and Te$^{-}$-ions that one of the other lattice. So every unit cell contains a net contribution of four atoms, two cations and two anions, determining five possible configurations of Cd and Hg as neighbor sites of Te: four Hg-ions, four Cd-ions, two Cd-ions and two Hg-ions, three Cd-ions and one Hg-ion or three Hg-ions and one Cd-ion. Regarding the electronic structure we have six valence electrons for the Te (their configurations are $5s^2$ or $5p^4$) and two both for Hg and Cd (respectively $6s^2$ and $5s^2$). The resulting chemical bonds – located along tetrahedral directions – are mainly covalent, but with a ionic component due to a net electron transfer between Cd- or Hg- and Te-ions that, in turns, depends on a deeper potential of the latter. This fact occurs despite Te has more electrons than Cd and Hg, because the lowest unfilled states in Te are less energetic then the filled states in the other two ions (a more complete discussion about the energetics in HgCdTe will be done in the next paragraph).

It must be said that there may occur some deviations in the situation depicted in Figure 1.5 - a): firstly they can be represented by changes in the average composition in the range of microns due to statistical fluctuations; secondly we can find also short-range deviations. Let’s consider the Bernoulli distribution $Z$ as a function of the molar fraction $x$ for all the possible crystal configurations [3]:

$$Z(x) = \binom{4}{i} x^i (1 - x)^{4-i},$$  \hspace{1cm} (1.4)

where the index $i = 0, 1, 2, 3, 4$ indicates one of the five arrangements above mentioned. When we have to study the cation-cation interaction, small but non-negligible, the probabilities to find Cd-Cd or Hg-Hg couples of nearby atoms are respectively:

$$P_{\text{Cd-Cd}}(x) = x + \beta(1 - x)$$  \hspace{1cm} (1.5)

and

$$P_{\text{Hg-Hg}}(x) = (1 - x) + \beta x,$$  \hspace{1cm} (1.6)
1.2 Review on some Physical Aspects in HgCdTe

Figure 1.5. Left: schematic representation of the HgCdTe zincblende “fcc” structure with lattice constant $a_0$ and (Right) the truncate octahedron of its first Brillouin zone, where the standard symmetry points are annotated. The $\Gamma$ point ($k = 0$) represents the center of the cell in the reciprocal space. Adapted from [3].

where $\beta$, the so-called short-range order parameter – SRO, lies in the interval between $-1$ and 1 and depends on this expression of neighboring energies:

$$
\varepsilon = \frac{2E_{Hg-Cd}}{E_{Hg-Hg} + E_{Cd-Cd}}.
$$

(1.7)

In the following table we resume all the possible situations depending on the parameter values:

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$\beta$</th>
<th>configurations allowed/enhanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>only $i = 0$ (HgTe) or $i = 1$ (CdTe) allowed</td>
</tr>
<tr>
<td>$&lt; 1$</td>
<td>$&gt; 0$</td>
<td>$i = 0$ or $i = 1$ populations enhanced</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>random fluctuations described in (1.4)</td>
</tr>
<tr>
<td>$&gt; 1$</td>
<td>$&lt; 0$</td>
<td>$i = 2$ (both HgTe and CdTe) population enhanced</td>
</tr>
<tr>
<td>$\gg 1$</td>
<td>$-1$</td>
<td>ordered crystal</td>
</tr>
</tbody>
</table>

Table 1.2. Different types of lattice configurations in HgCdTe alloys described in terms of energy (through the parameter $\varepsilon$), probability of cation-cation neighboring bonds and final configuration (adapted from [3]).

When Lawson and his co-workers prepared the first sample of HgCdTe alloy, in 1959, they did grow separately two different crystals, of CdTe and HgTe. Successively they mixed them in a vacuum chamber heated slightly above the melting temperature. Nowadays there are essentially four primary methods to synthesize directly the HgCdTe from its constituents: bulk crystal grow, liquid phase epitaxy (LPE), molecular beam epitaxy (MBE) and vapor phase epitaxy (VPE).

Among the bulk techniques we find the well known Bridgman method, that through a furnace with a certain temperature gradient grows the structure by slowly shifting it within this gradient. This movement provides the progressive crystallization at the solid/liquid
interface, starting from a seed of nucleation. A second method is the *solid state recrystallization*, which exploits a stoichiometric mixture of Hg, Cd and Te combined firstly near the Te-HgTe eutectic point (∼970 K) and then, with a sudden (perfectly controlled) rise of the temperature above the melting point (>1050 K), HgCdTe begins to grow spontaneously. Finally the furnace sees a rapid quench cooling that provides a very fast (∼250 μm/hr) crystallization. We can reach a precision of Δx < 0.002.

The LPE, and in general all the epitaxial techniques, offer lower growth temperatures with respect to the bulk ones. Thanks to the reduced energy involved in the process it is possible also to reach higher levels of homogeneity with relatively short growth times (5-60 μm/hr) and the inclusion of dopant atoms (if any) in the grower. The process is based on the precipitation of a saturated solution (containing the growing constituents) on the substrate surface. The molar fraction is well pre-determined by the abundance of each reagent and by the process temperature (often about 620–820 K) with a Δx < 0.002.

For the production of HgCdTe thin films, instead, the most used technique is the MBE (low growth rate: 1-5 μm/hr). In our case the sources are three and are represented by CdTe, Te₂ and Hg, while the substrates can be CdZnTe, Si or GaAs (with a ZnTe buffer layer). The sticking coefficients of these three precursors (and obviously the reaction temperature, less than 470 K) will determine the correct molar fraction (0.0006 < Δx < 0.01).

The last category of growing processes is represented by the VPE. The first method, invented in 1968, was the ISOVPE, a quasi-equilibrium technique in which HgTe is transported (670-870 K) from the source (HgTe or Hg₁₋ₓCdₓTe) to the substrate by evaporation-condensation mechanisms. Nowadays one of the most used is the MOCVD (Metal-Organic Chemical Vapor Deposition): at about 620 K the substrate (often CdZnTe, sapphire or GaAs) is put ion contact with some precursors (Hg₁(ₚₑₜ) and, for instance, Dimethylcadmium – DMCd and diisopropyltelluride – DIPTe or diethyltelluride – DETe) in a non-equilibrium conditions are carried by H₂ into the reactive chamber, providing a powerful method in which it is possible to dynamically control growth factors, presence of dopants or preparation and passivation of surfaces/interfaces. The process is not so fast (2-10 μm/hr). In this technique the precision in the final composition, pretty similar to Δx_{MBE}, is given by some factors, like the dynamics of the adsorption mechanisms, the flux of carrier and precursor gases and the mass of the compounds introduced in the reaction chamber.

### 1.2.3 Band Structure

The lattice characteristics, its symmetry and chemical configuration, with the particular value of aₒ(x), the temperature and, especially, the molar fraction x itself are the most powerful parameters driving the trend of the energetic dispersion E(k) in the reciprocal space of the Hg₁₋ₓCdₓTe alloy family.

We know from Quantum Mechanics that a single (i.e. ideally) isolated atom is characterized by electrons with well-determined energetic levels. This quantization derives from the eigenvalue solution of the Schrödinger’s equation

\[
-\frac{\hbar^2}{2m} \Delta_r \Psi + U(r)\Psi = E\Psi
\]  

(Δ ≡ ∇² is the Laplace differential operator) that provides a discrete energy spectrum for bound states. But when different atoms are arranged in a lattice structure the energetic levels are modified in order to fulfill the Pauli exclusion principle and the new nature of the term U(r): being no longer the only Coulomb attraction of the nucleus towards
the electrons, it represents now the lattice potential, expression of the interaction of all the atoms forming the crystal, and usually treated as a perfectly periodic term. This periodicity of $U(r)$ has a crucial role in the Solid-State Physics. In fact, according to the *Bloch Theorem*, every valence electron in the solid responds to the lattice potential with a superposition of wavefunctions like

$$\Psi_k(r) = u_k(r)e^{i\mathbf{k} \cdot \mathbf{r}}$$  \hspace{1cm} (1.9)$$

that form a complete set of basis in the momentum space and where the subscript $k$ indicates that it is possible to label univocally each $\Psi$ with a particular wavevector of the first Brillouin zone (as if $k$ were a true quantum number). The (1.9) states that the electrons are assumed as plane waves moving in the solid with a periodicity modulated by the spatial term $u_k(r)$, which has the same period of $U(r)$ (remember that for free electrons, in general, $U$ is a constant term). In this regime the electrons lose their quantization energy and the levels typically degenerate into a pattern so dense of levels that we can define “band”: as a matter of fact, single atom’s valence electrons cannot be thought of as belonging to a particular atomic site, but they behave as belonging to the whole periodic crystal, therefore feeling a periodic potential $U(r)$. The description of bands is provided by the *Energy Band Theory*, that merges the Quantum Mechanics and the Solid-State Physics in a unique model able to predict the behavior of a crystal by the electronic standpoint. Without enter in the details of this theory we can say that combining the Bloch wavefunction with the potential (typically in literature the Kronig-Penney potential is used because of its simplicity) we obtain a set of periodic solutions for the energy $E(k)$ characterized by real or pure imaginary wavevectors $k$. The regions connoted by a real wavevector represent allowed states (in the plain $E \times k$), while the imaginary $k$ belong to forbidden energy states which correspond to attenuated wavefunctions in the crystal. This fact, combined with the periodic boundary conditions, generates the forbidden bandgap in semiconductors and the band shape as well.

There are many methods used to model the trend of the bands in solids and all of them are only approximations of the true problem. This one should consist in solving the Schrödinger’s equation for a particular lattice configuration. In fact the energy eigenvalues can be correctly obtained without any approximation by substituting the potential $U(r)$ in the Hamiltonian operator of the system and then solving the (1.8) for a given wavefunction (1.9). Despite this is the most efficient and rigorous way, such a process could be too expensive and complicate, in a computational sense. In fact typically it is difficult to describe the real shape of the potential, hence some models have been elaborated to overcome these difficulties, such in the case of the *Virtual-Crystal Approximation – VCA*, developed because the fluctuations in the $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ structure previously discussed generate the absence of a true potential periodicity, a necessary requirement for the Bloch treatment. In the VCA framework we make use of a virtual potential, weighted on the Cd and Hg potentials, of the form

$$U_{VCA} = xU_{Cd} + (1-x)U_{Hg}, \hspace{1cm} (1.10)$$

where $x$ is always the molar fraction. But unfortunately it has been proved that VCA fail just in the case of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ away from the critical points in the reciprocal space [3], nonetheless there are small but important effects as well. A possible solution is represented by the *Coherent Potential Approximation – CPA*, an effective but high-computational-cost numerical solution.

For such reasons there exist several theoretical approaches in the determination of the band structure that essentially differ in the way in which they approximate the shape
of the potential. Two of the most important are the *tight-binding approximation* and the \( \mathbf{k} \cdot \mathbf{p} \) *perturbation method*: the first one describes the system accounting for only particular selected atomic orbitals and only for nearby atoms (it puts to zero the hopping parameter for atoms not first-neighbors), while the second one exploits an expression of the Hamiltonian which contains some perturbative terms. But before to examine the latter of them let’s explain some intuitive concepts.

Without even going into details, a first source of energetic levels degeneration for valence electrons (or holes) comes from the theory of orbitals. In direct gap semiconductors, typically, the atoms are connected through the presence of \( \sigma \)-bonds, i.e. with a superposition of \( p \)-orbitals. So, the expression of the spatial term \( \psi_{\mathbf{k}}(\mathbf{r}) \) and, thus, of the probability density \( |\psi_{\mathbf{k}}(\mathbf{r})|^2 \) is affected by the shape of these orbitals. Since in truth our orbital is constituted by a three-fold \( sp^3 \) hybridization (there are \( p \)-orbitals lying on the three directions of the real space: \( p_x, p_y, p_z \)) this means that we must expect an energy differentiation for valence electrons in at least three sub-bands. Moreover, if we consider non-negligible the spin-orbit interaction, that is the interaction between the intrinsic electron angular momentum (the spin \( s \)) and the magnetic field generated by the nuclei or, in other words, the overlapping \( (s \times \mathbf{L}) \) between the spin and the nucleus momentum \( \mathbf{L} \) (almost equal to the electron orbit momentum), the splitting along the energy axis for the states occupied by these electrons (or holes) become more evident.

![Simplified band structure](image)

**Figure 1.6.** Simplified band structure for a direct gap semiconductor at the point \( \Gamma \) (\( \mathbf{k} = 0 \)) of the first Brillouin zone. The bottom of the conduction band is often called \( \Gamma_6 \), while the top of the heavy hole and light hole valence bands is commonly named \( \Gamma_8 \). Finally \( \Gamma_7 \) is the top of the spin-orbit SO valence band. There are two typical gaps in this structure: the first is the well-known \( E_g \), i.e. the amplitude of the classical forbidden region, then there is also the gap \( \Delta \) defined as \( \Delta = \Gamma_8 - \Gamma_7 \).

A similar argumentation can be made for the conduction state of carriers. Since in this case the spatial term of \( \Psi \) is dominated by the shape of the \( s \)-orbitals (with almost spherical symmetry) we can conclude that, combining the \( s \)-orbitals of two nearby atoms,
the conduction band of the semiconductor can be obtained from the degeneration of these orbitals without, in a first approximation, any sub-band splitting: in fact in this case there in not spin-orbit interaction if we suppose that electrons have no orbital momentum \( \mathbf{L} \). The only degeneration, at this stage of approximation, could be a slight contribution of the \( p \)-orbital to the conduction band. But as long as we assume \( \mathbf{L} \simeq 0 \) the \( s-p \) two-fold splitting in this case will be inappreciable.

In II-VI direct gap semiconductors like Hg\(_{1-x}\)Cd\(_x\)Te the extreme values of both valence and conduction band are located at the critical point \( \Gamma \) of the first Brillouin zone (for \( k = 0 \)). As we have already seen, the top of the valence band is made up of \( p \)-symmetry states (mostly \( p \)-Te ones), including also the spin-orbit SO term. The \( s \)-symmetry at the bottom of conduction band is mainly due, instead, to \( s \)-Cd and \( s \)-Hg states and for this reason here it is two-fold degenerated. By the way, we will neglect this feature. Let’s analyze through the \( \mathbf{k} \cdot \mathbf{p} \) perturbation method how these bands are generated.

This method is probably the most useful to describe the energetic bands in narrow-gap semiconductors near \( \Gamma \) and the other symmetry points and it starts from the substitution of the Bloch wavefunction into the Schrödinger’s equation:

\[
\left( -\frac{\hbar^2}{2m} \Delta_r + \frac{\hbar}{m} \mathbf{k} \cdot \mathbf{p} + \frac{\hbar^2 k^2}{2m} + U(r) \right) u_{n,k}(r) = E_{n}(k) u_{n,k}(r) \tag{1.11}
\]

where \( k \equiv |\mathbf{k}| \), \( \mathbf{p} = -i\hbar \Delta \) is the momentum operator and the subscript \( n \) represents the band quantum number. Now we suppose to know the Hamiltonian and the wavefunction at an arbitrary \( \mathbf{k}_0 \):

\[
H_{k_0} = -\frac{\hbar^2}{2m} \nabla^2_r + \frac{\hbar}{m} \mathbf{k}_0 \cdot \mathbf{p} + \frac{\hbar^2 k_0^2}{2m} + U(r) \tag{1.12}
\]

then

\[
\left( H_{k_0} + \frac{\hbar}{m} (\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{p} + \frac{\hbar^2}{2m} (k^2 - k_0^2) \right) u_{n,k}(r) = E_{n}(k) u_{n,k}(r) \tag{1.13}
\]

and

\[
u_{n,k}(r) = \sum_{n'} C_{n',n} (|\mathbf{k} - \mathbf{k}_0|) u_{n',k}(r) \tag{1.14}
\]

If we substitute the (1.14) into the (1.13), multiply both sides by \( u_{n,k_0}(r) \) and integrate over the volume of a unit cell we obtain a secular equation for eigenvalues:

\[
\sum_{n'} \left[ \left( E_{n}(k_0) + \frac{\hbar^2}{2m} (k^2 - k_0^2) \right) \delta_{n',n} + \frac{\hbar}{m} (\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{p}_{n',n} \right] C_{n',n} = E_{n}(k) C_{n,n} \tag{1.15}
\]

where

\[
\mathbf{p}_{n',n} = \int u_{n,k_0}^*(r) \mathbf{p} u_{n',k_0}^*(r) \, dr. \tag{1.16}
\]

The method consists, hence, in considering \( \frac{\hbar}{m} (\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{p}_{n',n} \) as a perturbative term when we want to solve the secular equation (1.15) for \( k \simeq k_0 \), determining the coefficients \( C_{n',n} \) and then the eigenvalues of the system with the well-known quantum perturbation theory.

Now we want to build a more complete expression of the Hamiltonian, taking into account also relativistic (i.e. SO) terms, because of the presence in Hg\(_{1-x}\)Cd\(_x\)Te of a high atomic number of Hg, Cd and Te that provides a relativistic modification in the electron
orbital motion. Our Hamiltonian will be in the form:

\[ H = H_0 + H_{mv} + H_D + H_{SO} \]

\[ = \left( -\frac{\hbar^2}{2m} \Delta_e + U(r) \right) - \left( \frac{E - mc^2 - U(r)}{4m^2c^2} \hbar^2 \Delta \right) - \left( \frac{\hbar^2}{4m^2c^2} \nabla U(r) \cdot \nabla \right) + \]

\[ + \left( \frac{\hbar^2}{4m^2c^2} (\nabla U(r) \times \mathbf{p}) \cdot \mathbf{\sigma} \right) \]

where the first term \( H_0 \) is the non relativistic contribution, \( H_{mv} \) is the mass-velocity interaction, \( H_D \) the Darwin interaction, \( H_{SO} \) the spin-orbit term (which includes also the \( \mathbf{k} \cdot \mathbf{p} \) perturbation) and \( \mathbf{\sigma} \) represents the spin operator. Solving the secular equation (1.15) with the Hamiltonian reported in (1.17) means finding the eigenvalues of the problem and then predict the trend of the energetic bands (i.e. compute the value of the energy \( E \) for each wavevector \( \mathbf{k} \) of the reciprocal space in the first Brillouin zone).

Once we solved the system for the Hg_{1-x}Cd_xTe we will obtain more than one expression for the energy dispersion \( E(\mathbf{k}) \), accordingly to the degeneration of bands previously discussed. As represented in Figure 1.6 for a generic narrow- direct-gap semiconductor as in our case we have a valence band splitted into three sub-bands: one named heavy-hole band (HH) where, since the effective mass of the carriers is proportional to the curvature of the energy dispersion \( E \) around \( \Gamma \) point (we neglect here the tensorial notation) through

\[ \frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2E}{dk^2}. \]  

the low concavity of this sub-band around the \( \Gamma \) point yields a pretty high \( m^*_{HH} \); therefore we find a light-hole band (LH) determined, instead, by a lower hole mass \( m^*_{LH} \); finally there is also a spin-orbit split-off band (SO) due to the relativistic effect between orbital momentum and spin yet discussed. The conduction band is, as predicted, represented by a unique non-degenerate band. Here we report the various dispersion equations for each band (see also Figure 1.6) in a first-order perturbative framework [14]:

\[ E_C(k) = \frac{E_g(x)}{2} \left( 1 + \sqrt{1 + \frac{8k^2Q^2}{3E_g^2(x)}} \right) + \frac{\hbar^2k^2}{2m_0} \]  

(1.19)

\[ E_{HH}(k) = -\frac{\hbar^2k^2}{2m^*_HH} \]  

(1.20)

\[ E_{LH}(k) = -\frac{E_g(x)}{2} \left( 1 - \sqrt{1 + \frac{8k^2Q^2}{3E_g^2(x)}} \right) + \frac{\hbar^2k^2}{2m_0} \]  

(1.21)

\[ E_{SO}(k) = \Delta(x) + \frac{\hbar^2k^2}{2m_0} - \frac{\hbar^2Q^2}{3(E_g^2(x) + \Delta(x))} \]  

(1.22)

where \( Q \) is the so-called momentum matrix element, \( m_0 \) the free-electron mass and \( \Delta(x) = E_{HH}(\Gamma, x) - E_{SO}(\Gamma, x) \). Possible values for these parameters are \( Q = 8.4 \cdot 10^{-8} \) eV and \( \Delta(x) = 1.08 - 0.12x \) eV. It’s worth noting, at first, that there are some parameter \( x \)-dependent, such the energy gap and the parameter \( \Delta \). This means that the configuration of the bands plotted in Figure 1.6 should change with respect to the molar fraction. Moreover it is also evident from the Figure that in Hg_{1-x}Cd_xTe is \( m^*_{LH} \approx m^*_C \) (where \( m^*_C \) is the effective electron mass in conduction band that dominates, with \( m^*_HH \), the generation/recombination processes for this alloy). Moreover, typically, is \( m^*_HH \approx 0.55m_0 \).
Especially $m_C^*$ is a physical quantity with a crucial role and by analyzing in detail the equation (1.19) – which is an approximated solution for a narrow-gap and small $k$ – we can distinguish two cases: when $E_g^2 \gg \frac{8}{3} k^2 Q^2$ we have the relation

$$m_C^* \simeq \frac{\hbar^2 E_g}{2 Q^2} \quad (1.23)$$

and, moreover, the dispersion is almost parabolic

$$E_C(k) \sim k^2; \quad (1.24)$$

when $E_g^2 \ll \frac{8}{3} k^2 Q^2$, instead, we obtain

$$E_C(k) \simeq E_g + \frac{\hbar^2 k^2}{2 m_0} + \sqrt{\frac{2}{3}} Q k \quad (1.25)$$

that is almost linear for intermediate/large wavevectors w.r.t. the critical point $\Gamma$ (see the second picture in Figure 1.7). This case yields

$$m_C^* \simeq m_0. \quad (1.26)$$

Figure 1.7. Schematic representation on how the band structure changes in the Hg$_{1-x}$Cd$_x$Te alloy by varying the molar fraction $x$. The two dashed lines represent a linear approximation of the bandgap tuning, from the semi-metal state in HgTe to the semiconductor one of CdTe (from [3]). Note that in the condition of very narrow-gap (i.e. for $x$ near the value $x_0$) the dispersion of bands $\Gamma_6$ and $\Gamma_8$ becomes pretty linear.

A general relationship can be also obtained through experimental fitting, yielding [3]:

$$\frac{m_0}{m_C^*} = \frac{1}{0.05966} \frac{E_g(x) + 0.667}{E_g(x)(E_g(x) + 1)}, \quad (1.27)$$

which is a slight modification of the previous (1981) Weiler formula [15]

$$\frac{m_0}{m_C^*} = 1 + 2 F + \frac{E_g(x)}{3} \left( \frac{2}{E_g(x)} + \frac{1}{E_g(x) + \Delta} \right), \quad (1.28)$$
1.2 Review on some Physical Aspects in HgCdTe

Figure 1.8. Electronic structure of Hg$_{0.7}$Cd$_{0.3}$Te computed by using the nonlocal empirical pseudopotential method (from F. Bertazzi, M. Moresco, M. Penna, M. Goano and E. Bellotti, “Full-band Monte Carlo simulation of HgCdTe APDs”, J. Electr. Mat. 39 (2010)). Note the direct bandgap at Γ points.

Figure 1.9. The $x$- and $T$-dependence of the energy gap for different Hg$_{1-x}$Cd$_x$Te alloys, depicted in the same plot through the fitting curves of experimental data (adapted from [3]).
where $F = -0.8$, $E_p = 19$ eV and $\Delta$ is assumed equal to $1$ eV. Again we find a dependence of the band parameters by the molar fraction $x$. The most evident result of this phenomenon comes from the variation of the energy gap as a function of $x$ (see Figure 1.7). In HgTe ($x = 0$) the extremal edge of the conduction band $\Gamma_6$ lies at a less energetic level than $\Gamma_8$: since they are inverted in the energy axis the $E_g$ is negative, yielding a semi-metal alloy. In the case of CdTe we find, instead, a situation already seen in Figure 1.6. Between these two configurations there are all the possible arrangements of the Hg$_{1-x}$Cd$_x$Te and, thus, all the possible values of $x$ from 0 to 1. At a certain value ($x_0 \approx 0.15$ at $T = 77$ K and $x_0 \approx 0.1$ at $T = 300$ K) the bands $\Gamma_6$ and $\Gamma_8$ change the sign of their curvature determining the transition point between the semiconductor state ($x > x_0$), where the gap opens, and the semi-metal one ($x < x_0$).

Many experimental models have been developed for the function $E_g(x)$, based upon both theoretical studies and experimental measurements. Since the late 1960s several scientists proposed different methodologies to compare, directly or indirectly, the almost linear theoretical trend of $E_g$ as a function of $x$ for instance by measuring interband recombination rates in photo luminescence experiments, photovoltaic/photoconductive spectral response of of Hg$_{1-x}$Cd$_x$Te devices or through optical absorption coefficient $\alpha$ measurements. In fact [16]:

$$\alpha(x) \approx \frac{\sqrt{2}c}{\tau_e} \sqrt{1 - \frac{\lambda}{\lambda_c(E_g(x))}} \left( \frac{m^*_c}{m_0 \hbar} \right)^3$$

(1.29)

where $\alpha$ is measured in m$^{-1}$, $\tau_e$ is the electron lifetime, $\lambda_c$ the cut-off wavelength [14]:

$$\lambda_c(E_g(x)) \approx \frac{1.24}{E_g(x)}$$

(1.30)

One of the most reported models in literature for $E_g(x)$ is due to Hansen et al. (1982) and comes from magneto-optical experiments at low temperature [17]:

$$E_g(x, T) = -0.302 + 1.93x - 0.81x^2 + 0.832x^3 + 5.35 \cdot 10^{-4} (1 - 2x) T$$

(1.31)

where also the temperature dependence is expressed. In 1990 Seiler [18] and his co-workers developed an improved version of the previous formula, resulting

$$E_g(x, T) = -0.302 + 1.93x - 0.81x^2 + 0.832x^3 + 5.35 \cdot 10^{-4} (1 - 2x) \left( \frac{-1822 + T^3}{255.2 + T^2} \right)$$

(1.32)

valid for $0.2 < x < 0.3$.

### 1.2.4 Optical and Electric Properties

The density of states in both bands and, thus, the carrier concentration are in semiconductor photon detectors heavily affected by generation and recombination processes. These mechanisms, combined with the transport characteristics of the material and the optical and thermal excitation processes (already seen in the previous paragraphs) determine the behavior of the photo-generated signal. Before analyzing the possible sources of noise, like in the case of tunneling, now we briefly discuss the most important HgCdTe properties under the optical and electronic standpoint and, finally, the principal generation/recombination processes affecting HgCdTe-based infrared photodetectors.
A – Mobility and Intrinsic Concentration

One of the most important transport properties in semiconductors is the carrier mobility. We have already mentioned the behavior of the carrier effective masses near the symmetry point $\Gamma$ of HgCdTe and, furthermore, we saw all the relevant features concerning the behavior of the energy gap (and band shapes) as a function of the molar fraction. So, thanks to the small $m^*_{C}$ values, mobilities are pretty high in this alloy, and electrons have a mobility two or three orders of magnitude greater than the heavy-hole one. This is particularly true, in general, when the forbidden gap is very narrow or in the vicinity of the famous value $x_0$, i.e. the composition at which occurs the transition between the semi-metal state and the semiconductor state. In that condition, as previously demonstrated, the dispersion relations for conduction and LH bands are almost linear, yielding low masses and, finally, high mobilities.

From this general introduction it is already clear that also mobilities are $x$-sensitive parameters thanks to their dependence on $E_g$, that is a function of $x$ itself. By the physical standpoint, instead, mobilities are determined by the competition of some scattering processes. In HgCdTe the $\mu_e$ is mainly due to the scattering with ionized impurities (especially in the low-temperature regime, i.e. less than some decades of K) and polar longitudinal optical (LO) phonons, which dominate at high temperatures. Other sources of mobility reduction are represented, for instance, by lattice defects.

Since GR processes are governed by electrons and heavy-holes, $\mu_h$ is almost confusable with $\mu_{HH}$, and being $\mu_{HH} \ll \mu_e$ the hole mobility is less studied than the electron one. Nonetheless, typically for IR photodetectors it is assumed [14]:

$$\mu_h(T, x) \simeq \frac{\mu_e(T, x)}{100}$$

Furthermore, it must be noticed that $\mu_{HH}$, as in the case of $\mu_e$, is dominated by scattering processes according to the lattice temperature: we have scattering with dislocations up to 50 K, with ionized impurities for intermediate $T$ and polar phonon scattering above 200 K. The $\mu_{LH}$, instead, is dominated by the acoustic phonon scattering. By the practical point of view $\mu_h$ at room temperature is between 40 and 80 cm$^2$V$^{-1}$s$^{-1}$, while at 77 K it is almost one order of magnitude higher.

There are many empirical expressions for $\mu_e$ and this is partly due to the large number of parameters in Hg$_{x-1}$Cd$_x$Te alloys affecting this physical quantity. In the composition range $0.2 \leq x \leq 0.6$ and for $T > 50$ K it is often exploited the law [1]

$$\mu_e(T, x) \simeq \frac{9 \cdot 10^8 \cdot \left(\frac{0.2}{x}\right)^{7.5}}{T\left[2\left(\frac{0.2}{x}\right)^{0.6}\right]}$$

while at $T = 300$K it approximates to

$$\mu_e(x) \simeq \frac{1}{8.754 \cdot 10^{-4}x - 1.044 \cdot 10^{-4}}.$$  \hspace{1cm} (1.35)

Below 200 K and for $0.29 \geq x \geq 0.285$ the electron mobility has been measured in the range $2 \cdot 10^4$ to $4 \cdot 10^3$ cm$^2$V$^{-1}$s$^{-1}$, while at 77 K (and $x = 0.2$) it rises to $2 \cdot 10^5$ cm$^2$V$^{-1}$s$^{-1}$.

Also the minority carrier mobility in doped HgCdTe samples is an important parameter under the transport standpoint. For $p$-type materials $\mu_e$ lies in the same order found for $n$-type one, until the acceptor concentration remains $< 10^{15}$ cm$^{-3}$. As the latter increases the electron mobility drops.
1.2 Review on some Physical Aspects in HgCdTe

Figure 1.10. Electron mobility in HgCdTe measured in various experiments and plotted as a function of the molar fraction and for two different temperatures: 77 K and 300 K (from [1]). Solid lines represent the theoretical curves, computed for the electron concentration of $2 \cdot 10^{15}$ cm$^{-3}$ at 4.2 K.

Finally we report the most exploited approximated formula for the intrinsic concentration, derived from the $k \cdot p$ theory [19]:

$$n_i(T, x) \simeq (A + B x + C T + D x T + F x^2 + G T^2) \cdot 10^{14} \sqrt{E_g^3 T^3} e^{-\frac{E_g}{2k_BT}} \tag{1.36}$$

where $n_i$ is in cm$^{-3}$, $A = 5.242566$, $B = -3.5729$, $C = -4.74019 \cdot 10^{-4}$, $D = 1.25942 \cdot 10^{-2}$, $F = 5.77046$ and $G = -4.24123 \cdot 10^{-6}$.

B – Carrier Lifetime and GR Processes

In solids, and in particular in semiconductors, the description of equilibrium is generally expressed in terms of carrier density. This state, its perturbations and how the system reacts restoring the initial equilibrium condition represent the basis of the electronic behavior of every semiconductor device. In the particular case of photodetectors, almost all the operation mechanisms arise from those non-equilibrium phenomena in which an electron-hole pair is created by converting the thermal or absorbed energy in a so-called generation process or, inversely, the pair is annihilated (the recombination) through, for instance, radiative processes. The generation/recombination mechanism changes the availability of carriers, both majority and minority, in the conduction and valence band determining relaxation processes towards equilibrium or perturbation ones from the equilibrium.
Moreover, they affect an important physical quantity of the material that will be analyzed, i.e. the carrier lifetime, a concept linked to the net GR rate. A precise description of generation/recombination is not possible except in terms of statistical mechanics, since the exact time and position at which a phenomenon can occur is not well predictable. We will give in the following a detailed insight of the physical processes and its formalism used in solid-state physics of semiconductors. Those concepts will become very useful to the comprehension of how it is possible to model and simulate such processes, but at the moment we confine ourselves to a mere phenomenological description of what happens outside equilibrium in semiconductors such as HgCdTe.

There are essentially three ways according to which GR processes can take place: if an electron drops directly from the CB into an unoccupied state in VB within a characteristic lifetime $\tau$, and emitting a photon with energy $\geq E_g$, one speaks of radiative direct band-to-band recombination process (where the inverse process represents the corresponding generation); secondly if in the bandgap there is an impurity-related energy level the transition is mediated by this level and, thus, also enhanced resulting in a shorter lifetime; this is the Shockley-Read-Hall recombination, where a photon with energy equal to the difference between the energies of the electron and the hole captured by the impurity level can be emitted (as before, the inverse stands for generation). Alternatively, this energy can be dissipated within the lattice in the form of phonons or imparted as kinetic energy to a third carrier, determining in the latter case the so-called Auger process (also here, the mechanism has two directions, determining both Auger recombination and generation).

If at equilibrium the two competitive phenomena of generation and recombination pair are perfectly balanced, in presence of external influences on the semiconductor (such as an electric field combined with lighting conditions) the resultant net rate $U = G - R$, i.e. the difference between the generation rate $G$ and the recombination one $R$, is a function of the equilibrium carrier densities $n_0$ and $p_0$ but also of the non-equilibrium densities $n' = n_0 + \Delta n$ and $p' = p_0 + \Delta p$. The discrepancies $\Delta n$ and $\Delta p$ are the quantities responsible of an $U \neq 0$. In the relaxation time approximation for the minority carriers we have:

$$
\begin{align*}
U_n &\approx \frac{n' - n_0}{\tau_n} = \frac{\Delta n}{\tau_n} \\
U_p &\approx \frac{p' - p_0}{\tau_p} = \frac{\Delta p}{\tau_p}
\end{align*}
$$

(1.37)

then, from the Boltzmann’s transport equation where we neglected the current density term (i.e. for small excess carrier concentrations),

$$
\begin{align*}
\frac{\partial \Delta n}{\partial t} &\approx -\frac{\Delta n}{\tau_n} \\
\frac{\partial \Delta p}{\partial t} &\approx -\frac{\Delta p}{\tau_p}
\end{align*}
$$

(1.38)

yielding

$$
\begin{align*}
n(t) &= \Delta n e^{-t/\tau_n} + n_0 \\
p(t) &= \Delta p e^{-t/\tau_p} + p_0
\end{align*}
$$

(1.39)

From the Equations (1.39) the role of $\tau$ as a relaxation time is evident, since it represents – as a first approximation – the time needed to restore the equilibrium concentrations in the case of recombination, or to rise at the perturbed ones in generation processes. Under
the same assumption given before, each lifetime contribution due to its respective process (radiative, SRH, Auger and others) can be treated as independent, leading to a resultant lifetime of the following form

$$\frac{1}{\tau} = \sum_j \frac{1}{\tau_j}$$

(1.40)

where the index represents the $j$-th process contribution.

Minority carrier lifetimes in HgCdTe lie in a huge range, from $10^{-10}$ to $10^{-5}$ s, with increasing values for lower doping and dislocations concentrations ($< 5 \cdot 10^5$ cm$^{-2}$). For instance, $n$-type samples ($x \simeq 0.2$, 80 K) with carrier concentrations less than $10^{15}$ cm$^{-3}$ exhibit a range of values (0.4-8 µs) depending on the growth technique [1], while the highest lifetime was measured in high-quality undoped samples grown by low temperature LPE and MOCVD. Regarding the temperature, it is well known that the Auger mechanism governs the high-$T$ lifetime, while the Shockley-Read-Hall is mainly responsible for low-ones. This is particularly true in $n$-type HgCdTe where this two-fold trend is pretty evident, but less in $p$-type. In the latter case we have a very broad range of lifetime values at low-temperature, then an increase for intermediate temperatures, where there is a maximum, and again a decrease for high-temperatures. Another fact to be accounted for is the nature of the dopants: it has been observed a $\tau$ increased by a factor 2-5 for Au and As-doped HgCdTe with respect to the native doping of the same level. This probably arises from a reduction of SRH impurity levels.

Once the general behavior of lifetimes in HgCdTe has been discussed, in the following sections we will derive in detail the expressions for the net rate $U$ (usually named also net recombination rate $R$) and its corresponding $\tau$ for each mentioned GR process, when we will reconsider all the mechanisms affecting the photodetector performance.

C – Photon Absorption Properties

We already discussed about the absorption coefficient $\alpha$ and in particular we have seen with the Formula (1.29) its dependence on the parameter $x$. Despite the almost parabolic band structure of HgCdTe should determine a square-root dependence of $\alpha(E)$ it is worth stressing that in literature there are experimental results somehow in disagreement with this theoretical behavior except for high-quality samples. The presence of unexpected tails in the absorption curves at energies lower than $E_g$ (see Figure 1.11) is partially explainable with the presence of lattice impurities or defects, conduction or valence intraband transitions, valence interband transitions and also with the already mentioned compositional fluctuations, that determine regions in the lattice of $E_g$ different from the average.

The first description is due to Urbach (1953), that stated the exponential behavior of $\alpha$ with respect to the frequency in AgBr. In 1957 Martienssen extended the validity of that rule to KBr, by introducing the formula

$$\alpha = \alpha_0 e^{\frac{\sigma (E - E_0)}{k_B T}}$$

(1.41)

where $\sigma$, $\alpha_0$ and $E_0$ are fitting parameters. In 1966 Marple found that the (1.41) is also applicable to II-VI compounds, like Hg$_{1-x}$Cd$_x$Te, if we put

$$\begin{cases}
E_0 = -0.355 + 1.77 x & \text{in eV} \\
\ln \alpha_0 = -18.5 + 45.68 x \\
\sigma = 3.267 \cdot 10^4 (1 + x) & \text{in K/eV}.
\end{cases}$$

(1.42)
1.2 Review on some Physical Aspects in HgCdTe

\[ \log(\alpha) \times \]

Figure 1.11. Behavior of the absorption edge in Hg\textsubscript{0.557}Cd\textsubscript{0.443}Te with respect to the photon energy, plotted at different temperatures (adapted from [3]). It is well appreciable the two-fold trend of the curves: Urbach-like below the knee (at \( E_g \)) and Kane-like above it. Moreover it is evident that all the curves converge to the same point, called “focal point”, that is \( x \)-sensitive and in this case is located at \((0.441; 10)\). If we plot the curve \( \log(\alpha)(E) \) at a fixed temperature and for different molar fractions we should note an horizontal shift (as indicated by the arrow) towards greater energy values.

When the photon has \( E = E_g \) we have a particular \( \alpha = \alpha_g \), yielding:

\[
\frac{\alpha}{k_B T} = \ln \frac{\alpha_g}{\alpha_0} - \frac{E_g - E_0}{E_g - E_0} \tag{1.43}
\]

It should be noticed that \( E_g \) is often derived from the experimental corresponding of (1.43) by imposing \( \alpha_g = 500 \) or 800 cm\(^{-1}\). From this equation and fitting experimental data one can obtain also

\[
\begin{align*}
\alpha_g(x) &= 500 + 5.6x & 0.17 \leq x \leq 0.443, \ T = 300 \ \text{K} \\
\alpha_g(x) &= -65 + 1.88T + (8694 - 10.31T)x & \text{otherwise}
\end{align*} \tag{1.44}
\]

This Urbach-like model is able to explain the behavior of \( \alpha(E) \) for \( E \lesssim E_g \). When the band-to-band transitions overcome this first zone it is better to use the Kane rule:

\[
\alpha = \beta \sqrt{E - E_g} \tag{1.45}
\]

where

\[
\beta = 2.109 \cdot 10^5 \sqrt{\frac{1 + x}{81.9 + T}} \tag{1.46}
\]
and $E = h\nu$ is the photon energy

The absorption strength of the material generally decreases as the gap becomes smaller due to the decrease of $m^*_C$. Moreover, the degree of energy band nonparabolicity increases as the temperature or $x$ decreases, resulting in a discrepancy between experimental data and theoretical $\alpha(E)$. Then, recently, a new model has been suggested [1]:

$$\alpha \sim \sqrt{(E - E_g + c)^2 - c^2 (E - E_g + c)}$$

where $c$ is a parameter related to the band hyperbolic curvature.

The refractive index in HgCdTe is usually described by the Kramers-Kronig formalism as a function of the temperature, but for $0.276 \leq x \leq 0.540$ and in the range $4.2 \leq T \leq 300$ K (or $0.205 \leq x \leq 1$ at room temperature) the following empirical expression is widely exploited [1]:

$$n^2(\lambda, T) = A + \frac{B}{1 - (\frac{C}{\lambda})^2} + D\lambda^2$$

where $A$, $B$, $C$ and $D$ are fitting parameters varying with $x$ and $T$.

The high frequencies dielectric constant $\epsilon_\infty$ and the static one $\epsilon_0$, derived typically by reflective experiments, are instead only dependent on $x$ and scarcely on the temperature [1], so:

$$\epsilon_\infty = 15.2 - 15.6x + 8.2x^2$$

$$\epsilon_0 = 20.5 - 15.6x + 5.7x^2.$$  

---

**Figure 1.12.** Trend of the variation $\Delta \lambda_c$ and molar fraction $x$ with respect to the cut-off wavelength for a fluctuation $\Delta x = 0.001$ in Hg$_{1-x}$Cd$_x$Te growth process and at $T = 77$ K (from [1]). From the plot it is evident that a good growth process is a stringent requirement for the LWIR detection (right side of the figure).

Finally, besides the considerations already done about the cut-off wavelength with the equations (1.30) and (1.31), we can substitute the latter into the first one and, rearranging it, we obtain:

$$\lambda_c(T, x) = \frac{1}{-0.244 + 1.566x + (4.31 \cdot 10^{-4}) T (1 - 2x) - 0.65 x^2 + 0.671 x^3}$$
\[ \frac{\Delta \lambda_c(T, x)}{\Delta x} = \lambda_c^2 (1.566 - 8.62 \cdot 10^{-4} T - 1.3 x + 2.013 x^2) \] (1.52)

If we plot this relationship between the variation in the cut-off wavelength \( \Delta \lambda_c \) and \( \lambda_c \) itself at a fixed uncertainty \( \Delta x = 0.001 \) (see Figure 1.12) we recover some considerations discussed at the beginning: in fact, if the technological process ensures a high-quality HgCdTe we can manage the absorption band spectrum of our IR photodetector.

### D – Doping, Impurities and Defects

We have already discussed the role of non-constant lattice composition by the optic and electronic point of view. Let’s browse now a brief description on what kind of variational features there could be in HgCdTe-based devices.

Typical acceptors for Cd or Hg sites (substitutive dopants) are elements from group IA or IB (Ag, Cu, Au), while group IIIB (Ga, Al, In) has shown a donor predisposition for the same sites. For the Te, instead, some elements from group VIB (or acceptors in interstitial sites) and VIIIB (Br, Cl, I) are substitutive donors, whereas elements from the VB (Te-substitutional) are acceptors. Concerning the whole material the most effective \( p \)-dopant is Arsenic while Indium is the best \( n \)-dopant.

Some elements (B, Al, Ga, In, Si, O, Cl, Br, I) behave either as impurity or as donor and the same occurs for some acceptors (Li, Cu, Ag, Au, As, Sb, P). The role of point defects (like vacancies or interstitials) and lattice defects (dislocations and grain boundaries) is two-folded, since sometimes they behave as acceptors, sometimes as donors. Moreover they can interact with other elements such as dopants or defects themselves.

In general, reducing defects is one of the most difficult challenges in HgCdTe growth, being sources of carrier lifetime and mobility decrease, current leakage and 1/f-noise. They mainly can act as SRH recombination centers determining an increased dark current in photodetector for concentrations \( > 5 \cdot 10^5 \text{ cm}^{-2} \) and also dead pixels. Unfortunately at this dislocation level the usual dimension of such devices yields a threshold of about 12 dislocations per photodiode [14]. This impose a particular attention during the growth process in order to avoid damages, contaminations and other lattice imperfections like mismatches, stacking faults, precipitates, grain boundaries, pinholes, striations, terracing, interstitials and inclusions.

Vacancies are considered a kind of dominant native defect in undoped HgCdTe, especially with respect to the metal sites, acting as acceptors (native donor defects are negligible) and high-diffusivity centers. In this framework the partial pressure of Hg during the growth process is of crucial importance, being the final hole concentration proportional to \( 1/p_{\text{Hg}} \) [19]. A subsequent annealing is required to eliminate unwanted impurities but the choice of the temperature is again very important, since the appearance of additional problems like \( p \)-to-\( n \) conversion, surface layers formation or strain effects depend on it.

### 1.3 Fundamental Concepts concerning IR Photodetectors

In the past sections we have already cited some mechanisms involving semiconductors in thermal or optical energy interactions: firstly we saw the general distinction in extrinsic, intrinsic or free carrier absorption, where we underlined the importance of the second one as a very common energy transfer process in direct gap semiconductors; then the most important generation/recombination processes in photodetectors have been briefly discussed. The aim of this paragraph is to set the principal functioning (and limiting)
1.3 Fundamental Concepts concerning IR Photodetectors

physical mechanisms at the origin of a photodetector description and, consequently, of devices modeling.

1.3.1 Theory of Photon Detection

The simpler macroscopic description of a photodetection system should consist in a reverse biased \( pn \) diode (the positive voltage is at the \( n \)-contact). Let’s firstly consider the structure without light and without bias. Due to the built-in potential, at the junction between the two doped regions the excess of holes in valence band of the \( p \) side tends to compensate the excess of free electrons in the conduction band of the \( n \) side generating a two-fold free carriers diffusion current: electrons diffuse into \( p \)-type region and holes into \( n \)-type region up to an equilibrium condition. The resultant of this diffusion is the formation at the junction of a free carrier depletion with remaining ionized atoms. This ionization, in turn, generates an electrical potential (and field \( \mathcal{E} \)) acting against the further diffusion through a drift current mechanism according to which any free carrier in the depleted region is drifted into its appropriate side (electrons are swept in \( n \)- and holes in \( p \)-type region). The reverse bias acts as a further driving force in the drift mechanism, enhancing it with respect to the diffusion. This leads to an increasing of the potential, through a so-called space charge field, and also of the width of the depleted region. If we introduce now the presence of light, entering for example from the \( p \) side, intrinsic or extrinsic absorption processes can occur by the production of electron-hole pairs. Electrons generated in the \( p \)-type region diffuse in this region and here recombine with holes. Pairs generated in the \( n \)-type region diffuse in \( n \)-region and there recombine with electrons. Pairs generated in the depleted region (and outside it, but at a distance from the depletion region edge less than the diffusion length), instead, are separated: thanks to the electric field these new carriers are drifted into their respective sides, resulting in a small detectable reverse current proportional to the radiation intensity. Since this current unbalance depends also on the semiconductor absorption efficiency we must choose a material in which these mechanisms are enhanced: for the infrared detection one of the best materials is without doubts the HgCdTe, as already demonstrated in the present work.

Now we want to answer the question: what happens in the radiation-matter interaction at the microscopic level? In the simplest case a photon can give its energy \(( \geq E_g \) \) to a valence electron promoting it in the conduction band (see the intrinsic absorption of Figure 1.1). This represents a direct band-to-band generation and its opposite is the radiative direct band-to-band recombination previously seen: a very important process mainly in emitting devices, but also in detectors composed by an array of photodiodes, where the emission produced by one diode can affect the performance of another one.

So, if this kind of generation is important because it contributes to the photocurrent signal production, the corresponding recombination is a limiting factor in photodetector and for this reason it needs to be analyzed under the recombination lifetime standpoint. Exploiting the theory developed by van Roosbroeck and Shockley the recombination rate can be written as

\[
R_{\text{rad}} = \frac{8\pi}{\hbar^3 c^3} \int_0^\infty \frac{\epsilon(E) \alpha(E) E^2}{e^{\frac{E}{kT}} - 1} \, dE
\]

where often the high-frequency dielectric constant \( \epsilon_\infty \) is used in place of \( \epsilon(E) \). For small \( \Delta n \) and \( \Delta p \) we can approximate the lifetimes with

\[
\tau_{n,\text{rad}} = \tau_{p,\text{rad}} = \tau_{\text{rad}} = \frac{n_0^2}{R_{\text{rad}}(n_0 + p_0)}.
\]
1.3 Fundamental Concepts concerning IR Photodetectors

Figure 1.13. Summary of the most important competitive radiative mechanisms in semiconductors:
a) **Direct band-to-band** generation (intrinsic absorption) due to the photon energy imparted to a valence electron promoted into the conduction band; this results in a vacancy in VB and a new occupied state in CB determining the appearance of an electron-hole pair, the exciton. b) **Radiative direct band-to-band** recombination, in which the drop of an electron from CB to VB generates the emission of a photon with energy equal to the energy difference between the initial and the final state and, thus, the electron-hole pair annihilation.

that holds for undoped materials while for *p*-type semiconductors (*p₀ ≫ n₀*) \( \tau_p = n_i^2 / R_{rad}p₀ \) and for *n*-type semiconductors (*n₀ ≫ p₀*) \( \tau_n = n_i^2 / R_{rad}n₀ \). The (1.53) can be computed by using the Bardeen expression (1956) for \( \alpha(E) \)

\[
\alpha(E) = \frac{3\sqrt{2\pi} m q^2}{3\sqrt{\epsilon\infty} h^2} \sqrt{\left( m_e^* m_h^* \right)^3} \left( 1 + \frac{m_0}{m_e^*} + \frac{m_0}{m_h^*} \right) \sqrt{\frac{E - E_g}{m e^2}} \tag{1.55}
\]

and integrating it, that yields:

\[
R_{rad} = n^2 \cdot 5.8 \cdot 10^{-13} \sqrt{\frac{\epsilon\infty \left( \frac{m_0}{m_e^* + m_h^*} \right)^3 \left( 1 + \frac{m_0}{m_e^*} \right)}{\epsilon\infty \left( \frac{m_0}{m_e^* + m_h^*} \right)^3 \left( 1 + \frac{m_0}{m_e^*} \right)}} \times \left( \frac{300 K}{T} \right)^3 \left( E_g^2 + 3k_BTE_g + 3.75 k_B^2 T^2 \right). \tag{1.56}
\]

Note that again the previous equation is, obviously, x-sensitive, being \( E_g \), \( n_i \) and \( \epsilon\infty \) functions of the molar fraction. Even if the above relation holds especially for parabolic bands and non-degenerate semiconductors, it has been experimentally demonstrated that they represent, almost always, a good model for Hg\(_{1-x}\)Cd\(_x\)Te. Finally, it is worth stressing that in the direct band-to-band process the couples generation can also come from thermal energy, as well as optical. In this case, and with very narrow bandgap, the high-temperature (especially when \( E_g \) and \( k_B T \) are of the same order of magnitude) is a crucial parameter since it can induce unwanted generation, resulting in noise or dark current increment.

When the semiconductor is doped, or there are some impurities resulting in the presence of a concentration \( N_t \) of SRH centers in the bandgap, external excitations in general can interact with one of these traps in at least four ways (see also Figure 1.14):

a) an empty SRH trap (with and energy \( E_i \)) captures a conduction electron with a cross-section \( \sigma_n \) and a capture rate \( c_n = \sigma_n v_n \) (where \( v_n = \sqrt{\frac{8k_B T}{\pi m_e^*}} \) is the thermal speed of the electron);
b) an occupied trap emits electrons onto the CB with emission rate $e_n$;

c) an occupied trap captures a valence hole with a cross-section $\sigma_p$ and a capture rate $c_p = \sigma_p v_p$ (where $v_p = \sqrt{8k_B T/\pi m_p}$ is the thermal speed of the hole);

d) an empty trap emits holes onto the VB with emission rate $e_p$.

![Figure 1.14. Schematic representation of the four possible generation/recombination processes involving a single SRH center within the bandgap. All these mechanisms can convert thermal or optical energy into the exciton generation, or vice-versa for the annihilation: for instance process b) is the standard extrinsic absorption when the electron transition from the trap to the conduction band is induced by the energy conversion of an incident photon.](image)

Using a Fermi-Dirac statistics for small excess of carriers $\Delta n$ and $\Delta p$ we find that the evolutions in time of $n$ and $p$ (i.e. the net SRH recombination rates) are:

$$
\begin{align*}
\frac{dn(t)}{dt} &= -c_n n(t) N_t (1 - f_t(t)) + e_n N_t f_t(t) \\
\frac{dp(t)}{dt} &= -c_p p(t) N_t f_t(t) + e_p N_t (1 - f_t(t))
\end{align*}
$$

(1.57)

(in which the terms $N_t f_t = N_t^-$ refer to the fraction of occupied trap concentrations and $N_t (1 - f_t) = N_t^0$ of the unoccupied ones) that, with the charge neutrality law

$$
n(t) + N_t f_t(t) = p(t),
$$

(1.58)

represent a nonlinear set of first-order differential equations. Under steady-state equilibrium ($\frac{dn}{dt} = 0$ and $\frac{dp}{dt} = 0$, with $n = n_0$ and $p = p_0$):

$$
\begin{align*}
\frac{e_n}{c_n} &= n_0 e^{\frac{E_l - E_C}{k_BT}} \equiv n_1 \\
\frac{e_p}{c_p} &= p_0 e^{\frac{E_V - E_l}{k_BT}} \equiv p_1
\end{align*}
$$

(1.59)

where $n_1$ and $p_1$ are the carrier densities for $E_l \equiv E_F$, and the (minimum) lifetime are

$$
\begin{align*}
\tau_{n_0} &= \frac{1}{c_n N_t} = \frac{1}{\sigma_n v_n N_t} \\
\tau_{p_0} &= \frac{1}{c_p N_t} = \frac{1}{\sigma_p v_p N_t}
\end{align*}
$$

(1.60)
Since the net rate of recombination can be rewritten as

\[
\begin{align*}
\frac{dn(t)}{dt} &= -c_n (n N_i^0 - n_1 N_i^-) \\
\frac{dp(t)}{dt} &= -c_p (p N_i^- - p_1 N_i^0)
\end{align*}
\] (1.61)

at the steady-state condition we have

\[
\frac{N_i^-}{dt} = \frac{dn}{dt} + \frac{dp}{dt} = c_n (n N_i^0 - n_1 N_i^-) - c_p (p N_i^- - p_1 N_i^0) = 0
\] (1.62)

yielding:

\[
\begin{align*}
N_i^0 &= \frac{c_n n_1 + c_p p}{c_n (n + n_1) + c_p (p + p_1)} \\
N_i^- &= \frac{c_n n + c_p p_1}{c_n (n + n_1) + c_p (p + p_1)}
\end{align*}
\] (1.63)

When we account for an external excitation the densities of electrons and holes are perturbed, then they begin to decrease towards their equilibrium values according to

\[
\frac{dn'}{dt} = \frac{dp'}{dt} = \frac{(n_0 + \Delta n) (p_0 + \Delta p) - n_i^2}{(n_0 + \Delta n + n_1) + \tau_{\text{rad}} (p_0 + \Delta p + p_1)} = U_{\text{SRH}}
\] (1.64)

where \( n \rightarrow n' = n_0 + \Delta n \) and \( p \rightarrow p' = p_0 + \Delta p \) are the excited carrier densities. So, for the electrons we find

\[
\tau_n = -\left( \frac{dn'}{dt} \right)^{-1} \Delta n
\] (1.65)

\[
= -\frac{\tau_{\text{rad}} (n_0 + \Delta n + n_1) + \tau_{\text{rad}} (p_0 + \Delta p + p_1)}{(n_0 + \Delta n) (p_0 + \Delta p) - n_i^2} \Delta n
\]

then

\[
\tau_{\text{SRH}} = -\frac{\sigma_n v_n (n_0 + \Delta n + n_1) + \sigma_p v_p (p_0 + \Delta p + p_1)}{\sigma_n v_n \sigma_p v_p N_i^2 [(n_0 + \Delta n) (p_0 + \Delta p) - n_i^2] \Delta n}
\] (1.66)

and similarly for the holes.

For \( x < 0.28 \) and at room temperature the SRH (recombination) lifetime is slightly greater than \( \tau_{\text{rad}} \) in HgCdTe, with an order of magnitude of \( 10^{-6} \) s. But since it is a trap-assisted process, at 77 K the radiative lifetime becomes the higher term with \( 10^{-5} \) s. Nonetheless, there is another relevant mechanism affecting the kinetics of photogenerated signal: the Auger process. This three-particle fundamental mechanism dominates generation and recombination especially at near room temperature, determining the smallest lifetime for almost all \( x \) values.

The Auger generation is essentially a kind of impact ionization, that is when an external stimulus gives to a carrier the required energy to create an electron hole pair through collision (Coulomb interaction), while the Auger recombination is, obviously, its inverse process. In 1962 Beattie reported ten possible photonless band-to-band Auger mechanisms for HgCdTe, involving an electron dropping from conduction to valence band that recombines with a hole in the valence band. According to their conservation principle, energy and momentum variation of this first electron is taken up by a second electron or a hole, and the way in which this recombination occurs determines the particular kind of such process (see Figure 1.15).
1.3 Fundamental Concepts concerning IR Photodetectors

Besides the mechanisms reported in Figure 1.15 there are also radiative or trap-assisted Auger process but in the case of narrow gap semiconductors the most relevant phenomena (see Figure 1.16) are the “Auger 1”, especially in \( n \)-type HgCdTe (particularly for \( x \sim 0.2 \) and at high temperatures), and the “Auger 7”, very common in \( p \)-type HgCdTe (and quite less the “Auger 3”) if the SO is ignored but occurs also in \( n \)-type. The Auger 1 generation is the impact ionization by an electron, generating a second electron and a heavy hole, while in the Auger 1 recombination (also labeled CHCC) an electron in conduction band, recombining with a hole in HH valence sub-band, transfers the energy lost to another electron in CB which is temporarily promoted into a higher excited state. Afterwards this excited electron returns to its initial state, often by emitting a phonon in the lattice. The Auger 7 generation is the impact generation of an electron-hole pair, involving one heavy hole, one light hole and one electron, while its corresponding Auger 7 recombination mechanism (also labeled CHLH) involves the recombination between a conduction electron and a heavy hole, while the excess energy is absorbed by an electron in the LH band which is promoted to an empty state into the HH band. This fact explains the prevalence of A7 processes in \( p \)-type semiconductors, which are highly populated of heavy-holes in VB and much less of electrons.

Before analyzing the formalism required to obtain suitable expressions for the Auger net recombination rate and for its lifetime we expect, being a heat-induced mechanism, a strong dependence on temperature and thus on \( E_g \); in fact, high temperatures and narrow bandgap should increase the Auger process, as well as the other thermal transitions.

The complete treatment on this kind of generation/recombination mechanisms has been developed in 1959 by Beattie and Landsberg [20]. In the case of Auger 1, they start from the rate expressions in the parabolic band approximation

\[
\begin{align*}
G_{A1} &= \frac{n'}{n_0} g_{A1} \\
R_{A1} &= \frac{n'^2 p'}{n_0^2 p_0} r_{A1}
\end{align*}
\]

(1.67)

where the coefficients \( g_{A1} \) and \( r_{A1} \) (i.e. the rates without external excitations) must be
determined. From the steady state condition, with $\Delta n = \Delta p$, one derives

$$
\tau_{A1} = \frac{n_i^4}{n' p_0 (n' + p_0)} g_{A1}.
$$

(1.68)

Through the evaluation of the transition probability from the initial to the final state and by properly integrating it over the wavevector $k$, they computed the $g_{A1}$ assuming negligible all screening effects and accounting for non-degenerate parabolic bands as:

$$
g_{A1} = \frac{8 (2\pi)^{\frac{7}{2}} q^4 m_0 \frac{m_e}{m_h} |F_1F_2|^2}{\hbar^3 \epsilon^2 \sqrt{1 + \mu (1 + 2\mu)}} \frac{n_0}{n_0} \left( \frac{k_B T}{E_g} \right)^\frac{3}{2} \exp \left[ \frac{-1 + 2\mu}{1 + \mu} \frac{E_g}{k_B T} \right]
$$

(1.69)

where

$$
\epsilon = \frac{m^*_e}{m^*_HH},
$$

(1.70)

$\epsilon$ is the relative dielectric constant and the term $|F_1F_2|$, here assumed constant, is the product of the overlap integrals between the periodic part of wavefunctions involving the bottom of the conduction band with the valence band and with the above mentioned excited state (again in CB, as represented in Figure 1.16). Exploiting the (1.69) and (1.68) we can compute the Auger 1 (recombination) lifetime under the previous assumptions.

Figures 1.16. Schematic description of the two most relevant direct band-to-band Auger recombination mechanisms occurring in HgCdTe. In the CHCC are involved two electrons and one heavy hole, while the CHLH involves the recombination between a conduction electron and a heavy hole, while the excess energy is absorbed by an electron in the LH band which is promoted to an empty state into the HH band.

Now, since for intrinsic HgCdTe samples ($\mu \ll 1$)

$$
\tau_{A1,i} = \frac{n_i}{2 g_{A1}}
$$

(1.71)

we can approximate the previous lifetime as

$$
\tau_{A1} = \frac{2n_i^2}{(n_0 + p_0)n_0} \tau_{A1,i}
$$

(1.72)

or, better,

$$
\left\{
\begin{array}{ll}
\tau_{A1} \simeq 2 \frac{n_i^2}{n_0^2} \tau_{A1,i} \propto \exp \left[ \frac{\mu E_g}{(1+\mu) k_B T} \right] / n_0^2 & \text{for } n\text{-type} \\
\tau_{A1} \simeq 2 \frac{n_i^2}{\beta p_0^2} \tau_{A1,i} \propto \exp \left[ \frac{\mu E_g}{(1+\mu) k_B T} \right] / p_0 & \text{for } p\text{-type}
\end{array}
\right.
$$

(1.73)
where $\beta$ is a function of $\mu$, $T$ and $E_g$ and depends on other Auger processes (especially Auger 3). Since typically $\beta \ll 1$ the lifetime in $p$-HgCdTe is longer than in $n$-type. Removing, furthermore, the approximation of $|F_1 F_2|$ as a constant by considering its k-dependence (in literature $0.1 \leq |F_1 F_2| \leq 0.3$) and accounting also for the effect of non-parabolic bands we obtain a more familiar and simple expression for the lifetime in $n$-HgCdTe:

$$\tau_{A1} = \frac{1}{g_{A1} (n_0 + p_0 + \Delta n)(n_0 + \Delta n)}$$

(1.74)

For the Auger 7, again with these same assumptions, Beattie and Smith [21] stated

$$\tau_{A7} = \left[ 1 + \frac{2n_0 \Delta p_{HH}}{p_{0,HH} \Delta n} - \frac{\Delta p_{LH} n_0}{p_{0,LH} \Delta n} \right]^{-1} \frac{D(E_{th})}{I(\eta_{th}) \eta_{th}^2}$$

(1.75)

where

$$E_{th} \simeq \frac{2 + \mu}{2 + \mu - \frac{m_{HH}^*}{m_{HH}^*}}$$

(1.76)

is the threshold energy for this process, $D(E_{th})$ is a function of the effective masses, $\eta_{th} = E_{th}/k_B T$ and $I(\eta_{th})$ represents an integral over the conduction band energy. The equation (1.75) can be simplified to

$$\tau_{A7} = \frac{2n_i^2}{(n_0 + p_0)p_0} \tau_{A7,i}$$

(1.77)

where

$$\tau_{A7,i} = \frac{D(E_{th})}{2 \eta_{th}^2 I(\eta_{th})}$$

(1.78)

since the term in square brackets in (1.75) becomes 1/2 in the intrinsic case. So, comparing $\tau_{A1}$ and $\tau_{A7}$ one can obtain

$$\frac{\tau_{A7}}{\tau_{A1}} = \frac{n_0}{p_0} \frac{\tau_{A7,i}}{\tau_{A1,i}} = \frac{n_i^2}{p_i^2} \tau_{A7,i}$$

(1.79)

where

$$\gamma = 2 \frac{m_C^*(E_{th})}{m_0^*} 1 - \frac{\frac{5}{2} \eta_{th}}{1 - \frac{5}{2} \eta_{th}}$$

(1.80)

($m_C^*(E_{th})$ and $m_0^*$ are, respectively, the effective conduction mass at $E_{th}$ and at the bottom of the band, with $E_{th,1} \simeq E_{th,7} \simeq E_g$) that, according to the literature, is a value spanning the range from 3 to 8 for molar fractions $0.16 \leq x \leq 0.40$ and temperatures $50 \leq T \leq 300$ K.

Finally, the net recombination rate for Auger processes 1 and 7 is expressed by

$$R_A = \frac{np - n_i^2}{2n_i^2} \left( \frac{n}{(1 + an) \tau_{A1,i}} + \frac{p}{\tau_{A7,i}} \right)$$

(1.81)

with $a \sim 10^{-18}$ cm$^3$ and where the (1.81) can be rewritten in the more generic form

$$R_A = (C_n n + C_p p) (np - n_i^2)$$

(1.82)

introducing the Auger recombination coefficients $C_n$ and $C_p$, which depend on the material.
Once we have obtained the most common (approximated) expressions for the three principal GR mechanisms in narrow gap semiconductors as HgCdTe – radiative, SRH and Auger processes – the net lifetime can be written accounting for the formula (1.40), so:

\[
\frac{1}{\tau} = \frac{1}{\tau_{\text{rad}}} + \frac{1}{\tau_{\text{SRH}}} + \frac{1}{\tau_{\text{A}}} = \frac{1}{\tau_{\text{rad}}} + \frac{1}{\tau_{\text{SRH}}} + \frac{1}{\tau_{\text{A1}}} + \frac{1}{\tau_{\text{A7}}}.
\]

It is worth noting that, besides these fundamental mechanisms occurring in semiconductors there are many other physical effects: first of all we didn’t treat the case of radiative Auger or trap-assisted Auger but, as already seen referring to SRH mechanism, the presence of impurities can induce the enhancement of GR processes, even at low temperatures. Moreover, by a general point of view, the Auger mechanism suffers heavily the thermal energy since, as one can note remembering the \(T\)-dependence of carrier densities in the equations (1.81), it is particularly furthered at relatively high temperatures and also at room temperature. From this fact it arises the requirement of cooled devices for Auger suppression. Another source of GR signal, affecting the photogenerated current, could be the presence of surface states that can generate or recombine electron-hole pairs. To prevent this kind of phenomena the growth process of HgCdTe, in our case, must be accurate as much as possible in order to produce high-quality materials, with well-controlled molar fraction and without surface defects.

If these effects are almost pretty controllable, by the engineering standpoint, there is an important fundamental mechanism that can be kept under control with much more difficulty: the tunneling. Due to its quantum nature, it is an inescapable phenomenon of matter that allows transitions of electrons among the valence and conduction band thanks to the probabilistic essence of fundamental particle wavefunction with respect to a barrier potential as the energy gap is in the semiconductor framework. In the next chapter we will analyze the physics of tunneling and its modeling theory.

As explained before, the most simple semiconductor device for photon detection is the \(pn\) junction under reverse bias, in which the light power is converted into electrical power through the photovoltaic effect. When the light illuminates the junction region the photons enter into the semiconductor by a penetration depth \(1/\alpha\) (the particular distance \(z\) at which the optical power drops to \(1/e\) with respect to the illuminated facet) of several microns according to the absorption Lambert’s law

\[
P^a(z) = [P^i(1 - r)](1 - e^{-\alpha z})
\]

where \(P^a = n^a_{ph} h\nu\) and \(P^i = n^i_{ph} h\nu\) represent, respectively, the absorbed and the incident power, \(n^a_{ph}\) and \(n^i_{ph}\) are the number of photons absorbed and incident per second, \(r\) is the reflectivity and \(\alpha\) is the well-known absorption coefficient. Since \(\alpha(E_g)\) (see equation 1.45) is a physical quantity material-related, by properly choosing the semiconductor we are able to absorb, ideally, only the radiation such that \(h\nu \geq E_g\). This is true for intrinsic semiconductors (for the extrinsic absorption and in the case of impurities or defects the energy of absorbed photons can be also less than the energy gap) and except for the absorption band tails previously seen (see Figure 1.11). The generation/recombination processes, as already explained, affect the presence of electron-hole pairs and thus they perturb the drop voltage. Connecting the two electrodes, a current in the circuit will establish in order to restore the equilibrium minority carrier concentrations. Not only: if the light stimulus modulates, also the photocurrent changes accordingly to the incident power variation and to the circuit parameters (bias voltage \(V_b\) and load resistance \(R_L\)).
The way through which the semiconductor converts the optical power into the photocurrent $I_{ph}$ is well represented by one of the most important parameters of photovoltaic detectors: the (internal) quantum efficiency $\eta$, i.e. the ratio between the number of electron-hole contributing to photocurrent and the number of incident photons (the external quantum efficiency is, instead, the number of photo-induced carriers divided by the number of absorbed photons per second $n_{ph}$), or

$$\eta = \frac{I_{ph}}{q P_n} = \frac{I_{ph}}{q n_{ph}}, \quad (1.85)$$

where, if each incident photon is converted into a photocarrier we have $\eta = 1$, otherwise $0 < \eta < 1$. As well as the (1.84)

$$\eta(z) = (1 - r) \left( 1 - e^{-\alpha z} \right) \quad (1.86)$$

suggesting that the most effective configuration of the device structure should consist of an illuminated surface with low reflectivity $r$ and a shallow junction with respect to this surface, in order to absorb most of the optical power within the depletion region. In this way we have diffusion lengths larger than the junction depth, allowing carriers to reach the depletion region through diffusion processes. Nonetheless the formula (1.86) does not account for some noise phenomena like, for instance, surface recombination processes.

In Figure 1.17 we see the schematized structure of a $pn$ junction photodiode with its bias circuit and the corresponding $I-V$ characteristic, which shows the possible operating modes of the device. Increasing the voltage in reverse bias and dark conditions (the dashed line) the curve firstly exhibits a plateau (region II) at a current value called saturation current $I_0$ then, for higher negative voltages, greater than the breakdown point $V_{BD}$, the tunneling processes determine an abrupt increase of the reverse current (region I).

When the device is illuminated the shape of the $I-V$ remains the same but the net result of intrinsic and extrinsic absorption processes leads to a photocurrent component $I_{ph}$ that make the inverse current increasing, proportionally on the power $P_i$ (and, hence, $P_a$). The region III of the plot (at low or absent voltage) is often the operating mode of solar cells while region I is common for avalanche diodes, so in our work we will focus only on the region II, suitable for generic photodiodes.

Theoretically, the ideal dark current is of the form

$$I_{dark} = I_0 \left( \frac{q V_b}{e^{kT} - 1} \right) \quad (1.87)$$

where the saturation term $I_0$ represents at first approximation (neglecting the drift component in quasi-neutral regions) a diffusion current of electrons into the $p$-type region and holes into the $n$-type region [1]:

$$I_0 = \left( \frac{q D_n p_n}{L_h} \right) \frac{\left( 1 \right)}{L_h} \cosh \left[ \frac{x_n}{L_h} \right] + \sinh \left[ \frac{x_n}{L_h} \right] + \frac{\left( 1 \right)}{L_h} \sinh \left[ \frac{x_n}{L_h} \right] + \cosh \left[ \frac{x_n}{L_h} \right] + \frac{\left( 1 \right)}{L_h} \cosh \left[ \frac{x_n}{L_h} \right] + \sinh \left[ \frac{x_n}{L_h} \right] + \cosh \left[ \frac{x_n}{L_h} \right] \quad (1.88)$$

in which

$$w_{dep} (V_b) = \sqrt{ \frac{2 e \left( \frac{k_B T}{q} \ln \left[ \frac{N_a N_d}{n_i^2} \right] \pm V_b \right)}{q N_a N_d (N_a + N_d)} }, \quad (1.89)$$
\[ N_a \text{ and } N_d \text{ are respectively the acceptor and donor concentrations, } \epsilon \text{ is the dielectric constant, } p_n \text{ is the hole carrier concentration in } n\text{-side and } n_p \text{ the vice-versa, } D_e = \mu_e k_B T / q \text{ and } D_h = \mu_h k_B T / q \text{ are the minority carrier diffusion coefficients according to the Einstein relations, } L_e = \sqrt{D_e \tau_e} \text{ and } L_h = \sqrt{D_h \tau_h} \text{ the minority carrier diffusion length and } v_1 \text{ and } v_2 \text{ the recombination velocities of minority carriers respectively at the back (holes in } n) \text{ and illuminated surface (electrons in } p). \]

For the geometrical parameters \( x_n, t \) and \( d \) refer to the Figure 1.18. For a junction with thick quasi-neutral regions (\( E \simeq 0 \)), i.e. when \( x_n \gg L_n \) and \( t + d - x_n - w_{dep} \gg L_e \), we find

\[
I_0 \simeq \frac{q D_h p_n}{L_h} + \frac{q D_e n_p}{L_e}
\]

This kind of current is usually dominant at high temperatures since

\[
I_0(T) \sim n_i^2(T) \sim e^{-\frac{E_g}{k_B T}}.
\]

Therefore, the generic expression for the total current in light conditions consists of two terms: one is the \( I_{dark} \) and the second is represented by the photocurrent \( I_{ph} \), where its basic expression can be obtained from (1.85):

\[
I_{ph} = \eta q n_i^4.
\]
1.3 Fundamental Concepts concerning IR Photodetectors

so

\[ I_{\text{tot}}(V_b, T, n_{\text{ph}}^i) = I_{\text{dark}} + I_{\text{ph}} \]
\[ = I_0 \left( \frac{q V_b}{e k_B T} - 1 \right) + \eta q n_{\text{ph}}^i. \]  (1.93)

The (1.93) refers to the ideal condition but, in the real behavior of photodiodes, \( I_{\text{dark}} \) does not depend only on the applied voltage and on temperature: in fact it occurs a series of fundamental mechanisms like generation/recombination at defect levels or tunneling processes determining deviations from the ideal case (e.g., \( I_{\text{dark}}(V_b = 0) \neq 0 \)). These are the reasons of our interest in modeling such effects. Let’s analyze now some of the most important kinds of current flowing in the device in dark conditions. An important contribution is due to the surface leakage current, with surface meant as external surfaces and also interfaces between different layers of the device. This term is due to energy levels generated by lattice defects, acting as SRH traps that generate minority carriers and affect the geometry of the depletion region. Supposing a trap level at mid-gap and capture cross sections \( \sigma_n \approx \sigma_p \approx \sigma \) the surface leakage current is [22]

\[ I_s = q A w_{\text{ch}} n_i \sigma v N_t = \frac{q A w_{\text{ch}} n_i}{2 \tau}. \]  (1.94)

in which \( N_t \) is again the trap density (at the surface), \( v \) the carrier thermal speed, \( A \) is the area of involved surface and \( w_{\text{ch}} \) the width of an eventual surface channel. In order to prevent the formation of surface states devices are usually passivated with oxides.

**Figure 1.18.** Left: the same \( p n \) photodiode of Figure 1.17 (reverse conditions) in which the depletion region is drawn. Center: energy band diagram showing electrons (filled dots) and holes (circles) photo-generated in the depletion region thanks to the photon absorption (here is schematized only the intrinsic process) drifting through it. Right: Trend of the electric field as a function of depth.

Despite surfaces are critical structures seen by the lattice defect standpoint (especially regarding the growth processes), trap-induced generation/recombination mechanisms do not occur only through the surface states: in fact the generation \( G_{\text{SRH}} \) within the depletion region can affect heavily the photodiode performance, since at low temperatures its contribution could be more important than the diffusion one. Furthermore, since in
reverse bias conditions the free carrier concentrations in depletion region are less than the corresponding ones at equilibrium, generation processes are prevalent with respect to the recombination. So, for reverse biases

\[ I_{SRH} = qG_{SRH}A_{dep} = q\frac{n_i}{\tau_n n_1 + \tau_p p_1}A_{dep}, \] (1.95)
depending on the temperature through \( n_i \sim e^{-E_g/2k_BT} \). This leads to a temperature at which the two phenomena are balanced: under this point, so at lower temperatures, SRH generation/recombination current dominates with respect to the diffusion one.

The last fundamental mechanism affecting the behavior of the dark current and, thus, of the \( I-V \) characteristic is the tunneling process, i.e. the direct or trap-assisted transition of electrons from valence to conduction band within the depletion region. For a triangular potential barrier the direct band-to-band tunneling current can be expressed by

\[ I_{btb}(V_b, T) = V_b A q^3 \sqrt{\frac{2m^*}{E_g(T)}} \exp \left[ -\frac{4\sqrt{2m^* E_g^3(T)}}{3qhE(V_b)} \right] \] (1.96)

and in the trap-assisted case we have

\[ I_{tat}(V_b, E_t, T) = AN_{dep} \frac{\pi^2 q^3 m^*_t E_{dep}(V_b) M^2}{\hbar^3 (E_g(T) - E_t)} \exp \left[ -\frac{8\pi \sqrt{2m^*_t (E_g(T) - E_t)^3}}{3qhE_{dep}(V_b)} \right] \] (1.97)

where \( M \) is the matrix element of the trap potential energy. We will justify these expressions and explain the relative details in the next chapter. At the moment note that both processes do not depend fundamentally on temperature, except through the thermal gap tuning, thus at high \( T \) tunneling noise in enhanced.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure19.png}
\caption{Schematic structure of a \textit{p-i-n} photodetector in \textit{punch-through} conditions (left); corresponding energy band diagram in reverse bias conditions with the photogenerated carrier drifting through the fully depleted intrinsic region (center); trend of the electric field within the \textit{i}-type region (right).}
\end{figure}

Since most of the photogenerated carriers comes from the depletion region (or, at least, at a distance from its edges less than the diffusion length) we must take care in designing
1.3 Fundamental Concepts concerning IR Photodetectors

and engineering the structure of our photodiode. In the case of narrow-gap junctions, like for HgCdTe, operating at high temperatures it is difficult to find a compromise between the optimum absorption length \( z \approx 1/\alpha \), here much larger than in cooled devices, and the diffusion length. One solution should be the formation of multiple heterojunctions. An alternative consists in increasing the width of the depletion region by inserting between the two doped regions an intrinsic layer, i.e. an undoped or intentionally lightly doped layer \( (N \sim 10^{14}-10^{15} \text{ cm}^{-3}) \): the \( p-i-n \) structure.

Due to its low free carrier density the \( i \)-type region is fully depleted almost for any applied \( V_b \); this yields the advantages of a larger and tailored \( w_{dep} \) without the risk of breakdown phenomena. Moreover with a shallow \( pi \) junction (in the case of \( p \)-on-\( n \) device, or \( ni \) junction otherwise) with respect to the illuminated surface we can minimize the carrier diffusion outside the intrinsic layer.

The intrinsic layer width \( w_i \) is a crucial parameter in \( p-i-n \) photodiodes: in order to have a small response time – a parameter linked to the transit time of diffused and drifted carriers through the depleted region and the RC time constant of the whole device/circuit system – it is required a small width, but the need of a quantum efficiency as much as possible yields to a wider \( w_i \). Unfortunately the transit time can not be reduced incrementing the bias voltage, since also at low \( V_b \) the electrons (more rapid then holes) move at their saturation velocity, nonetheless a trade-off in the choice of \( w_i \) must be found, for example in device configurations with lateral light incidence or at a specific angle with respect to the device side.

The two main photocurrent components in \( p-i-n \) structures in the hypothesis of \( \eta = 1 \) are [22]:

\[
I_{\text{drift}} = q(1 - r)n_{ph}^i (1 - e^{-\alpha w_{dep}}) \tag{1.98}
\]

and

\[
I_{\text{diff}} = q(1 - r)n_{ph}^i \left( \frac{\alpha L_e}{1 + \alpha L_e} \right) e^{-\alpha w_{dep}} + \frac{q D_e n_p L_e}{L_e} \tag{1.99}
\]

determining

\[
I_{ph} = I_{\text{drift}} + I_{\text{diff}} \tag{1.100}
\]

\[
= q(1 - r)n_{ph}^i \left( 1 - e^{-\alpha w_{dep}} \right) + \frac{q D_e n_p L_e}{L_e}
\]

We already mentioned the presence of other possible structures for photon detection. An important issue is represented by the avalanche photodetectors (APDs), a modification of the basic \( pn \) junction operating at high reverse bias regimes near the breakdown voltage, as depicted in the inset plot of Figure 1.17. In APDs the carriers interact through impact ionization and the very high electric field at the junction enhances this process accelerating the carriers, which in turn excite other carriers in a multiplication mechanism. The thinner is the multiplication region, the higher can be the driving electric field. This leads to the separation between absorption layer (which has got the requirements just seen in \( p-i-n \) photodiodes) and the multiplication one.

Figure 1.20 shows the APD principle of operation: a photon can be absorbed in the intrinsic region, generating an electron-hole pair. The pair is separated, so the hole moves to the \( p^+ \)-type while the electron drifts towards the \( p \) avalanche region, where occurs the multiplication. In fact, thanks to the high electric field, this photogenerated electron can gain sufficient energy (\( \gtrsim 3/2 E_g \)) to produce other pairs through impact ionization processes with lattice atoms. In such way the photocurrent is amplified. The photocurrent, whose main contribution in APD is the drift one, is modulated by the multiplication factor \( M \)
1.3 Fundamental Concepts concerning IR Photodetectors

Figure 1.20. Structure and principle of operation of an APD. The upper scheme represents a possible configuration, composed by an heavily doped \( n^+ \) contact region, the avalanche \( p \) layer (here not in scale, it must be smaller than this), then the intrinsic absorption layer (the label \( \pi \) stands for lightly \( p^- \) doped but is also possible a \( \nu \) lightly \( n^- \) doped) and finally the other contact \( p^+ \) region. Lower graphs: band diagram in which is evident the carrier multiplication, trend of the electric field along the device depth \( z \).

through the expression

\[
I_{ph}^{av} = M I_{ph},
\]

in which \( I_{ph}^{av} \) is the photocurrent in avalanche regime and \( I_{ph} \) the same current but at low bias, with

\[
M = \frac{1}{1 - \left| \frac{V_b}{V_{BD}} \right|^m} \quad (1.101)
\]

where \( m \) is a coefficient depending on the material and its doping profile. Another relevant parameter in APDs is the ionization coefficient ratio

\[
\kappa_{\text{ion}} = \frac{\alpha_{e,\text{ion}}}{\alpha_{h,\text{ion}}} \quad (1.102)
\]

in which appear the ionization coefficients for electrons and holes \( \alpha_{e,\text{ion}} \) and \( \alpha_{h,\text{ion}} \) that are ionization probabilities per unit length directly proportional to \( \mathcal{E} \) and inversely with respect to the temperature.

Besides the basic configurations illustrated above there are other structures designed for specific applications, such as Schottky barrier photodiodes, Metal-Semiconductor-Metal (MSM), Metal-Insulator-Semiconductor (MIS) or nBn structures. We do not discuss these configurations because by the tunneling standpoint they represent an useless complication, except for some of them which we will encounter discussing the state of the art of the HgCdTe infrared photodetectors. Nonetheless in Figure 1.21 there is a representation of their operating energy band diagrams.
1.3 Fundamental Concepts concerning IR Photodetectors

Figure 1.21. Energy band diagram of different photodetector structures: a) forward biased Schottky barrier on n-type semiconductor showing some processes (1 – electron emission; 2 – tunneling; 3 – recombination in depleted region; 4 – recombination in neutral region); b) biased Metal-Semiconductor-Metal structure; c) biased Metal-Insulator-Semiconductor (n-type); d) biased nBn structure.

1.3.2 Figures of Merit and Performance Limitations

Let’s continue the analysis of photovoltaic detectors browsing the most important features that must be accounted for, especially with regard to the narrow gap semiconductors as HgCdTe alloys. In the previous section we cited the quantum efficiency, but regarding the simple case of a pn photodiode we still have to mention the different contributions coming from each region of the junction. It has been found [1] that:

\[
\eta = \eta_p + \eta_{dep} + \eta_n,
\]

where:

\[
\begin{align*}
\eta_p &= \frac{(1-r)\alpha_L}{\alpha^2 L_e^2} e^{-\alpha x_p} \times \\
&\times \left[ e^{-\alpha(t+d-x_p)} - \frac{\frac{\alpha L_e}{t} e^{\alpha L_e}}{\frac{\alpha L_e}{t}} \cosh \left( \frac{t+d-x_p}{\alpha L_e} \right) - e^{-\alpha x_n} \right] + \alpha L_e \\
\eta_{dep} &= (1-r) \left[ e^{-\alpha x_n} - e^{-\alpha x_n + w_{dep}} \right] \\
\eta_n &= \frac{(1-r)\alpha L_n}{\alpha^2 L_n^2} \left( e^{-\alpha x_n} \left[ \cosh \left( \frac{\alpha L_n}{t} \right) + \sinh \left( \frac{\alpha L_n}{t} \right) \right] - \alpha L_n e^{-\alpha x_n} \right)
\end{align*}
\]

where \( x_p = x_n + w_{dep} \). Since normally the device is designed so that most of the radiation is absorbed in the upper layer (through a thin lower n region or by using an heterojunction with \( E_{g,n} > E_{g,p} \)), but keeping always the back contact several minority carrier diffusion lengths away from the junction, the quantum efficiency is given by:

\[
\eta(\lambda) = (1-r) \frac{\alpha(\lambda) L_e}{1 + \alpha(\lambda) L_e},
\]

in which we assume no reflection at the back contact and no surface recombination phenomena. If, instead, the contact is closer to the junction the previous relation reduces
to
\[ \eta(\lambda) = (1 - r) \left[ 1 - e^{-\alpha(\lambda)t} \right], \] (1.105)

where \( t \) is again the thickness of the \( p \)-type region. In \( p-i-n \) photodiodes, when the intrinsic region is fully depleted (i.e. in punch-through conditions) and supposing that all the photocarriers in the \( i \)-type region – and outside it at a distance less than their \( L \) – are contributing to the photocurrent the (1.85) and (1.99), in which the last term is neglected, yield
\[ \eta_{pin} = (1 - r) \left( 1 - e^{-\alpha \frac{w_{dep}}{1 + \alpha L_e}} \right). \] (1.106)

Another relevant parameter in the description of such devices is the resistance at zero-bias
\[ R_0 = \frac{\partial V_b}{\partial I} \bigg|_{V_b=0}. \] (1.107)

Also typical is the following modification of the (1.107)
\[ R_0 A = \frac{\partial V_b}{\partial J} \bigg|_{V_b=0} \] (1.108)
in which we used the definition of current density \( J = I/A \). Substituting in this expression the various currents already analyzed we obtain different \( R_0 A \) values for different devices. This makes the zero bias resistance an useful parameter in determining the quality of a structure, being sensible to almost all of its important technological and electronic properties. In general for high performances a high \( R_0 A \) is required.

Two parameters strictly related to the optical behavior of a photodetector are the responsivity \( S \) and the detectivity \( D \). The first is defined as the ratio between the photo-induced current \( I_{ph} \) (or photo-induced voltage) and the incident power \( P_i \)
\[ S = \frac{I_{ph}}{P_i g} = q \eta \frac{g}{h \nu} \] (1.109)

where \( g \) is a photoelectric current gain term, accounting for the particular \( V_b \) and the electrical frequency at which the device operates. Before describing the second parameter let’s introduce the concept of noise. There are mainly two kinds of noise in photodetectors: quantum noise (shot noise) and thermal noise (Johnson or Nyquist noise). The shot noise, very determinant especially in \( p-i-n \) photodiodes, involves the random motion and photogeneration process of carriers and reflects itself on the photocurrent. The shot noise mean square current is
\[
\begin{cases}
I_{sn}^2 = 2 q I_{tot} \Delta f & \text{for small } f \\
I_{sn}^2 = \frac{4}{3} q I_{tot} \Delta f & \text{for higher } f
\end{cases}
\] (1.110)

where \( \Delta f \) is the frequency noise bandwidth (i.e. the FWHM of \( \alpha(f) \) peak). Thermal noise, very crucial in LWIR detectors because of the competition between narrow energy gap and thermal energy, is produced in the device and in the load resistance outside the thermal equilibrium condition and consists of a random carrier motion. In this case
\[ I_{Jn}^2 = 4 \frac{k_B T}{R_{eq}} \Delta f \] (1.111)
where the equivalent resistance holds the zero bias resistance $R_0$ in case of thermal equilibrium. Because of its statistic nature also generation/recombination processes contribute to the noise through

$$T_{GRn}^2 = 2q^2(G + R)\Delta f g^2$$ (1.112)

The generation term is often represented by the thermal $G_{th}$ (as in the Johnson noise) or optical $G_{op}$ generation rate. In the latter case

$$G_{op} = n_{ph}^i \eta = \Phi A \eta,$$ (1.113)
due to fluctuations of the incident flux $\Phi$. Neglecting recombination processes ($R_{op} \sim 0$) the mean square optical noise current becomes

$$T_{opn}^2 = 2q^2 n_{ph}^i \eta \Delta f g^2.$$ (1.114)

The generic $T_n^2$ allows us to define a new parameter, the noise equivalent power NEP$_\lambda$, i.e. the power of the incident radiation – for a fixed $\lambda$ – at which the ratio between signal and noise

$$s/n \triangleq \sqrt{\frac{T_{tot}^2}{T_n^2}}$$ (1.115)
equals to 1, then (if $g = 1$):

$$\text{NEP}_\lambda = \frac{1}{S} \sqrt{\frac{T_n^2}{\Delta f}}$$ (1.116)

Then we can define the detectivity

$$D = D_\lambda = \frac{1}{\text{NEP}_\lambda}$$ (1.117)

and the normalized detectivity

$$D^* = D \sqrt{A \Delta f} = \frac{\sqrt{A \Delta f}}{\text{NEP}_\lambda}.$$ (1.118)

Now, accounting for a noise current at thermal equilibrium dominated by optical and thermal contributions such as

$$T_n^2 = 2q^2 \Phi_B A \eta \Delta f + 4\frac{k_BT}{R_0} \Delta f,$$ (1.119)

where $\Phi_B$ is the background flux (we set $\Phi_B$ instead of the signal flux $\Phi$ in the (1.113)), the normalized detectivity (hereafter, only detectivity) holds:

$$D^* = \frac{\eta q}{h \nu} \left[ 4\frac{k_BT}{R_0A} + 2q^2\Phi_B \eta \right]^{-\frac{1}{2}}$$ (1.120)

from which is evident that a way to maximize $D^*$ is, for instance, to design the detector such that also $R_0A$ is maximized. Moreover the formula (1.120) states that there are two main situations:

- the Background LImited Performance – BLIP, in which $4\frac{k_BT}{R_0A} \ll 2q^2\Phi_B \eta$;
- the thermal noise limited performance, if $4\frac{k_BT}{R_0A} \gg 2q^2\Phi_B \eta$.
Note that, like the quantum efficiency, also the detectivity is sensitive to the absorption coefficient $\alpha$:

$$D^* \sim \frac{q \sqrt{A} \Delta f (1 - r) [1 - e^{-\alpha z}]}{h \nu \sqrt{T_n^2}}.$$  \hspace{1cm} (1.121)

An important limiting phenomenon that we take into account is the detector frequency response. Mainly affected by the transit time of carriers within the depletion region and by the CR-time constant, of the device and the bias circuit, it results in a limited operation bandwidth when the incident light is modulating. Suppose that the junction is exactly located at the optimum depth $z = 1/\alpha$, then the time required to reach the junction through diffusion is \[19\]

$$\tau_d = \frac{1}{2.43 \alpha^2 D_e}$$  \hspace{1cm} (1.122)

and the corresponding roll-off frequency (i.e. the frequency at which the signal reduces by $1/\sqrt{2}$)

$$f_d = \frac{1}{2\pi \tau_d}.$$  \hspace{1cm} (1.123)

The required time to cross the junction and then to transit within the whole depleted region is

$$\tau_{tr} = \frac{w_{dep}}{2.8 v_{sat}} = \frac{1}{1.2 v_{sat}} \sqrt{\frac{\epsilon V_{dep}}{q N_{tot}}}$$  \hspace{1cm} (1.124)

where $v_{sat}$ is the carrier saturation velocity determined by the electric field, $V_{dep}$ is the potential drop across the depleted region and $N_{tot}$ the carrier concentration, assumed to be the same on both sides of an abrupt $pn$ junction. Then, again

$$f_{tr} = \frac{1}{2\pi \tau_{tr}}$$  \hspace{1cm} (1.125)

(as an example in extrinsic $p$-type $\text{Hg}_{0.832}\text{Cd}_{0.168}\text{Te}$ has been found \[23\] $\tau_{tr} \sim 100$ ps, but further reductions are possible with thinner absorber layers). With a correct device design one can reduce the impact of these two contributions. The most serious limitations arise from the RC-time delay. Since the depletion layer acts as a capacitance, in a $p-i-n$ photodiodes at the punch-through and with an intrinsic layer of width $w_i$ we can write

$$C_j = \frac{\epsilon}{w_i}$$  \hspace{1cm} (1.126)

and then

$$f_{RC} = \frac{1}{2\pi C_j R_{||}}$$  \hspace{1cm} (1.127)

in which the resistance

$$R_{||} = R_j \parallel R_L = \frac{R_j R_L}{R_j + R_L}$$  \hspace{1cm} (1.128)

accounts for both junction and load resistance. A way to reduce $R_j$ is for example to operate after punch-through.

The last figure of merit in which we are interested is a quantity derived from two parameters already mentioned and is represented by the ratio $\alpha/G_{th}$. In general it is recommended a high value of $\alpha/G_{th}$, especially in the IR detection: under this point of views the $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ proves once again an excellent material for IR detection, since it is characterized by high coefficients $\alpha$ and low rates $G_{th}$. Moreover, at a given composition $x$ and temperature $T$, it shows a decreasing dependence of $\alpha/G_{th}$ on the cut-off wavelength $\lambda_c$. 
1.3.3 The State of the Art about HgCdTe IR Photodetectors

Direct bandgap photodiodes like Hg$_{1-x}$Cd$_x$Te-based IR photodetectors suffer mainly the influence of generation/recombination processes of minority carriers injected from the contacts to the absorber region but, next to the control of GR limiting factors, the greatest challenge are the BLIP operating mode and the Auger suppression. Moreover the LWIR detection without stringent cooling requirements is a further wanted property. If the influence of surface states can be reduced with passivation techniques or through the employment of heterojunctions and thermal/optical generation terms can be kept under control via Burnstein-Moss effect in presence of high doping regions, the tunneling remains a fundamental quantum effect that is difficult to manage.

![Figure 1.22. Summary of the most common effects determining performance an limitations in $pn$ or $p-i-n$ narrow gap IR photodetectors.](image)

All these features characterized the development of HgCdTe-based IR photodetectors in the last decades and the efforts directed towards the control of dark current modeled not only the operating modes, but especially the structures of such devices. Piotrowski [23], for instance, stated that at liquid nitrogen temperature (77 K) most of the phenomena affecting the raise of a dark current are non-fundamental mechanisms, while at higher temperatures the fundamental ones become protagonist. An improved fabrication technology and material quality should reduce SRH and GR processes, Trap-Assisted Tunneling, 1/$f$ noise and surface recombination; a correct choice in doping and doping profile should control Band-to-Band Tunneling and Auger effects while, finally, a proper design could lower the Johnson noise.

The first generation of HgCdTe photodetectors, based upon the simple $pn$ junction, saw both $p$-on-$n$ and $n$-on-$p$ solutions. Remembering that Auger 1 process is mainly present in $n$-type HgCdTe and Auger 7 in $p$-type we can write the two corresponding $R_0 A$ products as

\[
(R_0 A)_{A1} = \tau_{A1,i} \frac{2k_B T}{q^2 N_a (t + d)},
\]

and

\[
(R_0 A)_{A7} = \tau_{A7,i} \frac{2k_B T}{q^2 N_d (t + d)},
\]
where $d$ is the base layer width. Since $\gamma = \tau_{A7,i}/\tau_{A1,i} > 1$ then higher values of $R_0 A$ are achievable in $p$-type absorption layers (with respect to the $n$-type at the same doping concentration), therefore the relations (1.129) and (1.130) stated that a thin absorber region width $t$ is a possible requirement for maximizing the term $R_0 A$. But the latter and, contemporarily, the need of small widths for all layers (for the bulk dark current reduction) are difficult parameters to manage by the technological point of view. Moreover thin layers also introduce unwanted surface GR mechanisms. As already said, using degenerated $n^+$ regions, these mechanisms can be controlled: so $n^+-p$ homojunction structures began to be developed. In order to manage also the other fundamental and non-fundamental sources of dark current we moved towards the use of nBn structures or heterojunctions with an increasing number of layers, up to five-layers configurations.

Other ways to maximize $R_0 A$ in $p$-type absorption layers can be obtained by setting $p = p_0 = \sqrt{\gamma} n_i^2$, which should coincide with the minimum of thermal generation $G_{th}$, or – in the case of “double radiation pass devices” – choosing a base layer width $d = 0.63/\alpha$ [1]. For SWIR detection an optimum configuration is represented by a device operating above 200 K with doping in the range $10^{16} - 5 \cdot 10^{17}$ cm$^{-3}$, while for cooled MWIR detectors we must stay below $10^{16}$ cm$^{-3}$. Thus, according to their doping, actually there are three main kind of devices, divided into near intrinsic, just extrinsic and extrinsic photodetectors. The first one includes devices with a very short central zone ($\pi$ or $\nu$) devoted to the Auger suppression, the second one which shows an extrinsic behavior already at zero bias and the last one, especially $p$-type, characterized by light doping and lowest currents at low bias. Auger suppression at higher voltages is also present. This result is often obtained also through so-called non-equilibrium devices: the most used are $n^+-\pi-p$, $n^+-i-p^+$, $n^+-\nu-p^+$ or $n^+-i-p^+$ structures, which exploit both minority and majority carrier currents. These third-generation HgCdTe structures, besides the multi-spectra structures, represent the present state of the art in the field of IR detection.

Under the frequency response it is noticeable the result of minority carrier drift transit time in $\pi$ intrinsic region of 36 ps, referred to an HgCdTe-based photodetector for 10.6 $\mu$m and operating at 260 K. Unfortunately RC-like transit time are pretty greater and are situated in the order of a few nanoseconds. Nonetheless well-designed structures can operate in the gigahertz detection domain.

Regarding the quantum efficiency, it is also known that HgCdTe operating at near room temperature are very poor in performance. A solution is represented, again, by the use of multiple heterojunction structures, like $n^+-p-P$ (where the capital letter stands for a wider bandgap). In such devices $\eta$ can reach also the 90%.

From the previous information it appears so clear why nowadays HgCdTe is considered the best choice for IR detection. Its principal sources of dark current are, fortunately, non-fundamental effects or almost controllable physical phenomena through high quality growth techniques and device fabrication. In fact, once the Auger suppression no longer represents a challenge, also trap-assisted tunneling – which is maybe considered the most noising effect of the present devices operating at moderate bias and low temperature – is partially controllable avoiding the presence of lattice defects. The influence of band-to-band tunneling, by the way, still remains a matter of study especially at high reverse bias. Probably due to their marked unpredictable nature, tunneling processes are one of the most investigated under the simulation point of view.
References


[5] A. D. D. Dwivedi, “Analytical modeling and numerical simulation of $P^+\text{Hg}_{0.69}\text{Cd}_{0.31}\text{Te}/n\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}/\text{CdZnTe}$ heterojunction photodetector for a long-wavelength infrared free space optical communication system”, J. Appl. Phys. 110, 143101 (2011)


The trend of progressively shrink the dimensions of electronic devices according to the Moore’s law is making raise the need to design in a more accurate way not only the structures, but especially all physical effects concerning their operation. Since practical investigation on these aspects may be a priori too expensive, because they require realization and testing procedures, numerical simulations play a key role in predicting the behavior of devices. By contrast, modeling often implies a simplification of the real problem in order to perform a low-computational-cost algorithm. For this reason, and because of the complexity of the physics behind such phenomena, there exist multiple strategies which allow to model the operating mechanisms (wanted or not) of micro- and nano-devices, that will be analyzed in the following.

In this chapter we focus on one of the most important factor to be accounted for in describing the behavior of semiconductor devices: the tunneling effect. It involves generation/recombination processes that can limit its operation because of the rise of leakage currents, especially at low temperatures. Since the width $W$ through which the electron must tunnel is a function of the external electric field applied to the semiconductor $W = \frac{V_{\text{gap}}}{E}$ we can easily see that increasing $E$ the probability of tunneling, and hence, the tunneling current is enhanced. But the tunneling is not only an undesirable quantum effect by the electronic point of view: because it is an intrinsic phenomenon in the matter it can statistically occur even if the semiconductor device is in the off-state and, speaking in terms of photodetectors, in dark conditions.

Thanks to its quantum nature, tunneling can not be eliminate completely. For this reason and in order to reduce it as much as possible, the simulation concepts becomes so important in predicting the device behavior. The purpose of this chapter is just to state all the fundamental knowledge necessary to understand the theory of tunneling, its quantum formalism and to insert it into the general frame of the physics-based numerical simulations, at least for a mono-dimensional case. At the end we are able to write the theoretical and the modeled expression for the electron current generated by the tunneling process, with particular attention to the class of direct- and narrow-bandgap semiconductors.
2.1 The Semi-classical Approximation of the Quantum Mechanics

Since the quantum description of a physical system implies also the classic one, this yields that the boundary between the two formalisms consists in the theoretical limit $\hbar \to 0$. By the practical standpoint the latter corresponds to considering a system in which $\hbar$ (that is, physically, an action) is negligible with respect to all the other actions of the system itself. More simply, we can speak in terms of geometrical optics: the same transition occurs if the De Broglie wavelengths ($\lambda = h/k$) are small with respect to the typical dimensions of the system. A way to justify and demonstrate this assumption passes through the analysis of a simplified solution of the Schrödinger’s equation [24, ch.10]: this approach, known as the WKB approximation, was proposed independently by Wentzel, Kramers and Brillouin in 1926.

It is worth noting, firstly, that if we take $\hbar = 0$ in the generic expression of the Schrödinger’s equation
\[ i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \Psi + U(r)\Psi \] (2.1)
it becomes meaningless. But let’s see what happens if we take a wavefunction $\Psi = \Psi(r, t)$ like
\[ \Psi(r, t) = Ae^{iS(r, t)/\hbar}, \] (2.2)
where $A$ is a normalization constant and the function $S(r, t)$ has the same dimensions of $\hbar$ (it is an action). In this way, the equation (2.1) becomes
\[ -\frac{\partial S}{\partial t} = \frac{1}{2m} (\nabla S)^2 - \frac{ih}{2m} \Delta S + U(r). \] (2.3)
Neglecting now the term containing $\hbar$ we found the classical form of the (2.1), i.e. the so-called Hamilton-Jacobi equation for the Hamilton’s principal function, applied to the classical action $S_c(r, t)$, a real function depending on the Lagrangian through
\[ S_c(r, t) = \int_{t_0}^{t} L(r, \dot{r}, \tilde{t}) \, d\tilde{t}. \] (2.4)
So we can say that, for $\hbar = 0$, the wavefunction (2.2) represents the classical trajectory since thanks to the Hamilton-Jacobi equation
\[ -\frac{\partial S_c}{\partial t} = H(r, p), \] (2.5)
where
\[ H(r, p) = \frac{p^2}{2m} + U(r), \] (2.6)
the momentum $p$ depends on the action through the well known fundamental relationship
\[ p = h k = \nabla S_c. \] (2.7)
This means that, in the classical limit, all the possible trajectories are orthogonal to the surfaces with $S$ constant. In other words, being in this case $S_c$ a real function, the particles describe trajectories orthogonal to the surfaces with wavefunction’s $\Psi$ constant phase. Moreover, this limit is valid if the last term in (2.3) is negligible, i.e. if the potential $U(r)$ is sufficiently small with respect to the other terms (and, obviously, to the effect of
an eventual external electric field $E$) or for particles with high momentum $p$ moving in a slowly varying potential (w.r.t. the lattice constant).

Let’s now explicit the different dependencies within the action $S$ as

$$S(r, t) = \sigma(r) - Et \quad (2.8)$$

(where $\sigma$ is the Hamilton’s principal function and $E$ is the energetic term) and consider, for simplicity, the case of only stationary states. The WKB method consists in postulate the following development for $\sigma(r)$:

$$\sigma(r) = \sigma_0(r) + \frac{\hbar}{i} \sigma_1(r) + \left( \frac{\hbar}{i} \right)^2 \sigma_2(r) + \ldots \quad (2.9)$$

and in determine $\sigma$ (thus, the spatial part of $\Psi$) through successive approximations. Substituting (2.8) in the Schrödinger’s equation we obtain:

$$\left( \nabla \sigma_j \right)^2 - i\hbar \Delta \sigma_j = 2m \left[ E - U(r) \right], \quad (2.10)$$

where the subscript $j = 0, 1, 2, \ldots$ represents the $j$-th order in the power development (2.9). Then, inserting the (2.9) in (2.10) yields a system of coupled equations for the unknown $\sigma_j$ that must be solved in sequence, from the first order to the last one:

$$\begin{cases}
\left( \nabla \sigma_0 \right)^2 = 2m \left[ E - U(r) \right] \\
\left( \nabla \sigma_0 \cdot \nabla \sigma_1 \right) + \frac{1}{2} \Delta \sigma_0 = 0 \\
\left( \nabla \sigma_1 \right)^2 + 2 \left( \nabla \sigma_0 \cdot \nabla \sigma_2 \right) + \Delta \sigma_1 = 0 \\
\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \li
Note that $|\Psi| \propto p^{-1/2}$ means that the probability of a particle described by this kind of wavefunction to be located in the interval $[x; x + dx]$ is $|\Psi|^2 dx \propto dx/p \propto dt$; hence it is proportional to the time $dt$ necessary to travel into the segment $dx$.

The (2.15) is the most general form for the $\Psi$. It includes two linearly independent solutions (thanks to de twice sign) of the Schrödinger’s equation and it is formally valid in each region of the space in which the WKB is applicable. In the next paragraph we will distinguish the case of a classical allowed region ($E > U(x)$ and wave number $k$ real), where $\Psi$ is oscillatory, from the forbidden one ($E < U(x)$ and $k$ complex) where, instead, the wavefunction takes another form. The points $x = x_i$ such that $E = U(x_i)$ are defined turning points of the classical motion. Since $p(x) = \sqrt{2m[E - U(x)]}$ then $p(x_i) = 0$ and the classical solution has no sense, here; so for this reason we will compute the opportune calculations to connect the two solutions around this points in the case of a potential barrier in the tunneling process. Furthermore, the WKB will ensure us that we can design this connection without specifying exactly the shape of the barrier.

### 2.2 The Mechanism of Tunneling in Semiconductors

The simplest explanation of what the tunneling is involves first of all the description of an electron confined in the valence band of our semiconductor as a traveling wave (because of the wave-particle duality). Its motion consists of an oscillating trajectory between the bottom and the top of the band [25, ch.12] (see Figure 2.1), i.e. at $x = a$ (the opposite happens at $x = b$ for holes in conduction band, but for a less tiresome treatment we will discuss only the electrons). Those, $x = a$ and $x = b$, are our turning points. Following the semi-classical propagation theory, at the two band edges the wave number $k$ will be zero, real within each band (i.e. for $x < a$ or $x > b$) since states are propagating waves, and in the forbidden region ($a < x < b$) $k$ is imaginary. This means that an exponentially decreasing probability (located in the region II) to pass through the middle gap is not null. When the electron overcomes the turning point and reaches the conduction band then the tunneling occurs.

According to the WKB approximation we know that our electron in valence band (a classical allowed region) is represented by a 1D wavefunction of the form (2.15). Since this semi-classical solution for $\Psi$ is valid in the two allowed regions far from the turning points (several wavelengths), we must find an overall wavefunction that joins smoothly with this WKB approximation. To do this without knowing the actual shape of the energy barrier we suppose a continuous variation of wave vectors $k$ within the forbidden region with a maximum located at the mid-gap. The simpler analytical form for the barrier of a direct band-gap semiconductor that corresponds to these requirements is the following one:

$$E - U = - \frac{E_g^2}{E} - \frac{E_{\text{mid-gap}}^2}{E_g}$$  \hspace{1cm} (2.16)

Let’s introduce now an injection of $n$ electrons with rate $1/\tau$ on the left side of the scheme in Figure 2.1 and consider firstly the region III ($x > b$) where we have only a transmitted wave towards the increasing $x$. The way to find a compromisingly solution of the Schrödinger’s equation such that $\Psi$ is incident from the left in region I, then for $x = a$ is partially reflected and partially transmitted with probability $P$ in the subsequent regions II and III, is the change of variable $\Psi = \varphi \sqrt{z}$ where

$$z = \int_{b}^{x} \sqrt{\frac{2mF}{\hbar^2}} (x - b) \, dx$$  \hspace{1cm} (2.17)
2.2 The Mechanism of Tunneling in Semiconductors

Figure 2.1. Schematic sketch (adapted from [25]) of the 1D potential barrier \((E-U) = -(\hbar k_x/2m)^2\) (dashed curve) and the energy distribution \(E\) along a p-i-n junction with \(E_{mid-pep} = 0\). If the bandgap is \(E_g\) and the electric field is \(E\), then \(|a-b| = E_g/qE\). Note that if \(E_\perp > 0\) the turning points become \(a'\) and \(b'\) and the tunneling probability is reduced because of the widening of the forbidden region.

Then substituting into the Schrödinger’s equation evaluated near the turning point \(b\) (for \(E = 0\))

\[
\frac{d^2\Psi}{dx^2} + \frac{2mF}{\hbar^2}(x-b)\Psi = 0 \tag{2.18}
\]

where \(F\) has the dimensions of a potential, yields the Bessel’s differential equation

\[
\frac{d^2\varphi}{dz^2} + \frac{1}{z} \frac{d\varphi}{dz} + \left(1 - \frac{1}{9z^2}\right)\varphi = 0 \tag{2.19}
\]

so that \(\varphi\) assumes the meaning of a Bessel function of order \(\pm\frac{1}{3}\), i.e. the solution of an equation like the one in (2.19), and then \(\varphi = J_{\pm\frac{1}{3}}(z)\).

Remembering that, according to the definition of Bessel function of \(\nu\)-th order

\[
J_{\nu}(z) = \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{z}{2}\right)^{\nu+2n}}{n! \Gamma(\nu + n + 1)}, \tag{2.20}
\]

we can write the wanted wavefunction as

\[
\Psi = \sqrt{z} \left(AJ_{\frac{1}{3}}(z) + BJ_{-\frac{1}{3}}(z)\right) \tag{2.21}
\]

Since we are interested in an effective form of the \(\Psi\), such that near \(a\) and \(b\) the expression (2.21) is met and far from there the WKB approximation is valid, we rewrite the wavefunction as

\[
\Psi \approx \sqrt{\frac{\zeta}{k_x}} \left(AJ_{\frac{1}{3}}(\zeta) + BJ_{-\frac{1}{3}}(\zeta)\right) \tag{2.22}
\]
2.2 The Mechanism of Tunneling in Semiconductors

with

\[ \zeta = \int_b^x \sqrt{\frac{2m}{\hbar^2}}(E-U) \, dx \quad (2.23) \]

Note that the form (2.22) completely satisfies the requirements mentioned above about the behavior of \( \Psi \). This means that in \( x = b \) the \( \Psi \) evaluated in region II and III has the same asymptotic form. Moreover for arguments \( \zeta \in \mathbb{R} \) the \( \Psi \) acts in the region III and I, while for imaginary values it is valid in the interval \( x < b \). The matching condition for the wavefunction at \( x = a \) is similarly given by

\[ \zeta' = \int_x^a \sqrt{\frac{2m}{\hbar^2}}(E-U) \, dx \quad (2.24) \]

and through the substitution of coefficients

\[ A = B \leftrightarrow A' = \frac{ie^\Lambda}{\cos \left( \frac{\pi}{6} \right)} \quad (2.25) \]

where

\[ \Lambda = |\zeta| + |\zeta'| = \left| \int_a^b \sqrt{\frac{2m}{\hbar^2}}(E-U) \, dx \right| \quad (2.26) \]

we obtain the ratio of transmitted flux to incident flux (i.e. the probability) \( P = e^{-2\Lambda} \) so the tunneling current will be:

\[ I_{\text{tunn}} = \frac{q}{\tau} n P = \frac{q}{\tau} n e^{-2\Lambda} \quad (2.27) \]

The above calculations do not account for an important solid-state-physics principle, that is the presence or not of available states in both energy bands. This fact, for example, results in the behavior of the tunneling current in \( pn \) junctions or, more specifically, in the Esaki \( p-i-n \) diode: while at thermodynamic equilibrium the electron and hole current are balanced, since all the states below the Fermi energy are almost full and above are empty, with a small forward bias voltage electrons can tunnel from the \( n \) to the \( p \) side, and in presence of a reverse bias the lack of available states makes occur the opposite. For the first time L. Esaki noted in 1958 that incrementing the forward voltage firstly the current begins dropping to zero for the presence of a forbidden region in the \( p \) side adjacent to the electrons in \( n \), and then the \( I-V \) characteristic rises up and the diode behaves like a normal \( pn \) junction.

From these considerations emerges that the probability \( P \) must include a term related to the availability of states. Now taking into account the formula (2.16) where the energy \( E \) is given by the sum of a component \( E_\perp \) associated to the wavevector \( k_\perp \) perpendicular to the tunneling trajectory and the component \( E_\parallel \) associated to \( k_\parallel \), we can state that, for \( E_\perp > 0 \):

\[ E_\parallel - U = - \left[ \frac{E_{\mbox{2}}^2}{4} - \frac{E_{\mbox{mid-gap}}^2}{E_g} + E_\perp \right] \quad (2.28) \]

If we call the two turning point (i.e. where the (2.28) is null) \( a' \) and \( b' \) then:

\[ \Lambda = \int_{a'}^{b'} \sqrt{\frac{2m}{\hbar^2}} \left[ \frac{E_{\mbox{2}}^2}{4} - \frac{E_{\mbox{mid-gap}}^2}{E_g} + E_\perp \right] \, dx \quad (2.29) \]
2.2 The Mechanism of Tunneling in Semiconductors

Taking now \( E_{\text{mid-gap}} = 0 \) as the origin of the system we have, in general, \( E_{\text{mid-gap}} = q\mathcal{E}x \) with the turning points at

\[
x = \pm \frac{1}{q\mathcal{E}} \sqrt{\frac{E_g^2}{4} + E_g E_\perp},
\]

so

\[
\Lambda = \frac{\pi \sqrt{m^* E_g^3}}{4\sqrt{2q\mathcal{E}} \hbar} + \frac{\pi \sqrt{m^* E_g^3 E_\perp}}{\sqrt{2q\mathcal{E}} \hbar} = \frac{E_g}{2E} + \frac{2E_g E_\perp}{E}. \tag{2.31}
\]

Finally:

\[
P = \exp \left[ -\frac{E_g}{4E} \right] \exp \left[ -\frac{E_\perp}{E} \right] = P_0 \exp \left[ -\frac{E_\perp}{E} \right] \tag{2.32}
\]

where \( P_0 \) is \( P(k_\perp = 0) \) and \( \tilde{E} \) is proportional to the wavevector \( k_\perp \) of electrons involved in the tunneling process. Now, taking into account the presence of empty states, we must reduce \( P \) by a factor \( (1 - F'(E)) \), where \( F'(E) \) is the Fermi distribution function \( F(E) \) at the arriving side for tunneling electrons. So

\[
\frac{I_{\text{tunn}}}{A} = \Phi_{\text{inc}} \times P \left( 1 - F'(E) \right) \times A \, dx \]

\[
= \left[ \left( \frac{q\mathcal{E} m^*}{2\pi^2 \hbar^3} \right) F(E) \right] \left[ P_0 \exp \left[ -\frac{E_\perp}{E} \right] (1 - F'(E)) \right] \, dE_\perp \, dx
\]

that, integrated with respect to the energy under the assumption of discontinuous \( F(E) \), yields

\[
\frac{I_{\text{tunn}}}{A} = \frac{gm^*}{2\pi^2 \hbar^3} P_0 \tilde{E} \left\{ qV_a - 2\tilde{E} \left[ 1 - \exp \left( -\frac{qV_a}{2\tilde{E}} \right) \right] \right\}
\]

\[
\approx \sqrt{2m^* q^3 \mathcal{E}} V_a \frac{\pi \sqrt{m^* E_g^3}}{2\sqrt{2q\mathcal{E}} \hbar} \exp \left[ -\frac{\pi \sqrt{m^* E_g^3}}{2\sqrt{2q\mathcal{E}} \hbar} \right] \tag{2.34}
\]

with the average field

\[
\mathcal{E}_{av} = \frac{V_g + V_a}{W}, \tag{2.35}
\]

\[
qV_g = E_g \tag{2.36}
\]

with \( V_a \) the applied bias and where in the last passage we neglected the term \( E_\perp = (\hbar^2 k_\perp) / 2m^* \) in the expression of \( P \) (this is particularly true in direct band-gap materials due to the conservation of the wavevector \( k_\perp \) during the process). For \( \mathcal{E} \rightarrow 10^6 \frac{V}{cm} \) the \( I_{\text{tunn}} \) begins to be non-negligible. Note that for abrupt junctions – in which we are slightly away from the uniform field condition of equation (2.31) – it is verified that an optimum choice is given by imposing:

\[
\mathcal{E}_{av} = \sqrt{\frac{qN_A V_g}{2e}} \tag{2.37}
\]

that is no so much different from the one reported in (2.35).
2.2 The Mechanism of Tunneling in Semiconductors

All the calculations reported above are valid in general, that is for a generic electron tunneling through a potential barrier. The true phenomenological expressions of these physical quantities will be exposed in the following paragraphs, where more specific cases of how tunneling occurs will be treated. Since the tunneling mechanism involves the exchange of carriers – i.e. is a generation/recombination process – between the conduction band and the valence band (or, in some cases, through intraband transitions), it represents a fundamental part of the classical Shockley-Read-Hall theory. As we have just seen, by imposing on the semiconductor a certain value of the electric field we are able to stimulate such a kind of quantum phenomena. By the simplest phenomenological standpoint, one can notice that tunneling may occur directly or through a step-like process. Let’s for example consider an indirect band-gap semiconductor, as in the case of Si and Ge: the mismatch in the k-space between the top of the valence band and the bottom of the conduction band (that occurs for $k \neq 0$) implies that electrons must overcome the forbidden region without the conservation of the perpendicular wavevector. The process needs a further mechanism to keep constant the overall wavevector and this is only possible thanks to the phonon absorption/emission, that makes drastically reduce the tunneling current $I_{\text{tunn}}$.

There are many other possibilities for having a multi-step tunneling process and one of the most important is due to the contribution of a so-called “trap” or “SRH center”, that is a defect (deep impurity, energy level almost in the mid-gap, like a dislocation) and, in very particular conditions, a dopant atom (shallow impurity, energy level close to band edges), located in the energetic forbidden region of the semiconductor [26]. In the next subsections we will consider two dominant mechanisms of tunneling in the simple situation of a biased pn junction: the trap-assisted tunneling TAT and the band-to-band tunneling BTBT. They are some of the most studied because they dominate the behavior of the $I-V$ characteristic of the devices in particular bias conditions and they are temperature-sensitive effects.

### 2.2.1 Trap-Assisted Tunneling

At relatively low electric fields carriers in the valence band can reach an energy trap and then tunnel in the conduction band or, alternatively, undergo a multiphonon transition over only a part of the remaining path towards the conduction band then, finally, the tunnel may occur. It should be noted that, by the quantum standpoint, the electron trapped can stay in this energy level for a finite time before escaping towards the conduction band or, for instance, it may be trapped by another impurity with the same energy, or not, through the so-called percolation path before reaching the conduction band. The final – i.e. statistical – result is schematized in Figure 2.2, where we represented only one energy level (here supposed at the mid-gap) within the forbidden region. Being the tunneling probability $P$ linked to the rate $\mathcal{R}$ at which a particle can tunnel within a potential barrier through the $\mathcal{R} = P N_{(C;V)}$ (where $N_C$ and $N_V$ are respectively the carrier density in conduction and in valence band) the consequent Trap-Assisted Tunneling rate from the SRH center to the conduction band for a pn junction is given by [27]:

$$\mathcal{R}_{\text{tun}}(E_t, T) = \frac{\pi^2 q m^* M^2 E_{\text{dep}}}{\hbar^3 (E_g(T) - E_t)} \exp \left[ -\frac{8\pi \sqrt{2m^* (E_g(T) - E_t)^3}}{3q \hbar E_{\text{dep}}} \right]$$ (2.38)

(in which $M$ is the matrix element of the trap potential energy) where we made explicit the dependence of the rate $\mathcal{R}$ on the energy of traps $E_t$ (computed as the difference between the energy of the level and the intrinsic one) and the temperature, through the gap $E_g(T)$. 

Note that the tunneling probability – and at the same time also $\mathcal{R}(E_t, T)$ – decreases with the increase of the temperature, because of the gap widening and for the carrier mean free-path reduction. Now, being

$$I_{\text{tat}} = q A N_{t} w_{\text{dep}} \mathcal{R}_{\text{tat}}$$

then

$$I_{\text{tat}}(E_t, T, N_t) = A N_{t} w_{\text{dep}} \frac{\pi^2 q^2 m^* \xi_{\text{dep}} M^2}{\hbar^3 (E_g(T) - E_t)} \exp \left[ -\frac{8\pi \sqrt{2m^* (E_g(T) - E_t)^3}}{3q\hbar \xi_{\text{dep}}} \right]$$

In the expression (2.40) – quite similar to the general form written in (2.34) – it appears also the density of traps $N_t$ responsible of TAT (different from the SRH trap density $N_{SRH}$), which is an important parameter in the TAT mechanism since it may affect not only the behavior of the $I$-$V$ characteristic but also the low-frequency noise of the device. The TAT current (like any other tunneling current) may be modified by varying the shape of the energetic bands and this fact is strictly to be accounted for in the description of the depletion region of a particular device. Therefore we can note that the current $I_{\text{tat}}$ decreases with the rising of $E_t$ with an exponential-like behavior. This mechanism tends to decline if we consider higher temperatures, since the higher is the energy of the system the greater is the ionized trap fraction (note the numerator in the negative exponential term depending on $T$).

It must be remarked that, as previously described, the TAT process is a multi-step mechanism, and the nature of each step can be energetically and physically different. The first transition, the one that involves the valence band and the trap level, may be a thermal

![Figure 2.2.](image-url)
or a tunnel process itself. Referring, for example, to the Figure 2.2, the mechanism denoted with arrows (1) and (2) is a thermal-tunnel process (remember that in the sketch of energy bands \( E(x) \) an horizontal transition means a movement of carriers in the real space, that is the tunneling, while a vertical transition is a change of energy). In general there can exist three other paths: tunnel-thermal, tunnel-tunnel and thermal-thermal, where the latter two represent respectively the BTBT and a kind of classical SRH mechanism, no more a TAT. The particular tunneling path can affect, for instance, the net GR rate \( U_{\text{nat}} \) of the process. If thermal transitions are always possible for all traps, the tunneling ones are not: in fact an essential condition within the depletion region – where the bands are bent – is a trap energy higher than the adjacent bottom of the conduction band (the opposite happens for the holes). Obviously, the greater is the electric field (thus, the applied voltage) in the depleted region, the bigger is the slope of the bands and, consequently the higher will be the tunneling probability because of the spatial reduction between the starting and the ending point of the tunneling path.

### 2.2.2 Band-to-Band Tunneling

If the electrons reach the conduction band directly (see Figure 2.3), without any mediator – this requires much more energy (thus, field \( \mathcal{E} \)) with respect to the TAT case – we are in presence of the so-called Band-to-Band Tunneling BTBT. As we can imagine, TAT is more important if the device temperature is not so high (a few tens of K) because impurities are energetically confined in the lattice structure. If the energy of the system, or its temperature (remember that \( E_g \propto T \)), increases these levels are promoted to the conduction state and carriers must tunnel mainly through BTBT processes, given that practically all traps are ionized not yielding more TAT. At the same time the quantum probability of tunneling increases too, so at high temperatures BTBT is the dominant tunneling effect detected. This distinction, thus, is valid not only for the temperature but also for the specific bias value at which the device is operating, since the bias affects the electric field: for example (at room temperature) in high reverse bias condition the BTBT prevails, while for low-to-medium reverse bias values TAT is the predominant mechanism.

Following the same procedure seen in the TAT case let’s write firstly the tunneling rate \( R_{\text{btbt}} \) for a \( pn \) junction [28]:

\[
R_{\text{btbt}} = \exp \left\{ -\sqrt{\frac{4em^*}{q^2\hbar^2N_A}} E_g \left( 1 - \frac{E}{E_g} \right) \left( \frac{\pi}{2} - \sin^{-1} \sqrt{\frac{E}{E_g}} \right) - \sqrt{\frac{E}{E_g}} \right\} 
\]

(2.41)

where \( R_{\text{btbt}} = R_{\text{btbt}}(E, T) \). Now, since

\[
I_{\text{btbt}} = \frac{qA}{4\pi\hbar^2} \sqrt{\frac{E_g k_BT}{Q^2}} \int_{-E_{\text{max}}}^{0} R_{\text{btbt}} \frac{E}{2} dE, 
\]

(2.42)

where \( E_{\text{max}} = -qV - E_F \) and \( Q \) is the momentum matrix element, we finally find:

\[
I_{\text{btbt}}(V, T) = \sqrt{\frac{2m^*}{E_g(T)}} \exp \left[ -\frac{4\sqrt{2m^*E_g^3(T)}}{3q\hbar \mathcal{E}(V)} \right]. 
\]

(2.43)

In the equation (2.43) we have explicitly specified the dependence of \( I_{\text{btbt}} \) with respect to the applied voltage just to recall that the magnitude of the electric field is one of the most responsible of the rise of this kind of current. The previous formula suggests also
2.2 The Mechanism of Tunneling in Semiconductors

Figure 2.3. Schematic sketches for a BTBT process. Upper: we assume a two-steps tunnel-tunnel transition (arrows (1) and (2) even if, ideally, BTBT occurs through a single-step process) mediated by an impurity with energy $E_t$. Lower: ideal scheme of a direct band-to-band tunneling from valence to conduction band through a single-step transition (arrow (3)).

Figure 2.4. Example of the BTBT $I-V$ dark characteristic of an Hg$_{0.7054}$Cd$_{0.2946}$Te $pn$ junction (solid line) at $T = 109$ K evaluated through a 1D simulation with TCAD (Technology Computer Aided Design) Sentaurus exploiting the simple Kane model of equation (2.53). Note the deviation from the standard behavior without the tunneling effect (dashed line) for a certain bias value.
that the behavior of the $I-V$ characteristic should have a trend almost similar to that in equation (2.40) for low $E(V)$, then a linear relation between $I$ and $V$ should be established (see Figure 2.4). This phenomenon is hindered by the temperature increase (and, hence, the $E_g$ increase) since in this case the numerator in the exponential term imposes a higher electric field at the denominator in order to experience a current rise. Thus the higher is the temperature the smaller is the BTBT contribution – i.e. the direct transition of electrons from the valence band on the $p$-side to the conduction band on the $n$-side – since the electric field necessary to counteract the energy gap widening must be stronger.

[Figure 2.5. Experimental $I-E$ dark characteristic of an Hg$_{0.757}$Cd$_{0.243}$Te metal-insulator-semiconductor ($p$-type) photodetector at $T = 55$ K. Note the trend similarity – except for the horizontal axis here shown in terms of electric field – with Figure 2.4, denoting the presence of BTBT effect (from [9]).]

In conclusion we can say that, despite the BTBT – unlike the TAT mechanism – takes place within a wide temperature range, its relevance is greater at low temperatures.

### 2.2.3 Phonon-Assisted Tunneling

Impurities in the semiconductor play an important role by the tunneling point of view. Not only they determine the presence of traps in the gap, but, in the case of shallow or dopant atoms they affect, through the doping concentration, the extension of the depleted region within the junction. In fact, high doping corresponds to a narrow depleted region and, thus, to a higher tunneling probability.

Besides the role of absorbed/emitted phonons in the tunneling mechanism already mentioned in previous paragraphs it should be noticed for completeness that, in indirect gap materials, another mechanism exists involving phonons: the Phonon-Assisted Tunneling, PAT. In this process electrons are scattered by phonons during the tunneling path. This fact may reduce abruptly the current, but not its analytical expression.
2.3 Tunneling Simulations

2.3.1 Brief Introduction to the Physics-based Numerical Simulation of Electronic Devices

We give here a general panoramic of what simulating means, just to introduce the reader to the field of semiconductor device modeling through a very short discussion of the most important transport (or quantum) models used in the scientific literature. Thus, the present introduction is not exhaustive, but it tends to cover the basic concepts needed to understand the subjects of this first chapter, and the framework in which they work. A more detailed insight will be done later.

When we have to build a physical model the first step is to ascribe the constituent particles of the ensemble under study to a particular energy distribution of states, that is the Boltzmann statistics for non-degenerate systems or the Fermi statistics for degenerate ones (when the Fermi energy $E_F$ is in conduction band). Therefore this choice depends on the intrinsic nature of the elements under consideration. Referring to the semiconductor transport theory the simpler starting point of the analysis is represented by the Poisson-Boltzmann equation, that in a generic 1D (non-degenerate) system at thermodynamic equilibrium has the form

$$\frac{\partial^2 [\epsilon \phi(x)]}{\partial x^2} = -q \left[ N_D^+(x) - N_A^-(x) + N_V(x) e^{\frac{q\phi(x) - \chi(x) - E_g(x)}{k_BT}} - N_C(x) e^{\frac{-q\phi(x) - \chi(x)}{k_BT}} \right]$$

(2.44)

with the boundary conditions imposed at $x = 0$ and $L$ (the device edges) and where $\phi$ is the electrostatic potential, $\epsilon$ the dielectric constant, $N_D^+$, $N_A^-$, $N_V$ and $N_C$ are, respectively, the concentrations of donors and acceptors and the effective densities of states at valence and conduction band, $\chi$ is the electron affinity and $q$ the elementary charge. The use of the Boltzmann statistics implies that the distribution function $f(k, r, t)$ of each particle (i.e. carriers) must obey to the so-called Boltzmann Transport Equation – BTE

$$\frac{\partial f}{\partial t} + v(k) \cdot \nabla_r f + \frac{F}{\hbar} \cdot \nabla_k f = \frac{df}{dt}|_{\text{coll}}$$

(2.45)

where $F$ is the force field acting on the particle, such that $F = \hbar(dk/dt)$, the left-hand side expression represents the drift term and $(df/dt)|_{\text{coll}}$ the collision term. Furthermore, in the relaxation time approximation:

$$\left. \frac{df(k, r, t)}{dt} \right|_{\text{coll}} = -\left. \frac{|f(k, r, t) - f_0(k)|}{\tau(k)} \right|_{\text{coll}}$$

(2.46)

where $f_0(k)$ is the distribution function at the thermodynamical equilibrium.

There are several methods to solve the problem: the most accurate (and for this reason, also the most expensive by the computational viewpoint) is the MonteCarlo (MC) approach that consists in bypassing the direct solution of the BTE by simulating the behavior of a given amount of carriers, or all of them if necessary, with the advantage of having a physical description almost complete of the system without any mathematical approximation of the function $f$. Since the description of the whole system in its microscopical entities is too complex in terms of analytical solutions, MC represents the behavior of particles according to a stochastic approach, through a sort of a loop process where the steps are: 1 – computation of the electric field $\mathcal{E}$ and the potential $\phi$ and then generation of a random free flight for each particle, 2 – imposition of a particular scattering process and 3 – computation of the final energy $E$ and wavevector $k$ after the scattering. In this
case the transport regime is considered ballistic. While the routine is computing this loop the macroscopic quantities are extracted and stored.

Moving towards faster and less accurate techniques we meet the Hydrodynamic Model (HD), suitable for systems with many electron-electron scattering events, like in submicrometric and micrometric devices, heterostructures or quantum wells. Its unknowns are the carrier concentrations $n$ and $p$, the energy $E$ and the electrostatic potential $\phi$. Finally we find the family of Drift-Diffusion Models (DD), used in the analysis of micrometric devices (down to 0.5 $\mu m$) in diffusive transport regime, by solving the PDE in (2.44) and the continuity equation for the electron and hole current density $J_{\{n,p\}}$

$$\frac{\partial \{n; p\}}{\partial t} = \pm \frac{1}{q} \frac{\partial J_{\{n,p\}}}{\partial x} - U_{\{n,p\}}$$  \hspace{1cm} (2.47)

(where $U$ is the generation/recombination net rate) over the statistical moments of BTE estimated through the method of moments. The DD model is valid under the assumption according to which all phenomena are slower than the energy relaxation time of the system (under a step variation in the electric field) and its unknowns are $n$, $p$ and $\phi$.

Once the approach has been chosen it is necessary to describe all the parameters related to the device and then discretize PDE equations in the space domain (and also in time, if necessary) in order to deal with ordinary differential equations ODE. Various discretization methods are available and some of the most commonly used are the well-known finite elements (FEM or upwind) and finite boxes (FB) and also the Scharfetter-Gummel method. The latter is an improvement of the first two schemes: let’s consider an electron current density $J_n$ of the form

$$J_n = D_n \frac{\partial n}{\partial x} + v_{\text{drift}}(x,t) n = D_n \frac{\partial n}{\partial x} - \mu_n \frac{\partial \phi(x)}{\partial x} n$$  \hspace{1cm} (2.48)

(and similarly for holes), where $D_n$ is the diffusivity and $\mu_n$ the mobility. Suppose now that $J_n$ is constant and $\phi(x)$ linearly interpolated within the mesh interval $[x_i; x_{i+1}]$, with boundary conditions $n_i = n(x_i)$ and $n_{i+1} = n(x_{i+1})$. Neglecting the GR net rate $U$ and integrating the equation (2.48) in $dx$ we obtain:

$$J_{i+\frac{1}{2}} \simeq -q v_{i+\frac{1}{2}} \left[ \frac{n_i}{1 - e^{-\frac{v_{\text{drift}} \Delta x}{D_n}}} + \frac{n_{i+1}}{1 - e^{-\frac{v_{\text{drift}} \Delta x}{D_n}}} \right]$$  \hspace{1cm} (2.49)

where $\Delta x = (x_{i+1} - x_i)$. Notice that in the limit of $v_{\text{drift}} \Delta x \gg D_n$ we turn back to the upwind scheme.

After the discretization and the imposition of proper boundary conditions, the final solution of the transport problem can be extracted.

It must be said that in addition to the models listed above, that belong to the class of Semi-Classical models, exists also a quantum approach which involves the description of some (few) atoms of the semiconductor lattice through the direct computation of Schrödinger or Green equation. The use of these models, however, is particularly crucial because of the high computational cost and the limited field of application due to the required good knowledge of the physical properties of the material subjected to simulation.

In order to translate the theoretical tunneling current into suitable generation rates we need to make use of the Semi-Classical models. The rates are used in transport models (like in DD) in order to simplify the treatment with respect to the full quantum approach, even if they tend to include as much effects as possible. They are reconverted (for example, with the continuity equation (2.47)) in a more useful expression of the current (i.e. the
model) that can be exploited, for instance, to build the I-V characteristics of the device. We will enter in details of this computational mechanisms in one of the next chapters while in this section we will refer to the state of the art in the simulation of tunneling meant as generation/recombination processes by chronologically browsing the most important models used in this field of research.

### 2.3.2 Kane Model

In the early 60’ of the past century E.O.Kane developed one of the first theory that modeled the tunneling mechanism both in direct and indirect semiconductors [29]. For the first case and starting from the Hamiltonian written in the Bloch basis under the influence of a constant electric field, Kane described the electrons in a very similar way to that seen previously regarding the WKB approximation, yielding a tunneling probability $P$ equal to the equation (2.32) except for a multiplicative factor $\frac{\pi}{9}$. Solving the Schrödinger’s equation in $k$ and since it is known from the Landauer formula that

$$G_{btbt} = |\mathcal{E}| \frac{\Delta J_{tunn}}{\Delta E} = |\mathcal{E}| \frac{\int \sum_{k} P(E, k_{\perp})(f_{v} - f_{c}) dE}{\Delta E}$$

(2.50)

(with $J_{tunn}$ the tunneling current density over the area $A$ and where $f_{v}$ and $f_{c}$ are respectively the Fermi functions $F(E)$ and $F'(E)$ of equation (2.33) in valence and conduction band) the resultant net rate of generation for direct (without phonon assistance) BTBT is [30]:

$$G_{btbt}^{Kane} = \frac{q^{2} \sqrt{m^{*}}}{18\pi \hbar^{2} \sqrt{E_{g}}} \mathcal{E}^{2} \exp \left[-\frac{\pi \sqrt{m^{*}E_{g}^{3}}}{2q\mathcal{E}\hbar} \right].$$

(2.51)

In the Kane’s paper the dispersion relation $E(k)$ (where $k \equiv |k|$), connecting the conduction with the valence band states, is calculated as:

$$E^{\pm}(k) = \frac{E_{g}}{2} + \frac{\hbar^{2}k^{2}}{2m_{0}} \pm \sqrt{E_{g}^{2} + \frac{E_{g}\hbar^{2}k^{2}}{2m_{c}m_{V}}}.$$ 

(2.52)

Now note that the equation (2.51) can be rewritten in the more general form

$$G_{btbt} = A \mathcal{E}^{2} \exp \left[-\frac{B}{\mathcal{E}} \right]$$

(2.53)

where the constant $A$ is

$$A^{Kane} = \frac{q^{2} \sqrt{m^{*}}}{18\pi \hbar^{2} \sqrt{E_{g}}}$$

(2.54)

and B (it has the dimensions of an electric field):

$$B^{Kane} = \frac{\pi \sqrt{m^{*}E_{g}^{3}}}{2q\hbar}$$

(2.55)

It is worth stressing that the equation (2.51) holds for a 3D carrier gas (i.e. without any quantum confinement) under the Fermi-Dirac statistics and in the approximation of a constant electric field $\mathcal{E}$ (Local model). This means that electrons and holes generations profiles $G_{\{n,p\}}(x)$ along the dimension $x$ of the junction are the same, yielding a possible overestimation of $G_{btbt}$. 
2.3.3 Hurkx Model

The Kane model represents a standard reference, since it was one of the first to be formulated. The two following models – the Hurkx and the Schenk one – are probably the most used both in computational simulators and in theoretical studies. In order to better understand them we must now take in consideration the classical formalism of SRH generation/recombination. In particular the net GR rate for an electron-hole couple is given by [26, 31]:

\[ U_{SRH} = \frac{np - n_{i,eff}^2}{\tau_p (n + n_1) + \tau_n (p + p_1)} \]  \hspace{1cm} (2.56)

The values of \( n_1 \) and \( p_1 \) depend exponentially by the energy trap \( E_t \) (if any) and lifetimes \( \tau_{p,n} \) are modeled as a product of a doping-, field- and temperature-dependent term:

\[ \{n; p\}_1 = n_{i,eff} \exp \left[ \pm \frac{E_t}{k_B T} \right] \]  \hspace{1cm} (2.57)

and

\[ \tau_{dop} = \tau_{min} + \frac{\tau_{max} - \tau_{min}}{1 + \left( \frac{N_A + N_D}{N_{ref}} \right)^\gamma} \]  \hspace{1cm} (2.58)

\[ t(E) = \frac{1}{1 + g(n,p)(E)} \]  \hspace{1cm} (2.59)

\[ t(T) = \left( \frac{T}{300 \text{ K}} \right)^\alpha \]  \hspace{1cm} (2.60)

where \( \gamma \) and \( \alpha \) are user-defined parameter. In TCAD (Technology Computer Aided Design) simulation tools the so-called Hurkx trap-assisted tunneling is implemented simply exploiting a “tunneling factor” \( g(E) \) of the form ([31]–[34]):

\[ g(E) = \int_0^E \exp \left[ \frac{-u^2}{3 E(E)} \right] du \]  \hspace{1cm} (2.61)

with the integral approximated solution

\[ g(E) = \begin{cases} \sqrt{\pi} E \exp \left[ \frac{E^2}{3} \right] \left( 2 - \text{erf} \left[ \frac{1}{2} \left( \frac{E_n}{E} - 1 \right) \right] \right) & E \leq E_n \\ \sqrt{\pi} \sqrt{E_n} \exp \left[ -E_n + \sqrt{E_n} \sqrt{\frac{E}{E_n}} \right] \text{erf} \left[ \sqrt{\frac{E}{E_n}} - \sqrt{\frac{E_n}{E}} \right] & E > E_n \end{cases} \]  \hspace{1cm} (2.62)

where \( E = E(E) = (E q \hbar) / \left( \sqrt{8 m_0 m_{tunn}(k_B T)^3} \right) \) and

\[ E_n = \begin{cases} \frac{0.5 E_g}{k_B T - \ln \left[ \frac{n}{n_i} \right]} & E_t \leq k_B T \ln \left[ \frac{n}{n_i} \right] \leq 0.5 E_g \\ \frac{0.5 E_g - E_t}{k_B T} & E_t > k_B T \ln \left[ \frac{n}{n_i} \right] \\ 0 & \text{otherwise} \end{cases} \]  \hspace{1cm} (2.63)
So, in order to obtain the net generation/recombination rate due to TAT in the Hurkx formalism, we should include the conventional SRH effect and substitute the carrier lifetimes \( \tau_{\text{np}} = \tau_{\text{dop}} t(E) t(T) \) and the \((2.57)\) in the expression \((2.56)\):

\[
U_{\text{H}urkx}^{\text{tat}} = \frac{np - n_{i,\text{eff}}^2}{n + n_{i,\text{eff}} e^{E_n/T}} + \frac{p + n_{i,\text{eff}} e^{E_p/T}}{p + n_{i,\text{eff}} e^{E_p/T}}
\]

(2.64)

Also the Hurkx Band-to-Band tunneling is pretty common in numerical solvers. It is usually implemented through a generation rate of the form \([31, 34]\):

\[
G_{\text{H}urkx}^{\text{bdtt}} = A \cdot E^\beta np - n_{i,\text{eff}}^2 (n + n_{i,\text{eff}}) (p + n_{i,\text{eff}}) \exp \left[ -\frac{B}{E} \sqrt{\frac{E_n^3(T)}{E_p^3(300 \, \text{K})}} \right]
\]

(2.65)

where \( A \) and \( B \) (and also \( \beta \)) are the typical parameters yet discussed with respect to the Kane model (in that case it was \( \beta = 2 \)) and which depend on the material. We will return in the following on these features, but for now let’s notice that in \([33]\) Hurkx highlights that for direct gap semiconductors, like the HgCdTe, it must be \( \beta = 2 \), where a triangular potential barrier yields a better fit of \( I-V \) characteristic with respect to the parabolic one (see \([34]\)). Thus, the case \( \beta = 2 \) seems to be valid for an abrupt junction with the above barrier assumptions or in a uniform field junction approximation (see \([31\text{, p.388}]\) and \([35]\)). For indirect tunneling, otherwise, \( \beta = 2.5 \). The latter case can be also applied in linearly graded junctions (see \([32]\)). If the shape of the potential barrier changes because of a different doping profile probably \( \beta \) should adapt its value, but how this should be done is still a topic of study. In fact most recent papers refer to the parameter \( \beta \) only in the twofold way just explained \((35\text{–}37)\).

For a triangular potential barrier and dealing our calculations with a unique lattice direction (e.g. \( x \)) we can finally obtain the \( I-V \) characteristic by reversing equation \((2.50)\) or remembering that:

\[
I_{\text{tunn}} = q \int_a^b G(E(x)) \left( f_v - f_c \right) dx
\]

(2.66)

(\( x = a \) and \( x = b \) are the two turning points, i.e. the band edges along the tunneling path) that is similar to the Landauer formula used in \((2.50)\) except for the integration, that here is performed in the space and no more in the energy domain.

### 2.3.4 Schenk Model

The third model \([38]\) that we examine is perhaps one of the most complete by the physical standpoint. In his 1993 study A.Schenk built a very rigorous theory – as declared in the title of the paper – concerning the trap-assisted tunneling and the phonon-assisted-band-to-band tunneling for silicon, that includes the scattering between electrons and phonons and calculated using the equilibrium Green’s function formalism (now we know that it represents a kind of quantum model) for the six conduction band valleys. The overall Hamiltonian of the system is written as a sum of an unperturbed term plus the electron-phonon and hole-phonon contributions, then the current density and generation rate are calculated.

The complete Schenk formalism deals with the second quantization treatment of the electrons, described as Bloch states, and contains four unknown matrix elements. For this reason Schenk provided a more flexible formulation of his model in order to make it
more implementable into device simulation programs. This procedure passed through the conversion of the Airy function present in the expression of the tunneling rate with their asymptotic approximations (recently corrected in [39]). This yields the compact versions of the Schenk trap-assisted tunneling net rate, expressed in the same form of the Hurkx one in (2.64), but with [31]:

\[ g_n(\mathcal{E}) = \sqrt{1 + \frac{(h\theta)^3 (E_t - E_0)}{E_0 \hbar \omega_0}} \sqrt{\frac{(h\theta)^3 (E_t - E_0)}{2E_tE_0}} \left[ -\frac{E_t}{\hbar \omega_0} \ln \frac{E_t}{E_0} - \frac{1}{\epsilon_R} \ln \frac{E_t}{E_0} + \frac{E_t - E_0}{kT} \ln \frac{E_t}{E_0} \right] e^{\Omega} \]  

(2.67)

(similarly for holes) with

\[ \Omega = \left[ \frac{(E_t - E_0)}{\hbar \omega_0} + \frac{h\omega_0 - kT}{2h\omega_0} + \frac{2E_t + kT}{2h\omega_0} \ln \frac{E_t}{\epsilon_R} - \frac{E_0}{\hbar \omega_0} \ln \frac{E_0}{\epsilon_R} \right] + \frac{(E_t - E_0)}{kT} - \frac{4}{3} \left( \frac{(E_t - E_0)}{\hbar \omega_0} \right)^2 \]  

(2.68)

and where \( E_0 \) denotes the energy for an optimum horizontal transition path, \( \epsilon_R \) is the lattice relaxation energy and \( \Theta = (q^2 \epsilon^2 / 2m_tunn) \) the electro-optical frequency. With a so determined tunneling factor, Schenk defines a new expression for the carrier lifetimes as follows:

\[ \tau_{\{n;p\}} = \frac{T}{300 \text{ K}}^{3/2} \frac{1 + g_{\{n;p\}}(\mathcal{E})}{\tau_{dop}} \]  

(2.69)

in which appears the same \( \tau_{dop} \) of equation (2.58). Also the carrier densities are modified with respect to the (2.57), and for the electrons we have:

\[ n_{\text{schenk}} = n \cdot e^{\frac{n - n_{\text{ref}} \left| \nabla E_{Fn} \right| (E_t - E_0)}{kT}} \]  

(2.70)

and similarly for holes (the value of \( n_{\text{ref}} \) can be found in literature or in TCAD software manuals, as in [31]).

Finally, the usual generation rate for the Schenk band-to-band tunneling is ([31], [38, pag.29]):

\[ G_{\text{schenk}} = A \cdot \mathcal{E}^\frac{7}{2} \frac{np - n_{\text{i,eff}}^2}{(n + n_{\text{i,eff}})(p + n_{\text{i,eff}})} \left[ (\mathcal{E}_c^+) - \frac{3}{2} e^{\left( \frac{-E_{\text{cr}}}{h \omega_0} \right)} \right] + (\mathcal{E}_c^-) - \frac{3}{2} e^{\left( \frac{-E_{\text{cr}}}{h \omega_0} \right)} 

\]  

(2.71)

where \( \hbar \omega_0 \) represents the TA phonon energy and \( \mathcal{E}_c^\pm = B \sqrt{E_{g,eff} \pm \hbar \omega_0} \) is a critical field strength (the upper and the lower signs stand for, respectively, the reverse and the forward bias condition) that depends on the material (like the parameters \( A \) and \( B \)). For this model the modified electron density is:

\[ n_{\text{btt}} = \frac{n}{n_{\text{i,eff}}} \left( \frac{N_C}{n} \right) e^{\frac{n_{\text{i,eff}} \left| \nabla E_{Fn} \right|}{kT}} \]  

(2.72)

with a similar relation for \( p_{\text{btt}} \).
2.3.5 Other models

In the case of a strongly non uniform electric field within the depleted region, like in the case of heterostructures, several authors have taken into account a supplementary formalism, also included in TCAD simulators: the *Non-local model* of tunneling. Such assumption on the electric field needs a more realistic computation of the tunneling path since the latter is strongly dependent on the tunneling path that, in turns, depends on the bands shape. Through an integration of the rate along the $x$ direction (for a 1D modeling) due to the inclusion of an electron or hole capture/emission from the defect level at any location between the starting position and the ending position, we are able to compute the new tunneling path no more confined only along the shortest distance between the two turning points (i.e., which minimizes its energy). This facts puts the basis for the investigation on the tunneling mechanism in several crystallographic directions and so on the study of the full-band description of the material under consideration.

2.4 Overview on the Tunneling Simulating Tools in Synopsys® TCAD Sentaurus Device

Once the general theory of tunneling has been exposed, now we will give for completeness a short essay about the capabilities of one of the most powerful and exploited commercial softwares, TCAD Sentaurus Device, produced by Synopsys. All the SRH-like processes are implemented by using the standard GR net rate already mentioned in equation (2.56) because the Boltzmann statistics is employed by default. Otherwise, if the particular device conditions need a more realistic description (for instance, at very high carrier densities and up to the limit of degenerate systems) we have to switch to the *Fermi-Dirac statistics*. So, in place of equation (2.57) or, instead of the classic formulation (here outside the equilibrium)

$$\begin{align*}
 n^{\text{Boltz}} &= N_C \cdot e^{\frac{E_{F_n} - E_C}{k_B T}} \\
p^{\text{Boltz}} &= N_V \cdot e^{\frac{E_{F_p} - E_p}{k_B T}}
\end{align*}$$

\hspace{1cm}(2.73)

$(N_C$ and $N_V$ are, again, the effective densities of states, $E_{F_n}$ and $E_{F_p}$ the quasi Fermi energies for electrons and holes), we must use these:

$$\begin{align*}
 n^{\text{Fermi}} &= N_C \mathcal{F}_{1/2} \left( \frac{E_{F_n} - E_C}{k_B T} \right) \\
p^{\text{Fermi}} &= N_V \mathcal{F}_{1/2} \left( \frac{E_{F_p} - E_p}{k_B T} \right)
\end{align*}$$

\hspace{1cm}(2.74)

in which $\mathcal{F}_{1/2}(\eta)$ represents the *Fermi integral* of the $(1/2)$-th order:

$$\mathcal{F}_{1/2}(\eta) \triangleq \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{u}}{1 + e^{\eta u}} du,$$

\hspace{1cm}(2.75)

where $u$ is the the carrier energy. In this framework Sentaurus Device allows a new edition for the SRH rate, that is

$$U^{\text{Fermi}}_{\text{SRH}} = \frac{np - \gamma_n \gamma_p n_{i,\text{eff}}^2}{\tau_p (n + \gamma_n n_1) + \tau_n (p + \gamma_p p_1)}$$

\hspace{1cm}(2.76)
where
\[
\gamma_n = \frac{n}{N_C} e^{\frac{E_{F_n} - E_C}{k_B T}}
\]  \hspace{1cm} (2.77)
and
\[
\gamma_p = \frac{p}{N_V} e^{\frac{E_V - E_F_p}{k_B T}}.
\]  \hspace{1cm} (2.78)

Taking inspiration by the non-local version of all models described before also Sentaurus built its own corresponding models, labeled with the prefix Dynamic Non-local Path- (the definition of “non-local” is commonly referred to the non-uniformity of the electric field \(E\) in the semiconductor that yields the tunneling path strongly dependent on the bands shape). For the TAT case the software accounts for the exact barrier profile to compute the position-dependent recombination rate, and then searches for the tunneling path with the following criteria [31]:

- The tunneling path is a straight line with its direction equal to the gradient of the conduction band at the starting position;
- The tunneling energy is equal to the conduction band energy at the starting position and is equal to the defect level at the ending position;
- Electrons can be captured in or emitted from the defect level at any location between the starting position and the ending position of the tunneling path;
- When the tunneling path encounters Neumann boundary conditions or semiconductor-insulator interfaces, it undergoes specular reflection.

Thus, if normally the tunneling process is seen as a carrier appearance/disappearance event in the bands and not as a real particles exchange process, in this particular case and for the electrons the recombination rate will depend on an integration on the path along \(x\):

\[
R_{\text{tat}}^{\text{DNP}} = C_n \int_{0}^{1} \frac{\Gamma(x, E_C(0)) \left( T(0) + T(x) \right)}{2 \tau_n(x) \sqrt{T(0) T(x)}} \times
\left[ e^{\frac{E_{F_n}(x) - E_C(x)}{k_B T(x)}} f_p(x) - e^{\frac{E_F(x) - E_C(x)}{k_B T(x)}} f_n(x) \right] dx
\]  \hspace{1cm} (2.79)

(and similarly for holes) where the tunneling path is normalized such that \(0 \leq x \leq 1\), \(T(x)\) is the lattice temperature, \(f_{(n;p)}(x)\) are the electron/hole occupation probabilities at the defect level and

\[
C_n = |\nabla E_C(x)|_{x=0} \frac{N_C(0)}{k_B T(0)} \left[ \frac{1}{e^{\frac{E_{F_n}(x) - E_C(x)}{k_B T(x)}} + 1} \right]
\]  \hspace{1cm} (2.80)

(the term in the square brackets is unitary for the Boltzmann statistics),

\[
\Gamma(x, E) = W_C(x, E) e^{-2 \int_{0}^{x} k(x, E) dx},
\]  \hspace{1cm} (2.81)

\[
W_C(x, E) = \begin{cases} 
1 & \text{Hurkx TAT} \\
\frac{\sqrt{\hbar(E_C(x) - E_C(0)) W(E - E_t(x))}}{4 \pi m_C k_B T(x)} & \text{Schenk TAT}
\end{cases}
\]  \hspace{1cm} (2.82)
where

\[ W(E) = \frac{1}{\sqrt{\chi \cdot e^{E + \frac{S}{h\omega_0} + \chi}}} \left( \frac{S}{\frac{h\omega_0}{2k_B T}} \right)^{E \over h\omega_0}, \]  
\[ \chi = \sqrt{\frac{E^2}{(h\omega_0)^2} + \frac{S^2}{\sinh^2 \frac{h\omega_0}{2k_B T}}}, \]  

\[ S \text{ is the Huang-Rhys coefficient and } h\omega_0 \text{ the phonon energy.} \]

In a similar way we find in Sentaurus the Dynamic Non-local Path implementation for both Hurkx- and Schenk-BTBT models, recommended when the band bending caused by high electric fields – or in heterostructures with abrupt or graded junctions – is consistent. In such condition the tunneling path is determined dynamically on the energy band profile through these criteria:

- The tunneling path is a straight line with its direction opposite to the gradient of the valence band at the starting position;
- The tunneling energy is equal to the valence band energy at the starting position and is equal to the conduction band energy plus band offset at the ending position;
- When the tunneling path encounters Neumann boundary conditions or semiconductor-insulator interfaces, it undergoes specular reflection.

For direct semiconductors we can redefine the hole recombination rate (again in the interval \( 0 \leq x \leq 1 \)) as follows:

\[ R_{\text{bdt}}^{\text{DNP}} = |\nabla E_V(x)|_{x=0} C_p e^{-\frac{2}{\epsilon} \int_0^1 k(x,E) dx} \frac{1}{e^{-\frac{E-E_{F_n}(0)}{k_B T}} + 1} - \frac{1}{e^{-\frac{E-E_{F_p}(1)}{k_B T}} + 1}, \]

where

\[ C_p = \frac{\pi g}{36 h} \int_0^{1/\epsilon(x,E)} \frac{1}{x} dx \]  

and with \( p_{\text{max}} \) the maximum transverse momentum and \( g \) the degeneracy factor. It must be noted that the equations (2.79) and (2.85) are in the form (2.50), where the square brackets correspond to the term \( (f_v - f_c) \) in the Fermi-Dirac framework.

For a more complete modeling of the process the software provides also the possibility to account for some other collateral phenomena. One of the most important is the temperature dependence of the physical quantities involved, like the effective DOS

\[ N_C(T) = N_C(300 \text{ K}) \cdot \left( \frac{T}{300 \text{ K}} \right)^{3 \over 2} \]  

or carrier heating terms (for instance, in the Peltier effect). Also the trend of the bands \( E(x) \) within the device or the energetic dispersion \( E(k) \) can be manipulated, determining a variation of the carrier wavevectors and its effective masses in the two bands. A separate discussion must be made about the determination of lifetimes \( \tau_n \) and \( \tau_p \): thanks to the intrinsic quantum origin of this parameter and its crucial role in determining with a certain accuracy the GR rate in the tunneling process, the relation (2.58) could be too
much generic and also inadequate with respect to the particular system examined: e.g. over a typical electric field strength ($\sim 3 \cdot 10^5$ V/cm) we see a reduction in the values of $\tau$. For such reasons we will return in the detail of this problem in the following, with the study of a specific case.

Another interesting feature is the possibility to implement in Sentaurus every kind of custom recombination models through the so-called Physical Model Interface – PMI. This environment needs a C++ code for the definition of the new GR rate, which then will be inserted into the Drift Diffusion Transport Model during the simulation.

Finally, a mention about the definition of the SRH centers. In this regard Sentaurus offers the possibility to design the energetic and spatial distribution of traps, with several types, capture and emission rates (thus, cross-sections) and locations (bulk or interfaces). Each trap can be modeled singularly or globally with the others and these opportunities makes the simulating environment as various as the system analyzed needs, meliorating the realism and the effectiveness of the simulation itself.

2.5 Actual HgCdTe-based Devices Modeling

Tunneling models implemented in Sentaurus are without doubt the building blocks of a rigorous approach to the HgCdTe-based photodetectors simulation but, thanks to the advent of a third generation of such devices, each sensor must be considered as a specific case: indeed geometrical, growth and engineering features play an important role in developing an effective model capable to predict the behavior of the material. It should be noticed that recently one of the most studied configurations, already described in the previous chapter, is the double $p^+ - n- n^+$ homojunction, because of – as already seen – its remarkable noise suppression mechanism and the capability to operate at room temperature (like in HOT devices). From the complexity of this structure and of the related synthesis process, the need to abandon the conventional way of simulating arises. First of all, in order to have a better agreement between computation and measurements of real devices, it is important to build a 3D modeling. Secondly, just these measurements can improve the model accuracy, by extracting from experiments some important physical features and, finally, this features must be introduced in the computation as corrective parameters that determine the exact work point of the simulated device. So only through cooperation between theoretical approach and physical data we can improve the reliability of the model. One example for all: being the TAT process related to the nature, concentration and distribution of lattice dislocation, most powerful algorithms must take into account several parameters describing the grow process of the structure in order to match properly data and calculations of a real device.

Thanks to the improvement of growing techniques of semiconductor materials and of designing devices many recent models have not to consider some physical effects (like internal radiative recombination, band-to-band-tunneling [40]), because they have no such great effect on the behavior of the structure. Neglected this kind of technologically-suppressed phenomena today models generally account only for a small series of effects (that remain the most relevant): Auger recombination, SRH processes, trap-assisted-tunneling and in some cases also surface recombination [41]. A very remarkable result is obtained by Srivastav et al. [40, 42] that successful simulated an array of HOT diodes by using a set of just four physical parameters. As depicted in graph of Figure 2.6 the new model proposed well explains the behavior of some devices based on $p^+ - n$ $n^+$ homojunction. For this purpose Srivastav and his co-authors have implemented an ad-hoc 1D numerical method that includes generation-recombination processes like Auger, SRH and TAT mechanisms in the
2.5 Actual HgCdTe-based Devices Modeling

Figure 2.6. Schematic view of a HOT device structure, with $p^+ - n - p^+$ homojunction (left) obtained by boron implantation in the $p$-zone, and comparison of experimental and theoretical data for such device (right) as reported in [40, 42].

Figure 2.7. Measured dark current as a function of predicted dark current for four representative 150 K MWIR diodes. Maximum deviation is $< 10\%$ (from [45]).

The right hand side of carrier continuity equations. Then, the characteristic of the device was tested through Hall measurements and a set of parameters was extracted from $R_0 A$-$V$ curves: the energy of traps $E_t$, their concentration $N_t$, the carrier lifetime $\tau$ and the series resistance of the device $R_s$. The rate of the dominant physical processes was introduced by equations

\[
U_{SRH} = \frac{p(x)n(x) - n_t^2}{[\tau_p n(x) + \tau_n p(x)] n_e E_t} \quad (2.88)
\]

\[
U_A = (p(x)n(x) - n_t^2)(p(x)G_A7 + n(x)G_{A1}) \quad (2.89)
\]

\[
U_{tot} = \frac{\pi^2 q^2 m_e^* N_t M^2 \mathcal{E}}{h^3 (E_g - E_t)} \exp \left[ -\frac{4\sqrt{2m_e^*(E_g - E_t)^{3/2}}}{3qh\mathcal{E}} \right] \quad (2.90)
\]

where $M^2$ is the transition matrix and the field $\mathcal{E} = \frac{f_c}{W}(V_{bi} - V_a)$ accounts for a correction factor $f_c = 1.62$ computed by fitting the $I$-$V$ characteristic. The results of simulations are shown in Figure 2.7.
Many other works recently published deal with the same guideline: mixing theoretical methods and physical data [41, 43, 44, 45], even if each time equations change according to the specific device which they have been written for. In [43], for example, Li et al. considered in their HgCdTe photodiode simulations also the contribution of metal-semiconductor contact, deducing the Schottky barrier of electrodes from fitting curves. This leads to build a realistic model, more adherent to the true operating condition and less theoretical than what researchers made in the past years.

Kocer et al. [44], instead, created a MATLAB iterative finite-difference algorithm to solve one dimensional Poisson and continuity equations for a \( pn \) homojunction HgCdTe photodiode, developing a new \textit{ad-hoc} numerical solver, in which the physical rates did not differ so much from those reported in equations (2.88)–(2.90). They predicted the device behavior by changing values of two parameters, especially: \( E_t \) and \( N_t \). Besides the fact that these parameters seem to be the most influent in such kind of simulations (due to the physics of recombination and tunneling mechanisms) it is possible to increase the number of them: Tennant simulated his double planar heterostructure [45] by using a set of ten fit/input parameters. This fact gives the idea of complexity in today device simulation.
References


[31] Sentaurus Device User Guide H-2013.03, Synopsys


Once the main phenomena occurring in narrow gap semiconductor photovoltaic detectors have been exposed we pass now to analyze how they can be translated into suitable physics-based modeling formulations. Remembering that the device simulation isn’t only a strategy to predict the electrical behavior of the structures but that it also represents the modern workbench for scientists who want to investigate on the advanced device performance, it can be stated that modeling is a complex issue both in mathematical and physical terms. Hence, first of all we have to stress that, as we will see later, being the electric transport problem represented by a set of PDE (partial differential equations) the complete solution of such framework is often impossible without approximations (except in particular conditions). By the physical standpoint, again, a complete description of the whole system should be in general difficult to obtain, since it requires the knowledge of all the physical quantities involved at each point and at each instant. That’s why the mathematical models used in the simulations replace the complete knowledge of the structure with a dynamic description of the averaged variables. Not only: in order to implement such a simulating tool into TCAD software the numerical solution of the PDE system must be evaluated within a discretized representation of the model itself, through one of the several techniques available (remember the already mentioned FEM, FB or “Scharfetter-Gummel” methods). The spatial approximation of the PDE by means of a discretization scheme yields, in turn, a discrete system of ODE (ordinary differential equations) that can be linearized, if necessary, under specific physical simplifications.

Besides these principles, governing almost all semi-classical models, the difference between one method and the others arises from the particular choice of the quantities described. A separate category, then, is represented by the quantum models and the MonteCarlo approach (see Sect. 2.3.1), since they exploit two different strategies in describing the evolution of each particle composing the system.

The basis of each physics-based device model lies in the statistical description of its constituents, that by the electronics standpoint are the charge carriers: electrons and holes. Making the assumption according to which the holes follow the same statistics of electrons, we make use of the Fermi-Dirac distribution $f(E)$ for a carrier population at
temperature $T$ and energy $E$

$$f_{FD}(E) = \frac{1}{e^{\frac{E-E_F}{k_BT}} + 1}. \quad (3.1)$$

The (3.1) is justified by the fact that electrons are fermions (spin 1/2 particles) obeying to the Pauli exclusion principle, and at $T = 0$ K it holds

$$f_{FD}(E) = \begin{cases} 
1 & \text{for } E < E_F \\
0 & \text{for } E > E_F 
\end{cases}, \quad (3.2)$$

from which we can compute the carrier energy densities as

$$\begin{cases} 
\frac{dn}{dE} = \rho_C(E) f_{FD}(E) \, dE \\
\frac{dp}{dE} = \rho_V(E) (1 - f_{FD}(E)) \, dE 
\end{cases}. \quad (3.3)$$

where the density of states are

$$\begin{cases} 
\rho_C(E) = \frac{dN_C(E)}{dE} = \frac{4\pi \sqrt{8m^*} e}{h} \sqrt{E - E_C} \\
\rho_V(E) = \frac{dN_V(E)}{dE} = \frac{4\pi \sqrt{8m^*} h}{h} \sqrt{E_V - E} 
\end{cases}. \quad (3.4)$$

If the temperature increases the distribution function has no more a step-like behavior at $E = E_F$ but assumes a smooth trend which affects a region of energy amplitude $\sim k_B T$. The meaning of $E_F$ in a generic metal at $T = 0$ K is obvious: it represents the maximum level of all the filled states or, in other words, in a metal at zero K the electrons occupy all the possible states up to the Fermi level. In intrinsic semiconductors at $T = 0$ K (let’s suppose the non-degenerate case), since $E_F$ lies within the forbidden region, it represents instead the energy corresponding to a valence band completely filled and to an empty conduction band. Furthermore one can note that at a generic temperature it holds $f(E_F) = 1/2$, resulting in an occupation probability equal to 1/2 for an electron whose energy is just $E_F$ so, as long as the Fermi level remains within the energy gap and, in turn, the energy gap is sufficiently greater than $k_B T$, then the semiconductor is still in equilibrium conditions. This results in a more rigorous explanation of generation processes raising in narrow gap semiconductors for LWIR and VLWIR detection, like Hg$_{1-x}$Cd$_x$Te, where the thermal energy could be greater or comparable to the energy gap: in this case a finite probability for the electrons to occupy empty states in conduction band occurs, since $f(E)$ non-negligibly overlaps the band edges, yielding carrier exchange phenomena between the two main bands and, thus, a newcomer $G_{th}$ term.

The Fermi-Dirac distribution function (3.1) is always valid, also for degenerate semiconductors, but for the non-degenerate ones we usually exploit the Boltzmann approximation

$$f_B(E) = e^{-\frac{E-E_F}{k_BT}}. \quad (3.5)$$

so we describe the system as composed by a population of generic particles which do not obey to the exclusion principle. Comparing (3.1) and (3.5) one can note that $f_{FD}(E) \simeq f_B(E)$ as long as $\exp\left[\frac{E-E_F}{k_BT}\right] \gg 1$, i.e. for $E \gg k_BT$. At this limit the classical Boltzmann treatment in place of the quantum Fermi-Dirac one is well accepted. Through the
integration of electron and hole energy with respect to the corresponding band we obtain

\[
\begin{align*}
\begin{cases}
n = N_C e^{\frac{E_F - E_C}{k_B T}} , \\
p = N_V e^{\frac{E_V - E_F}{k_B T}} .
\end{cases}
\end{align*}
\]

(3.6)

Moreover, defining the intrinsic concentration as

\[
n_i = n p = \sqrt{N_C N_V e^{-E_g / k_B T}},
\]

(3.7)

an alternative expression to the (3.6) can be

\[
\begin{align*}
\begin{cases}
n = n_i e^{\frac{E_F - E_{Fi}}{k_B T}} , \\
p = n_i e^{\frac{E_{Fi} - E_F}{k_B T}} ,
\end{cases}
\end{align*}
\]

(3.8)

where

\[
E_{Fi} = \frac{1}{2} \left( (E_C + E_V) + k_B T \log \left[ \frac{N_V}{N_C} \right] \right)
\]

(3.9)

is the intrinsic Fermi level.

In general, the distribution function must obey to the already mentioned (see equation (2.45)) Boltzmann Transport Equation (BTE)

\[\frac{\partial f}{\partial t} + v(k) \cdot \nabla_r f + \frac{F}{\hbar} \cdot \nabla_k f = \frac{df}{dt} \bigg|_{\text{coll}}\]

(3.10)

whose complete derivation can be found, for instance, in Ghione’s book [46]. Now we define the generalized statistical moment as the integral of the BTE in the \( k \) -space multiplied by \( k \) itself, i.e.

\[
\frac{\partial}{\partial t} \int_k \psi f \, dk + \nabla_r \cdot \int_k v(k) \psi f \, dk - \frac{F}{\hbar} \cdot \nabla_k \int_k \psi f \, dk = \int_k \frac{d(\psi f)}{dt} \bigg|_{\text{coll}} \, dk.
\]

(3.11)

Remembering that \( f(E) \equiv f(k, r, t) \) is a function defined in the phase space and assuming a \( k \)-power trend for \( \psi(k) \) such that holds the following development

\[
\psi(k) = a_0 + a_1 k + a_2 k^2 + \ldots
\]

(3.12)

(the \( a_j \) are constants) we can express the (3.11) as a succession of functions like

\[
M_j = \int_k \psi_j(k) f \, dk
\]

(3.13)

where the \( M_j \) represents the so-called \( j \)-th order moment of \( f(k, r, t) \) characterized by the property \( M_j = j \langle \psi_j(k) \rangle \), and with

\[
\begin{align*}
\begin{cases}
M_0 = 1 \\
M_1 = \hbar k \\
M_2 = \frac{\hbar^2 k^2}{2 m_e^*} \\
\vdots
\end{cases}
\end{align*}
\]

(3.14)
The first three moments are the most relevant in this kind of problems since they are related, respectively, to the carrier concentration, the carrier average momentum and to the carrier average energy. This means that using \( M_0, M_1 \) or \( M_2 \) in the BTE corresponds to the definition of a *carrier transport equation*, a *momentum transport equation* and finally of an *energy transport equation*. In particular, exploiting the zeroth moment

\[
M_0 = \int f(k, r, t) \, dk = n
\]

one can obtain the charge conservation law for the electrons

\[
\frac{\partial n}{\partial t} + \nabla \cdot (n \mathbf{v}_e) = \frac{dn}{dt}_{\text{coll}}.
\]

The system composed by the first three moments is not always selfconsistent, so often the heat flux \( \Phi_Q \) of carriers is involved thanks to the fourth moment. Otherwise we can approximate \( \Phi_Q \), assuming it proportional to the carrier temperature gradient [46]

\[
\Phi_Q \approx -\kappa \nabla_T T
\]

where the thermal conductivity in the case of an electron gas is

\[
\kappa \approx \left( \frac{k_B^2 n \mu_e q}{q} \right).
\]

Merging the three moment equations with the Poisson equation we obtain the hydrodynamic model (HD). But the need of reducing the complexity of this approach leads to a series of approximations. Neglecting convective terms [47], assuming a slow average energy dynamics with respect to the average momentum dynamics and, finally, through the relaxation time approximation for the collision term (see equations (1.39) and (2.46)), arises the so-called *energy balance model*. In this model the carrier densities can be separated into a drift and a diffusion term, with its corresponding energy-dependent mobility and diffusivity (with also the lattice temperature).

Further approximations conduce to the basic – and also most commonly used – statistical model, i.e. the drift-diffusion, in which the carrier mobilities and diffusivities are related to the local value of the electric field and no more to its time and space variations. Since the DD model is the main tool in commercial TCAD softwares like Sentaurus Device, the next section will be devoted to the explanation of the model in order to complete the description of device behavior within the modeling framework, before the analysis of a simulating study.

### 3.1 Details about the Drift-Diffusion Transport Model

The drift-diffusion model is composed by an equation describing the interaction between carriers and electric field, i.e. the Poisson equation

\[
\nabla (-\mathcal{E}) = \nabla^2 \phi(r, t) = -\frac{\rho}{\epsilon}
\]

and a set of continuity equations for electrons and holes

\[
\frac{\partial \rho}{\partial t} + \nabla J_{n(p)} - U_{n(p)} = 0
\]
in which the current densities are treated as a sum of a drift and a diffusion component

\[
\begin{align*}
J_n &= -q n(r, t) \mu_n \nabla_r \phi(r, t) + q D_n \nabla_r n(r, t) \\
J_p &= -q p(r, t) \mu_p \nabla_r \phi(r, t) - q D_p \nabla_r p(r, t)
\end{align*}
\] (3.21)

The (3.20) states that each carrier is characterized by a conservation law where the outgoing charge (expressed by the transport term \( \nabla_r J \{n,p\} \)) is balanced by the generation/recombination contribute (which corresponds to \( U \{n,p\} = G \{n,p\} - R \{n,p\} \)). Not only are electrons and holes affected by this conservation principle: in fact, also fixed charges – like defects, dopants or trap levels – are conserved, even if in this particular case the transport term is not present. Defining the ionized effective concentration at level \( k \) as

\[
C_{\pm}^k = \begin{cases} 
N_{D,k}^+ & \text{for acceptors} \\
-N_{A,k}^- & \text{for donors}
\end{cases}
\] (3.22)

the corresponding continuity equation reads

\[
\frac{\partial C_{\pm}^k}{\partial t} + \sum_k (U_{n,k} - U_{p,k}) = 0
\] (3.23)

where the bracketed term includes, in truth, only indirect GR phenomena being the instantaneous net recombination rate for direct processes fully balanced between electrons and holes, and thus equal to zero. Hence also the overall density \( \rho \) must be rewritten in order to account for the fixed charge

\[
\rho = q \sum_k C_{\pm}^k + q p(r, t) - q n(r, t).
\] (3.24)

Finally we can write the complete set of equations composing the 3D time variant drift-diffusion model as

\[
\begin{align*}
\frac{\partial n(r, t)}{\partial t} &= -\nabla \cdot \left( n(r, t) \mu_n \nabla_r \phi(r, t) - D_n \nabla_r n(r, t) \right) - U_n \\
\frac{\partial p(r, t)}{\partial t} &= \nabla \cdot \left( p(r, t) \mu_p \nabla_r \phi(r, t) + D_p \nabla_r p(r, t) \right) - U_p \\
\frac{\partial C_{\pm}^k}{\partial t} + \sum_k (U_{n,k} - U_{p,k}) &= 0 \\
\nabla^2_r \phi(r, t) &= -\frac{q}{\epsilon} \left( \sum_k C_{\pm}^k + p(r, t) - n(r, t) \right)
\end{align*}
\] (3.25)

whose unknown are \( n(r, t), p(r, t) \) and \( \phi(r, t) \) or, through the change of variables

\[
\begin{align*}
n &= n_i e^{\frac{E_{F_n} - E_F}{k_B T}} \\
p &= n_i e^{\frac{E_{F_p} - E_F}{k_B T}}
\end{align*}
\] (3.26)

they become \( E_{F_n}, E_{F_p} \) and \( E_F \). This formulation as a function of the IMREF levels allows to better understand the assumption of the drift-diffusion model according to which the energetic terms now depend only on the local value of the electric field (here represented by the IMREFs) and no more on its space- and time-variations.
While such a system can not be solved analytically without introducing specific approximations, at the thermodynamic equilibrium \((U \equiv 0, \text{ thus } p n = n^2)\) and for non-degenerate semiconductors the system (3.25) drastically simplifies into a unique expression: the Poisson-Boltzmann equation
\[
\nabla^2 r \left[ \epsilon \phi(r, t) \right] = -q \left[ N^+_D(\phi) - N^-_A(\phi) + n_i e^q \frac{E_F - E_F^*}{kT} - n_i e^q \frac{E_F - E_F^*}{kT} \right] \tag{3.27}
\]
in which only one generic level \(k\) has been accounted for in the computation of the fixed charges. Being a non-linear equation, also the (3.27) needs some approximations to be solved. One of the most used, not only at thermodynamic equilibrium, is the small-signal model, in which the bias voltages applied to the device are assumed of the form
\[
\begin{align*}
  i(t) &= i_0 + \hat{i}(t) \quad \text{for current generators} \\
  v(t) &= v_0 + \hat{v}(t) \quad \text{for voltage generators} 
\end{align*} \tag{3.28}
\]
where the terms \(i_0\) and \(v_0\) are constant (in time) while the \(\hat{i}(t)\) and \(\hat{v}(t)\) are small fluctuations (in amplitude) with respect to the constant component. Thus:
\[
\begin{align*}
  n &\approx n_0 + \hat{n}(t) \\
  p &\approx p_0 + \hat{p}(t) \\
  \phi &\approx \phi_0 + \hat{\phi}(t) \\
  C^\pm_k &\approx C^\pm_{0,k} + \hat{C}^\pm_k(t)
\end{align*} \tag{3.29}
\]
in which the perturbative terms are obtained through the linearization of the DD equations. Making all the calculations (see [46]) by using a 1st-order series expansion we finally found the following version of the model:
\[
\begin{align*}
  \frac{\partial \hat{n}}{\partial t} &= -\nabla \cdot \left( \hat{n} \mu_n \nabla \phi_0 + n_0 \mu_n \nabla \phi - D_n \nabla \hat{n} \right) - \frac{\partial U_n}{\partial n} \bigg|_0 \hat{n} - \frac{\partial U_n}{\partial p} \bigg|_0 \hat{p} \\
  \frac{\partial \hat{p}}{\partial t} &= \nabla \cdot \left( \hat{p} \mu_p \nabla \phi_0 + p_0 \mu_p \nabla \phi - D_p \nabla \hat{p} \right) - \frac{\partial U_p}{\partial n} \bigg|_0 \hat{n} - \frac{\partial U_p}{\partial p} \bigg|_0 \hat{p} \\
  \frac{\partial \hat{C}^\pm_k}{\partial t} + \frac{\partial U_n}{\partial n} \bigg|_0 \hat{n} + \frac{\partial U_p}{\partial p} \bigg|_0 \hat{p} - \frac{\partial U_n}{\partial p} \bigg|_0 \hat{n} - \frac{\partial U_p}{\partial n} \bigg|_0 \hat{p} &= 0 \\
  \nabla^2 r \hat{\phi} &= -\frac{q}{\epsilon} (\hat{p} - \hat{n})
\end{align*} \tag{3.30}
\]
(the symbol \(\frac{\partial U_{\bullet}}{\partial \bullet} \bigg|_0\) refers to a zeroth expansion term) valid if the hypothesis of constant mobility and diffusivity has been adopted.

As we can observe in the two versions of the DD model here presented, the (3.25) and the (3.30), besides the unknowns also the net rates of GR processes \(U\) appear. And, since tunneling mechanisms have their own generation/recombination rates, we are able to implement into a TCAD routine these rates in order to account for tunneling. Hence the net rates can be modeled as already explained in Sect. 2.2 and 2.3, and then translated into suitable \(I(V)\) expressions in order to study the electrical performance of the simulated device by the tunneling standpoint.

### 3.1.1 Boundary Conditions

By the mathematical standpoint the drift-diffusion model requires a suitable ensemble of initial conditions (IC) and boundary conditions (BC) in order to found one particular solution of the PDE system. Generally, the IC are imposed by the physics of the
Details about the Drift-Diffusion Transport Model

The first kind of boundary conditions that can be found in a semiconductor device are the so-called homogeneous Neumann conditions

\[
\frac{\partial n}{\partial \hat{n}} = 0, \quad \frac{\partial p}{\partial \hat{n}} = 0 \quad \text{and} \quad \frac{\partial \phi}{\partial \hat{n}} = 0
\]  

(3.31)

where \(\hat{n}\) is the unit vector (not to be confused with the perturbed electron concentration \(\hat{n}\)) normal to the interface between the semiconductor and an eventual insulator surrounding the device. Imposing a zero normal derivative of carrier concentrations and electrostatic potential is equivalent – in fact – to state the requirement of a null normal component of the current density, which corresponds precisely to the ideal insulation. Note that the (3.31) is a particular case of

\[
\begin{align*}
\rho n \frac{\partial \phi}{\partial \hat{n}} &= D_n \frac{\partial n}{\partial \hat{n}} \\
p \rho p \frac{\partial \phi}{\partial \hat{n}} &= -D_p \frac{\partial p}{\partial \hat{n}} \\
\epsilon_s \frac{\partial \phi}{\partial \hat{n}} \bigg|_s &= \epsilon_d \frac{\partial \phi}{\partial \hat{n}} \bigg|_d
\end{align*}
\]  

(3.32)

if we choose \(\epsilon_s \gg \epsilon_d\), otherwise conditions (3.32) concern the case of a generic dielectric interfacing with the semiconductor; in this case the requirement at the boundary is the continuity of the normal electric displacement vector. It is worth stressing that both (3.31) and (3.32) neglect the presence of surface states.

Another kind of BC are the Dirichlet-like conditions at contacts. If we take into account ideal metallic contacts, i.e. equipotential, the electrostatic potential \(\phi\) is the same at each point of the metal, so \(\phi(\mathbf{r}, t) \equiv \phi(t)\). For an ideal ohmic contact, behaving as a short-circuit \((R = 0)\), there is no potential drop at the interface so the BC commonly reads [47]

\[
\begin{align*}
n &= \frac{1}{2} \left( \sqrt{\sum_k C_k^{\pm 2} + 4n_i^2} + \sum_k C_k^+ \right) \\
p &= \frac{1}{2} \left( \sqrt{\sum_k C_k^{\pm 2} + 4n_i^2} - \sum_k C_k^- \right) \\
\phi &= v_{\text{appl}} + \text{const}
\end{align*}
\]  

(3.33)

where the constant reflects the presence of a shift due to the internal reference potential level; but in the situation of a device with all identical contacts (made by the same material) we can take the constant as zero and then the potential equals the applied voltage \(v_{\text{appl}}\), otherwise a built-in potential must be included. If the \(v_{\text{appl}}\) does not come from an ideal generator the BC written in (3.33) become, unfortunately, non-linear.

If we are in presence of a Schottky rectifying contact a linear potential drop arises across the device boundary, in correspondence of the contact itself. Several models can be exploited for this circumstance and one of them is the so-called “Sze-Bethe theory”. This kind of thermionic DD model involves the surface recombination within the expression of
3.1 Details about the Drift-Diffusion Transport Model

the carrier current densities, such as [47]

\[
\begin{align*}
J_n^\perp &= q v_{\text{sur},n} (n - n_{\text{sur}}) \\
J_p^\perp &= q v_{\text{sur},p} (p - p_{\text{sur}}) \\
\phi &= v_{\text{appl}} + v_{\text{bi}} + \text{const}
\end{align*}
\]

(3.34)
in which \(J^\perp \equiv J \cdot \hat{n}\) and where the constant in the third equation follows the same argument seen for the ohmic contact and where \(v_{\text{bi}}\) is the Schottky barrier built-in potential, \(v_{\text{sur}}\) the surface recombination velocity and \(n_{\text{sur}}\) or \(p_{\text{sur}}\) the carrier concentrations at the surface.

3.1.2 Discretization Procedure

The drift-diffusion model, as the other PDE-based models, needs to be numerically solved (e.g. through the Newton method) but in order to reduce the complexity of the system of coupled equations two steps are required: a spatial discretization of the device that allows a locally averaged calculation of the unknowns, then the solution of the discretized equations composing the model, which become now ordinary differential equations (which often follows a linearization). The various techniques of discretization are distinguished by the method generating the so-called mesh, i.e. the grid in which the device is divided, and are chosen accordingly to the application or to the working point of the structure. Some of them have been already mentioned in Sect. 2.3.1, like the finite elements FEM, but one of the most involved in TCAD numerical simulations of PDEs is the finite boxes FB scheme. By using a discretization method also known as Delaunay grid (from the Russian mathematician and mountain climber Boris Delaunay who introduced this method in his 1934 work “Sur la sphère vide”), that generates a division of the spatial domain into non-obtuse triangles and consequently in nodes, the FB computes the physical quantities through a \(n\)-order approximation technique. At the first order all the differential operators occurring in the 2D drift-diffusion model transform in such a way

\[
\begin{align*}
\frac{\partial}{\partial t} \int_S f \, ds &\approx \frac{\partial f_i}{\partial t} S_i \\
\int_\Gamma F^\perp \, d\gamma &\approx \sum_j l_{ij} \langle F^\perp \rangle_{ij} \\
\int_S c \, ds &\approx c_i S_i
\end{align*}
\]

(3.35)

where \(f\), \(F\) and \(c\) are, respectively, a scalar unknown, a vectorial unknown and a scalar constant, \(S\) and \(S_i\) are the generic area (or volume in 3D, obviously) and the area of the particular finite box surrounding the node \(i\)-th, \(l_{ij}\) the box side located between nodes \(i\) and \(j\) and, finally, \(\langle F^\perp \rangle_{ij}\) represents the average value of the normal vector \(F^\perp\). Accordingly to the Delaunay grid, the boxes, thus, must be defined with a certain geometrical criterium like connecting the bisectors of each side composing the triangles (see Figure 3.1).

So the Poisson equation can be rewritten as

\[
\frac{q}{\epsilon} \left( \sum_k C_{k,i}^{\pm} + p_i - n_i \right) S_i = \sum_j l_{ij} \langle \mathcal{E}^\perp \rangle_{ij} \\
\approx \sum_j l_{ij} \frac{\phi_i - \phi_j}{d_{ij}}
\]

(3.36)
3.1 Details about the Drift-Diffusion Transport Model

in which $d_{ij}$ is the distance between nodes $i$ and $j$ and where the second line concerns the first-order finite difference approximation of the electric field, averaged around the box sides.

**Figure 3.1.** Schematic representation of a Delaunay grid for a finite boxes discretization. The dashed shapes are the edges of the finite boxes and are calculated through the bisectors of the triangle sides converging in the corresponding node. The gray area represents the finite box surrounding the $i$-th node.

Discretizing the three continuity equations needed to complete the drift-diffusion model requires a more refined treatment, but for the moment let’s start from the first-order finite boxes approximation just exploited and apply it to the electron continuity equation

$$\frac{\partial n_i}{\partial t} = \frac{1}{q} \sum_j l_{ij} \langle J_{n_i}^\perp \rangle_{ij} - U_{n_i} S_i.$$  

becoming

$$\frac{\partial n_i}{\partial t} S_i = \frac{1}{q} \sum_j l_{ij} \langle J_{n_i}^\perp \rangle_{ij} - U_{n_i} S_i.$$  

Now assume that, being the current density conserved, its normal component $J_{n_i}^\perp$ is almost constant along the direction connecting the nearby nodes $i$ and $j$, thus:

$$\frac{1}{q} \langle J_{n_i}^\perp \rangle_{ij} \approx \frac{1}{q} \nabla \cdot J_{n_i}^\perp$$  

$$= -\mu_n n \frac{d\phi}{dr} + D_n \frac{dn}{dr}$$  

$$\approx -\mu_n n \frac{\phi_i - \phi_j}{d_{ij}} + D_n \frac{dn}{dr}$$

where $r$ is a generic reference direction and where the total derivatives have been used in place of the partial ones thanks to the assumption of constant $J_{n_i}^\perp$ between two nearby nodes. Solving the (3.39) in $n$ we obtain the non-linear solution

$$n(r) = n_i \left[ 1 - g \left( r, \Delta_{ij} \right) \right] + n_j g \left( r, \Delta_{ij} \right)$$  

(3.40)
in which
\[
\begin{cases}
\Delta_{ij} = q \frac{\phi_i - \phi_j}{k_B T} = \frac{\phi_i - \phi_j}{V_T} \\
g (r, \Delta_{ij}) = \frac{1 - e^{\frac{\Delta_{ij}}{k_B T}}}{1 - e^{\Delta_{ij}}}
\end{cases}
\]  
(3.41)

Now the so-called Scharfetter-Gummel approximation is assumed, so
\[
\frac{1}{q} \langle J^\perp_n \rangle_{ij} \approx \frac{1}{q} \nabla \cdot J^\perp_n \approx D_n \left[ n_j B (\Delta_{ij}) - n_i B (-\Delta_{ij}) \right]
\]  
(3.42)

where
\[
B(\alpha) = \frac{\alpha}{e^\alpha - 1}
\]  
(3.43)

is the Bernoulli function. Finally:
\[
\frac{\partial n_i}{\partial t} = \sum_j D_n \frac{l_{ij}}{s_i} \left[ n_j B (\Delta_{ij}) - n_i B (-\Delta_{ij}) \right] - U_{n,i}
\]  
(3.44)

and similarly for holes
\[
\frac{\partial p_i}{\partial t} = -\sum_j D_p \frac{l_{ij}}{s_i} \left[ p_i B (\Delta_{ij}) - p_j B (-\Delta_{ij}) \right] - U_{p,i}
\]  
(3.45)

while, regarding the fixed charges
\[
\frac{\partial C^\pm_{k,i}}{\partial t} = -\sum_{j \neq k} (p_{k,ji} - p_{k,ij}).
\]  
(3.46)

Equations (3.36), (3.44), (3.45) and (3.46) represent the finite boxes discretized version of the drift-diffusion model for each node \(i\) of the Delaunay grid. From this set of equations is now possible to extract a suitable approximation for the current at the contacts (see detailed calculations in [47])
\[
i_{\text{cont}} \approx \sum_{i,j} \left[ q \frac{D_n}{d_{ij}} \left( n_j B (\Delta_{ij}) - n_i B (-\Delta_{ij}) \right) + q \frac{D_p}{d_{ij}} \left( p_i B (\Delta_{ij}) - p_j B (-\Delta_{ij}) \right) \right] W l_{ij} +
\]  
(3.47)

+ \left[ q \frac{\partial n_i}{\partial t} + \frac{\partial p_i}{\partial t} - U_{n,i} + U_{p,i} \right] s_i W,

where \(W\) is the contact length and
\[
l_{ij} = \begin{cases} l_{ij} & \text{for internal } j \text{ nodes} \\ \frac{l_{ij}}{2} & \text{for } j \text{ nodes on the boundary} \end{cases}
\]  
(3.48)

and also to implement a new version of the boundary conditions in which, e.g. for ohmic contacts, hold the (3.33) where the unknown carrier concentrations and the potential at the left-hand side must be evaluated at the node \(i\).

The treatment we have seen until now concerns the geometrical discretization of the structure and, consequently, a simplification of the physical model itself. This strategy
allowed to perform the solution of the PDE system in terms of ODE regarding its spatial part (e.g. see in the second equality of (3.39)). What happens, instead, if we consider a variation in time of the variables defining the experiment, for instance introducing a voltage ramp applied to the contacts? TCAD simulators like Sentaurus Device exploit a so-called quasi-stationary procedure in which the macroscopic time variant physical quantities are evaluated, accordingly to the formulas previously derived, in a sort of time-discretization: in fact, dividing the time scale into a finite number of steps we can treat each of them as a quasi-equilibrium state whose initial conditions are the system solution at the previous step. Iterating this scheme the software is able to generate the required voltage ramp, where the number of steps is an user-defined parameter of the simulation.

3.2 A Mono-dimensional Preliminary Study

After a review on infrared photodiode properties, on tunneling effect and on physics-based numerical model issues, we are ready to introduce the simulation of an HgCdTe-based IR photodetector starting from a simple case: a 1D case-study of an ideally abrupt pn-HgCdTe homojunction in dark conditions. We will build the $I-V$ characteristic exploiting the tunneling models studied in the previous chapter showing the physical effects of those models on the junction behavior.

Although at this stage of simulation the differences between the various tunneling models are negligible and mono-dimensional results have been demonstrated overestimating the reverse dark current $I(V)$ with respect to the 2D ones and thus to the real case [48], a 1D model can reveal very useful by a qualitatively point of view, being a perfect workbench for the analysis of some tunneling dependencies on several physical quantities. So, in the following we will explore the influence of Trap-Assisted and Band-to-Band Tunneling processes on the reverse dark $I-V$ characteristic by tuning the most relevant of them for HgCdTe-based devices. In Table 3.1 are summarized the parameters entered in Sentaurus Device for our simulations (see also the Appendix A).

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>molar fraction $x$</td>
<td>0.2946</td>
</tr>
<tr>
<td>tunneling effective carrier masses</td>
<td>$m_e^* = 0.017 m_0$ and $m_h^* = 0.55 m_0$</td>
</tr>
<tr>
<td>cut-off wavelength</td>
<td>$\lambda_c = 5.29 \mu m$ (@ 80 K)</td>
</tr>
<tr>
<td>donor/acceptor concentrations (constant doping profile)</td>
<td>$N_D = 4 \cdot 10^{14}$ cm$^{-3}$ / $N_A = 2 \cdot 10^{16}$ cm$^{-3}$</td>
</tr>
<tr>
<td>doping reference concentration (see eqn. (2.58))</td>
<td>$N_{ref} = 10^{16}$ cm$^{-3}$</td>
</tr>
<tr>
<td>tunneling lifetimes (see eqn. (2.58))</td>
<td>$\tau_{min} = \tau_{max} = \tau_{dop} = 10^{-6}$ s</td>
</tr>
<tr>
<td>device dimensions</td>
<td>$w_{tot} = 8.5 \mu m$, $w_p = 7 \mu m$, $w_n = 1.5 \mu m$</td>
</tr>
<tr>
<td>maximum mesh density (quasi-neutral regions / junction)</td>
<td>$\leq 0.1 \mu m^{-1}$ / $\leq 0.001 \mu m^{-1}$</td>
</tr>
<tr>
<td>device temperature and energy trap $E_t$</td>
<td>variable parameters</td>
</tr>
</tbody>
</table>

Table 3.1. List of the main simulation input parameters of HgCdTe used in Sentaurus Device.

3.2.1 Energy trap $E_t$ Dependence

We have already mentioned in this work the concept of energy trap (the difference between the impurity level and the intrinsic one), a quantity related to the presence of
impurities lying within the forbidden gap and affecting the rate of trap-assisted GR mechanisms. Our choice is to run the simulation modulating $E_t$ from 20 to 40 meV with eight steps 2.5 meV each, with the aim to investigate on its effects in the electrical performance of the device. In Figure 3.2 is shown the reverse dark $I-V$ obtained through the quasi-stationary iterative scheme which includes, besides incomplete ionization, radiative-, SRH- and Auger-recombination also the Trap-Assisted Tunneling in the Hurkx formalism, i.e. through the net recombination rate of equation (2.64).

![Figure 3.2](image1.png)

**Figure 3.2.** Reverse dark $I-V$ characteristics with Hurkx trap-assisted-tunneling calculated at 85 K for nine values of $E_t$ (the energy increases downwards).

![Figure 3.3](image2.png)

**Figure 3.3.** Behavior of the current vs. the energy trap calculated at 85 K for four voltages (they are approximated values, being extracted from the Sentaurus quasi-stationary stepped voltage ramp).
3.2 A Mono-dimensional Preliminary Study

The simulation is performed at 85 K because the $E_t$-dependence is more evident at low temperatures, while at higher $T$ the choice of the level becomes less crucial with respect to the trend of the characteristic. As a matter of fact, the TAT itself is less important when the temperature raises, since the tunneling probability $P$ for TAT decrease with the increase of the temperature, because of the gap widening and of the carrier mean free-path reduction:

$$P_{\text{tat}}(T, E_t) = \exp \left[ -\frac{4V^2}{3q\hbar E_{\text{dep}}} \left( E_g(T) - E_t \right)^\frac{3}{2} \right].$$  \hspace{1cm} (3.49)

From Figure 3.3 we can note that – as predicted by theory – the tunneling probability and, thus, the total current decreases with the rising of $E_t$ with an exponential-like behavior, as stated by the equation (3.49). Remember that also the TAT component of the current has a similar dependence on $E_t$ since, as already written in equation (1.97)

$$I_{\text{tat}}(V_b, E_t, T) = A N_{\text{t}} w_{\text{dep}} \frac{\pi q^2 m^*_e \epsilon_{\text{dep}}(V_b) M^2}{\hbar^3} \exp \left[ -\frac{8\sqrt{2m_e^*(E_g(T) - E_t)^3}}{3q\hbar E_{\text{dep}}(V_b)} \right],$$ \hspace{1cm} (3.50)

the higher is the thermal energy of the system the greater is the trap ionization, yielding a less current contribution.

The Band-to-Band Tunneling is, in principle, not affected by the impurity levels since it represents a kind of direct mechanism between bands.

3.2.2 Temperature Dependence

As we have just stressed, Trap-Assisted Tunneling tends to disappear as the temperature increases (indicatively, over about 100 K). In Figure 3.4 we find the $I-V$ dark characteristics at a fixed value of $E_t = 30$ meV with and without the Hurkx Trap-Assisted model, and also including the Dynamic Non-local Path model, thanks to the net recombination rate written in (2.79). Again the calculations are performed only for a temperature of 85 K because at higher $T$ the difference between the two curves, with and without TAT, becomes inappreciable. From the same figure it is clear that, due to the modification of the tunneling path discussed in the previous chapter, the current contribution coming from a non-local approach is well above the simple TAT model, by a factor of about $1 - 2$.

Although Band-to-Band Tunneling is unsensible to the presence of traps located in the forbidden gap, the other mechanisms accounted for in this simulations can be, instead, trap-assisted, like SRH generation/recombination are. In this sense the value of $E_t$ remains always important. Furthermore BTBT survives also at high thermal energies, contrarily to the TAT. For these reasons we declared in all the following simulations the value $E_t = 30$ meV, in order to work with homogeneous results, and we also used as input parameter the temperature $T$, by setting its value at 85, 109 and 143 K. Our model of BTBT exploited the Kane-like generation rate

$$G_{\text{btbt}} = A\epsilon_x^\beta \exp \left[ -\frac{B}{\epsilon_x^\beta} \right],$$ \hspace{1cm} (3.51)

where the subscript refers to the unique $x$-axis and in which we adapt the generic coefficients (see equations (2.54) and (2.55)) to the HgCdTe material according to [34] as

$$\begin{align*}
A &= -\frac{2\sqrt{2m_e^*}}{4\pi^2 \hbar^2 \sqrt{E_g}} \\
B &= \frac{\sqrt{m_e^* E_x^3 (300 \text{K})}}{2\sqrt{2q\hbar}} \\
\beta &= 2
\end{align*}$$ \hspace{1cm} (3.52)
3.2 A Mono-dimensional Preliminary Study

Figure 3.4. Comparison of Hurkx trap-assisted-tunneling dark reverse $I-V$ characteristics (solid line) calculated at $85$ K and $E_t = 30$ meV, also with the Dynamic Non-local Path model (the upper curve) and the characteristic without tunneling processes (the lower one).

The C++ code reported in Appendix B implements the equation (3.51) into the Sentaurus Physical Model Interface (PMI). Compiling it through the Compact Model Interface (CMI) the resulting *.so.linux_gnu shared object file was included in Sentaurus in order to enter this BTBT model into the simulation.

Figure 3.5. Comparison between Band-to-Band tunneling $I-V$ characteristic calculated at $85$ K and $E_t = 30$ meV with the no-tunneling one. Solid line: BTBT. Dashed: no-tunneling.
3.2 A Mono-dimensional Preliminary Study

Figure 3.6. Comparison between Band-to-Band tunneling $I-V$ characteristic calculated at 109 K and $E_t = 30$ meV with the no-tunneling one. Solid line: BTBT. Dashed: no-tunneling.

Figure 3.7. Comparison between Band-to-Band tunneling $I-V$ characteristic calculated at 143 K and $E_t = 30$ meV with the no-tunneling one. Solid line: BTBT. Dashed: no-tunneling.

Figures 3.5–3.7 represent, for each value of $T$, the comparison between the curve with the Band-to-Band model and that one with no tunneling effects.

Firstly one should note the threshold-like behavior of the characteristics: as a matter of fact, the direct tunneling between valence and conduction band should occur only if the electric field in the junction (for instance provided by a bias voltage $V_b$, as in our case) becomes equal or greater than a critical field strength, in such a way that the energy of carriers in the semiconductor is enough to the interband transition. So, the behavior of
$I_{BTBT}(V_b)$ consists of a first zone (i.e. in the low-voltage domain) in which there is no BTBT contribution (the two curves coincide perfectly) and a second one where, after an upwards bend, the tunneling produces a sudden rise in the current almost linearly in $V_b$.

Moreover, it’s worth noting the good trend with respect to the permanence of BTBT current contribution – at constant voltage – with the increase of $T$. This is well shown in Figure 3.8 where, as theoretically predicted, is evident the BTBT also at higher temperatures. Furthermore it is also evident that the higher is the temperature, the greater is

![Figure 3.8](image1.png)

**Figure 3.8.** $I$-$V$ characteristics (at $E_t = 30$ meV) with Band-to-Band tunneling calculated at 85, 109 and 143 K (the temperature increases upwards) for our Hg$_{0.7054}$Cd$_{0.2946}$Te $pn$ homojunction.

![Figure 3.9](image2.png)

**Figure 3.9.** Same plot of Figure 3.5 with the addition of $B_{0.5}$ and $B_{1.5}$ curves.
the "threshold" voltage at which the bend occurs (i.e. the tunneling contribution), since the electric field necessary to counteract the energy gap widening must be stronger. As reported in [5] and for high $V_b$:

$$I_{btbl}(V_b, T) = V_b \frac{A q^3 \mathcal{E}(V_b)}{4 \pi \hbar^2} \sqrt{\frac{2m^*}{E_g(T)}} \exp \left[ -4 \sqrt{\frac{2m^* E_g^3(T)}{3q \hbar \mathcal{E}(V_b)}} \right]$$

whence comes out the (quasi) linear behavior of $I_{btbl}(V_b)$, over a certain voltage, confirmed

**Figure 3.10.** Same plot of Figure 3.6 with the addition of $B_{0.5}$ and $B_{1.5}$ curves.

**Figure 3.11.** Same plot of Figure 3.7 with the addition of $B_{0.5}$ and $B_{1.5}$ curves. Here the voltage range (10 V) is not enough to show properly the rise of $I_{btbl}$ for the $B_{1.5}$ curve.
Another important feature about the threshold concerns its position: the voltage at which the bend occurs is quite sensitive – as well as to the temperature – also to the value of the coefficient $B$ within equation (3.51). The plots in Figures 3.9 to 3.11 demonstrate this behavior when $B_{0.5} = 0.5B$, $B_1 = 1B$ or $B_{1.5} = 1.5B$ are accounted for in place of $B$.

The simple 1D model here presented can be considered not so accurate by the quantitative point of view but, nonetheless, very instructive within a phenomenological analysis: it includes, as a matter of fact, all the main features connected to the tunneling process and, as predicted before, it represents a valid workbench for the study of the main factors affecting the behavior of tunneling dark curves. If one search for a more accurate description, the simulation should account for some new realistic features like the shape of the junction, here assumed perfectly abrupt (from which arises the assumption $\beta = 2$) with a triangular potential barrier, and also other information related to the device geometry might be included. This will be the aim of the following sections, were we will simulate a 2D HgCdTe-based $p$-$i$-$n$ IR photodiode, both under dark and lighting conditions.

### 3.3 Device Project and its Equilibrium Features

In this section we come finally to the description of a specific Hg$_{0.7054}$Cd$_{0.2946}$Te-based device modeled through Sentaurus for tunneling simulations. In Figure 3.12 it is shown the geometry of our structure. As one can see, the simulated 2D $p$-$i$-$n$ IR photodiode consists

![Diagram of a 2D p-i-n IR photodiode](image)

**Figure 3.12.** Schematic representation of our $p$-$i$-$n$ IR photodiode. The light-gray lines on the right side of the picture refer to the non-implemented part of the device: thanks to its symmetry, only the left side has been simulated. This allow a simulation less computational cost.

- from top to bottom – of a superficial $p$ doped layer ($N_A = 10^{17} \text{ cm}^{-3}$) $2 \mu$m deep on which it is present the ohmic anodic contact-ring (consider the picture as a transversal slice of the corresponding 3D device, resulting that the two upper contacts belong to the same metallic structure). Then, after an ideally abrupt $pi$ junction, there is the $3 \mu$m deep absorber intrinsic layer, which is lightly doped (in truth this is a $\pi$-type region with
Finally we find the second ideally abrupt junction and the 2 μm deep $n$-type base layer, doped with $N_D = 10^{17}$ cm$^{-3}$. At the bottom of the base layer we put the ohmic cathode contact, which closes the bias circuit employed during the simulations.

![Diagram](image)

**Figure 3.13.** The mesh of our structure used in the simulations. Note that in correspondence of the two junctions (located at 2 and 5 μm from the cathode) the grid becomes more dense: here the triangular region has an edge length in depth-axis direction of 1 nm (for this reason the discretization is not well appreciable at the junctions).

This described geometry has been implemented in Sentaurus Workbench through the tool SDE (Sentaurus Structure Editor) thanks to a *_sde_dev.cmd* command file, which contains all the geometry and doping instructions. Once this file has been compiled the grid and boundary file *tdr* is automatically generated. In order to perform more precise calculations in correspondence of the junctions – where the electric field change more rapidly – we built a Delaunay grid such that in quasi-neutral regions each triangular region has an edge length in depth-axis direction of 100 nm, while in the junction regions (i.e. 0.14 μm symmetrically around the junction, as indicated in Figures 3.12 and 3.13) this length is 1 nm. Through a second tool, the so-called SDEVICE, which takes as input the grid and boundary file *tdr*, the parameter file *par* containing all the material specifications (summarized in Appendix A) and also a second command file *des.cmd* in which are included the physical and mathematical instructions for the experiment, Sentaurus can perform the simulation.

The structure analysis at equilibrium conditions starts from the electric field along the device. Since the abrupt junctions induce pronounced and sharp spikes in the trend of $E$ we modify the doping profile by implementing a gaussian grading such that the total doping concentration as a function of the depth $N(z)$ is [49]

$$N(z) = \frac{N_{\text{peak}}}{e^{\frac{(z-z_{\text{peak}})}{\sigma_z}}^2}$$  \hspace{1cm} (3.54)

and the standard deviation is

$$\sigma_z = \frac{z_{\text{base}} - z_{\text{peak}}}{\sqrt{-2 \cdot \ln \frac{N_{\text{base}}}{N_{\text{peak}}}}}$$  \hspace{1cm} (3.55)

in which $N_{\text{peak}} = 10^{17}$ cm$^{-3}$, $N_{\text{base}} = 10^{14}$ cm$^{-3}$ and $z_{\text{base}} - z_{\text{peak}} = 0.14$ μm coincides with the more dense grid regions symmetrically defined around the junctions.
3.3 Device Project and its Equilibrium Features

**Figure 3.14.** Simulated vertical component of the electric field of our Hg$_{0.7054}$Cd$_{0.2946}$Te p-i-n IR photodiode for two kinds of doping profiles at the junctions, at zero-bias and in dark conditions at 85 K.

**Figure 3.15.** Simulated vertical component of the electric field of our Hg$_{0.7054}$Cd$_{0.2946}$Te p-i-n IR photodiode for two kinds of doping profiles at the junctions, at zero-bias and in dark conditions at 140 K.
3.3 Device Project and its Equilibrium Features

Figure 3.16. Simulated vertical component of the electric field of our Hg$_{0.7054}$Cd$_{0.2946}$Te $p$-$i$-$n$ IR photodiode for two kinds of doping profiles at the junctions, at zero-bias and in dark conditions at 200 K.

Figure 3.17. Summary of the simulated vertical component of the electric field of our Hg$_{0.7054}$Cd$_{0.2946}$Te $p$-$i$-$n$ IR photodiode for the gaussian graded doping profile at the junctions, at zero-bias and in dark conditions.
In order to have a spike reduction of the electric field at the junctions we choose to use the gaussian graded doping. This allows, also, a more realistic device simulation, under the physical standpoint.

In Figure 3.18 we report the energetic bands taken at $V_b = 0$ V in dark conditions for three different lattice temperatures. As expected the energy gap lies in the range $E_g \gtrsim 0.23$ eV and it is increasing with the raising of the temperature. As the temperature increases, the number of ionized charges increases too (see also Figure 3.19), resulting in a less effective doping. This, in turn, yields a lower electric field within the intrinsic region. The asymmetric behavior of $E$ and of the band structure – especially noticeable at lower temperatures – is due to the fact that the intrinsic layer is a $\pi$-type region (i.e. a $p$-type lightly-doped region), thus the changes in the behavior of the energy levels and of the electric field at the $pi$ junction result less evident than the ones at the $in$ junction.

![Simulated band diagram](image)

**Figure 3.18.** Simulated band diagram of our Hg$_{0.7054}$Cd$_{0.2946}$Te $p$-$i$-$n$ IR photodiode at zero-bias and in dark conditions at 85, 140 and 200 K. The plot shows also (right $y$-axis) the total doping concentration (in cm$^{-3}$ or, better, in cm$^{-2}$, since this is a 2D structure).

Figure 3.19 shows the effects of the *Incomplete Ionization* model at equilibrium and for various temperatures and both for the abrupt and graded junction: since at low temperature only a fraction of acceptor and donor atoms is activated, the charges do not reach the doping concentration. While the temperature increases, this fraction becomes more important and tends to the doping concentration value. As one can see in the figure, this holds especially for $p$-type HgCdTe, where the activation energy for acceptors is less than the one for donors in $n$-type material. Moreover, it is noticeable that the graded profiles allow to reduce also unwanted peaks in the doping concentration, as shown in Figure 3.20 where only this configuration is summarized for each temperature.

In Figures 3.21 – 3.23 we find, then, the dark reverse characteristics obtained at 85, 140 and 200 K in both doping profiles cases. Besides the incomplete ionization, these
Figure 3.19. Simulated comparison between the doping concentration (the total – “implanted” – contribution and the ionized one) for the abrupt junctions and in the gaussian graded case for three different temperatures. For this simulation the IncompleteIonization Sentaurus model has been activated through the *des.cmd command file.
simulations account also for the SRH, radiative and Auger recombination processes (the results with the tunneling will be presented in the next chapter). The test was executed also in forward bias conditions but the use the two different profiles does not affect so much the characteristic. In reverse conditions, instead, the difference is pretty evident.

Figure 3.20. Summary of the simulated concentrations of the doping ionized fraction compared with the doping concentration level of our Hg$_{0.7054}$Cd$_{0.2946}$Te $p$-$i$-$n$ IR photodiode at zero-bias and in dark conditions at 85, 140 and 200 K. The plot shows only the gaussian graded profile.

Figure 3.21. Simulated reverse dark characteristic at 85 K for the abrupt junctions and graded profile cases.
Figure 3.22. Simulated reverse dark characteristic at 140 K for the abrupt junctions and graded profile cases.

Figure 3.23. Simulated reverse dark characteristic at 200 K for the abrupt junctions and graded profile cases.
The current generated by the device with abrupt junctions is higher than the one of its graded version, whatever the temperature is. This is due, probably, both to the drift and the diffusion terms in the current density for electrons and holes. In fact, the higher electric field in correspondence of the junctions should give to the carriers an additional driving force contributing to the total $J$: here the diffused carriers are swept more strongly towards their respective regions (electrons to the $n$-type and holes to the $p$-type region). At the same time also the diffusion term increases, since in the abrupt junction case also the carrier concentration experiences an abrupt trend, resulting in a higher value of $\nabla n$ and $\nabla p$ within $J$ and then in the continuity equations. The combined effect of these two mechanisms should explain the increased current shown in Figures 3.21 – 3.23. In order to see the trend of the electric field for the abrupt junctions also outside the equilibrium we calculated it for a $V_b = 2$ V in reverse conditions. The results are reported in Figures 3.24 to 3.26.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3_24.png}
\caption{Simulated electric field for both doping configuration, a reverse bias of $V_b = 2$ V and 85 K.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3_25.png}
\caption{Simulated electric field for both doping configuration, a reverse bias of $V_b = 2$ V and 140 K.}
\end{figure}
3.3 Device Project and its Equilibrium Features

From the analysis of the electric field in reverse conditions at $V_b = 2$ V is evident once again that the best choice – between the abrupt junction and the gaussian graded doping profile – is represented by the smooth function $N(z)$ for the total doping concentration reported in the equation (3.54) although, in the real devices, the broadening of such a grading is more pronounced than this because of some technological reasons. Note also that at 2 V (we are already in the punch-through condition) the depleted region is larger in the case of abrupt junctions with respect to the graded ones. This fact seems suggesting a more attention on the modeling of tunneling processes, for instance via the analysis of the real trend of energetic bands, barrier potential or electric field within the depleted region.

Once the operating features at equilibrium of our p-i-n structure have been verified, we will focus in the following chapter on its performance in dark and light conditions and reverse bias regime under the tunneling standpoint. A parametric study, as already discussed for the 1D $pn$ junction, will be accomplished in order to investigate the eventual modifications of 2D simulation on the physical and analytical parameters of tunneling models. Then a new recent modification of the Kane Band-to-Band generation rate for p-i-n structures will be tested, given that almost all models reported in literature are written explicitly for $pn$ junctions and not for the p-i-n ones. Moreover, common models make use of the constant electric field approximation while in our case this is not true (see Figures 3.24 – 3.26), as remarked just above. This new model makes use of an electric field expression in which are included an average value $\mathcal{E}^{D}_{av} \equiv 1$ – function of the main physical quantities related to the depletion region, and weighted by a parametric factor $D$ – and the total field (simply reported as the usual $\mathcal{E}$).

**Figure 3.26.** Simulated electric field for both doping configuration, a reverse bias of $V_b = 2$V and 200 K.
References


CHAPTER 4

RESULTS, COMMENTS AND PERSPECTIVES

As we did for the 1D HgCdTe-based pn junction, in this chapter we will firstly test our p-i-n photodiode through the parametric response of tunneling models in dark conditions: the dependencies of the I-V characteristic on the energy trap and temperature are again the preliminary target of this last section. Then we will analyze more deeply the Hurkx TAT effects both for the abrupt junctions device and for the graded doping one. Since the Schenk version of Trap-Assisted Tunneling is designed for indirect-gap semiconductors – especially for the Silicon – we do not include such model.

After the introduction of a C++ routine computing the BTBT generation rate, very similar to the 1D one exploited in the previous chapter, we will see its impact both through the Kane formalism and also with a new edition of it, expressly conceived for p-i-n structures, in which the expression of the electric field in the depleted region appears slightly modified. The present work, thus, applies for the first time such Kane-modified model to a HgCdTe-based IR p-i-n photodetector.

Finally we will present a brief analysis of our device under illumination conditions, giving also some basic hints of performance optimization with respect to the tunneling phenomenon.

4.1 Tunneling Simulations in Dark Conditions

The most interesting feature in this kind of simulations is without doubt the dark characteristic. As we have already seen, tunneling phenomena are present also without illumination and so they contribute heavily to the dark current, making crucial the difference between the photodetector operating conditions with and without the IR light. The first step we will do is to verify in our new 2D structure the correct dependence of the dark current, including tunneling, on the two most important physical parameters: the energy trap $E_t$ and the temperature $T$. Clearly the first one concerns only the TAT mechanism, while we already know that the lattice temperature should influence both processes, TAT and BTBT, even if in a different way. Remember that, since our simulations include Auger, radiative and SRH generation/recombination processes, the value of $E_t$ is not irrelevant in BTBT simulations: in fact the energy of impurities affects all kinds of SRH processes, resulting in a different amount of dark current, that is inversely proportional to $E_t$ itself.
4.1 Tunneling Simulations in Dark Conditions

For almost all of the following simulations we will compare the results related to the graded junction with the abrupt one, in order to be well aware of this difference. This is an useful tool for designing such kind of photodetectors and also to improve the robustness of future models.

4.1.1 $E_t$ and $T$ Parametric Analysis

In Figure 4.1 are reported the simulated dark characteristics for a set of temperatures and also at three different values of the parameter $E_t$. As expected from theoretical calculations and from the previous 1D simulation, the Trap-Assisted Tunneling tends to vanish if the temperature increases and for high trap energies. Remember that $E_t$ is computed as the difference between the impurity level and the intrinsic one. The main qualitative difference with respect to the mono-dimensional case already reported is the tunneling permanence at higher temperatures: if in the 1D $pn$ diode the TAT is essentially negligible at about 100 K, here we can still observe a small deviance from the dashed curve.
at 120 K, especially at low $E_t$ values. Thus, the Figure 4.1 confirms also the trend of $I(E_t)$: in fact the higher is the $E_t$ value the smaller is the tunneling effect, since we are going away from the midgap level of HgCdTe.

![Figure 4.1](image1.png)

**Figure 4.1.** A zoom of the same plot of Figure 4.1 at 85 and 100 K, where the grey lines refer to the abrupt junctions. Dashed: without Hurkx TAT model. Solid: with Hurkx TAT model.

In Figure 4.2 we show the comparison between the gaussian graded case and the abrupt one. In the plot it is noticeable the same behavior also for this second configuration except for the currents magnitude that, as always, is greater with respect to the graded doping profile. Since we are interested in the quantitative importance of such a choice about the doping, in the next paragraph we will see how the profile at the junctions could affect the tunneling contribution.

Setting a constant energy trap value, $E_t = 30 \text{ meV}$, we tested also the BTBT behavior with respect to the temperature increase. In the new C++ routine implemented for these simulations (see Appendix C) the code was ameliorated through the addiction of an if statement which computes the exact $E_g$ and $m_t$ values according to the gap widening and the tunneling mass modification with the temperature increase (all the coefficients reported in the code are extracted from the HgCdTe parameter files *.par). In this way the energy gap lies in the range between $E_g(85 \text{ K}) = 0.235537 \text{ eV}$ and $E_g(200 \text{ K}) = 0.261219 \text{ eV}$,
while for the effective tunneling mass is \( m_t(85 \text{ K}) = 0.0171517 m_0 \) and \( m_t(200 \text{ K}) = 0.0188990 m_0 \). The simulation results are shown in Figure 4.3.

**Figure 4.3.** Simulated dark reverse current with (solid) and without (dot-dashed) Kane Band-to-Band Tunneling model (see Appendix C) for the gaussian graded doping profile device at three temperatures and \( E_t = 30 \text{ meV} \).

In Figure 4.3 we can find the comparison between the \( I-V \) characteristics, including BTBT, at various temperatures. The first interesting element is the good trend of the dark current with respect to the tunneling overcome at a given bias value: when the temperature increases the electric field (and, thus, the bias \( V_b \)) needed for the tunneling occurrence increases too, because of the bandgap widening. This effect generates also a downward shift of the \( I-V \) tunneling curve, after the bend, so the higher is the temperature the smaller is the BTBT current contribution.

The effects of combined TAT and BTBT will be analyzed only at the end of the chapter. Before this we will see some important features about these two mechanisms taken singularly. Under a general point of view, at the moment, we can say that the Trap-Assisted Tunneling is the prevalent mechanism determining the diode characteristic at low-medium bias voltages and low temperatures (less than \( \sim 120 \text{ K} \)). In the high bias range, instead, the Band-to-Band Tunneling dominates determining the current increase
of several orders of magnitude, even if the temperature tends to reduce its contribution.

### 4.1.2 Trap-Assisted Tunneling Simulations

Since a correct device design requires a detailed knowledge of the phenomena involved in its functioning, now we want to investigate the role of tunneling in the two doping profile configurations, again as a function of the physical parameters $E_t$ and $T$. Then let’s start from the analysis of TAT.

![Graph showing the ratio of the dark current with the Hurkx Trap-Assisted Tunneling model over the total dark current without TAT plotted as a function of the bias voltage at $T = 100$ K and $120$ K, both for graded profile (solid) and the ideally abrupt one (dashed).]

**Figure 4.4.** Ratio of the dark current with the Hurkx Trap-Assisted Tunneling model over the total dark current without TAT plotted as a function of the bias voltage at $T = 100$ K and $120$ K, both for graded profile (solid) and the ideally abrupt one (dashed).

In Figure 4.4 we present the function $\frac{I_{\text{tunn}}}{I_{\text{tot}}}(V_b)$, where $I_{\text{tunn}}$ is the dark current including the Hurkx Trap-Assisted Tunneling model while $I$ is the total dark current without TAT. Firstly one can notice the almost linear increase of the tunneling importance with the reverse voltage (this fact was a priori not so obvious and it represents a noticeable result by the simulation standpoint) except for a small low-bias range in which the trend seems to be slightly downward bowed. So, at first approximation (remember that usually a $p-i-n$ photodiode operates at $0.2 \text{ V} < V_b < 0.5 \text{ V}$) we can state that in reverse conditions the TAT contribution over the dark current is linear with respect to the bias voltage.
Furthermore the plot states unequivocally that TAT tends to vanish if the temperature increases: as a matter of fact, the black curves are denoted by a smaller ratio (keeping constant the bias) and also a smaller slope than the grey ones.

Nonetheless an unexpected result comes from Figure 4.4. if it is true that the current generated by the device with abrupt junctions is greater, now we have to notice that its TAT component (dashed curve) is smaller with respect to the graded solution. In other words, the abrupt junctions generate a higher current because of the high $\nabla n$ and $\nabla p$ values and not by reason of a Trap-Assisted Tunneling contribution. The tunneling component, in turn, is greater in the graded doping profile device, since in this case the depleted region is narrower (as already argued in the previous chapter) and then the tunneling probability is increased. The offsets found between the two doping profiles configurations, moreover, are enhanced by high $V_b$ (whatever other parameters), low $E_t$ (at the same temperature) and low $T$ (at the same energy trap).

![Figure 4.5. Dark I-V curves of our photodiode, with graded doping profile, comparing the Hurkx TAT model (solid lines) with its Non-local version (dashed lines) for a set of temperatures and at $E_t = 30$ meV.](image)

As a final test we computed the Non-local version of the Hurkx Trap-Assisted Tunneling model (already explained in detail in paragraph 2.4). In Figure 4.5 are shown the dark characteristic with and without the Non-local model for the gaussian grading case and at $E_t = 30$ meV. This model follows the same trends of the simple Hurkx TAT, although it
generates higher currents. This fact is mainly due to the particular choice of the tunneling path, according to which the electrons can be captured in or emitted from the defect level at any location between the starting position and the ending position.

\[
\frac{I_{\text{non-local}}}{I_{\text{local}}} = -\alpha(T) \cdot \log \left[ \beta(T) \cdot V_b \right]
\]

Figure 4.6. Ratio of the dark current with the Hurkx Non-local TAT model over the current with the local (standard) model as a function of the bias, both for graded profiles (solid) and for abrupt ones (dashed).

In order to have a quantitative idea of the Non-local TAT model contribution respecting to the standard one we chosen two configurations, with \( E_l = 30 \) meV at \( T = 100 \) K and \( T = 120 \) K. For each of them we computed the ratio respectively between the dark currents for the Hurkx model with and without the Non-local option. As shown also in Figure 4.5 the Non-local path method increases the dark current by a temperature-dependent factor: for instance at 85 K it corresponds to a whole order of magnitude, while at 100 and 120 K (see Figure 4.6) this factor is around a few units. Moreover Figure 4.6 shows an effect not well clear in Figure 4.5, that is the decrease of the current ratio with the rise of the bias voltage. This trend seems to have a log-like behavior such that

\[
\frac{I_{\text{non-local}}}{I_{\text{local}}} (V_b) \sim -\alpha(T) \cdot \log \left[ \beta(T) \cdot V_b \right]
\]

where \( \alpha(T) \) and \( \beta(T) \) are proper functions of the temperature (and obviously they are also \( E_l \)-dependent, even if here we do not investigate on it).

From the comparison between the curves in Figure 4.1 and 4.5 we can finally observe that, since the Non-local option overestimates the current thanks to its criterium of tunneling path determination, the decay of TAT with temperature is slower: in the second case, in fact, the tunneling is still evident at 160 K, while in the previous one at 120 K it was yet negligible (especially at high \( E_l \)). This fact must be kept in mind in design a device with such a kind of TAT model in order to correctly evaluate the dark current.
4.1.3 Band-to-Band Tunneling Simulations

In this section we will see some properties of the Kane\(^1\) Band-to-Band Tunneling model on the photodiode dark characteristic. In Figure 4.7 is reported the same trend of Figure 4.3 but in both doping profile configurations, and again with the assumption of \(E_t = 30\) meV. As already seen, the BTBT starts at medium-high bias values and decreases with the rise of temperature. There are mainly two physical quantities that contribute to the latter phenomenon: the effective electron tunneling mass \(m_t\) and the energy gap \(E_g\). The Kane coefficients used in our simulations were:

\[
A(T) = -\frac{q^2 \sqrt{2 m_t(T)}}{4\pi^3 \hbar^2 \sqrt{E_g(T)}}
\]  

\(1\)The BTBT models implemented in Sentaurus and described in Chapter 2 proved ineffective, so we proceeded to the use of a C++ routine, as in the 1D case, for the implementation of a Kane-like model (see the Appendix C).
and

\[ B(T) = \frac{\pi \sqrt{m_t(T) E_g^3(300 \text{ K})}}{2\sqrt{2} q h} \] (4.3)

in which \( m_t(T) \) and \( E_g(T) \) are increasing functions of the temperature, determining the fall of the BTBT generation rate

\[ G(T) = A(T) \cdot \mathcal{E}^2 \exp \left[ -\frac{B(T)}{\mathcal{E}} \right] \] (4.4)

and then of the current too, when the temperature increases. The ideal abrupt junction not only experiences a higher voltage for the tunneling onset, but also the magnitude of \( I(V_b) \) and its slope are different from the graded case: in the low-bias range the \( I_{\text{gauss}} \) is smaller than \( I_{\text{abrupt}} \) but, when the BTBT contribution starts, it overcomes the other one. Approximatively we can said that for our structure the abrupt junction induces a BTBT current underestimation almost equal to a temperature shift of \( \Delta T \approx 60 \text{ K} \): in fact the dashed curve at 140 K is very similar to the solid one at 200 K and, in turn, the dashed curve at 85 K follows the solid one at 140 K.

As we have already done for the TAT, now we investigate on the tunneling component over the total dark current. Figure 4.8 shows the trend of the ratio \( \frac{I_{\text{tunn}}}{I_{\text{tot}}} (V_b) \) for the BTBT model, both in presence of graded (solid curve) and abrupt (dashed curve) doping profile.
4.1 Tunneling Simulations in Dark Conditions

Total Generation rate $[\text{cm}^{-3}\text{s}^{-1}] \times 10^{25}$

Figure 4.9. Total Generation rate along the device thickness, including SRH, Auger, radiative and BTBT processes.

Figure 4.10. Ratio of the dark current with the Kane Band-to-Band Tunneling model over the total dark current without BTBT plotted as a function of the bias voltage for the gaussian graded doping profile. Here the vertical scale is linear.
Again, as well as for the TAT, the tunneling contribution above a given voltage is greater in the graded configuration. This seems a confirm of the depleted region narrowing when the total doping concentration is smoothed through a gaussian function. As a further proof we plotted in Figure 4.9 the total Generation rate obtained in both devices at a reverse bias of $V_b = 12$ V. Besides the depleted region narrowing, in in Figure 4.9 it is also evident the greater generation of carriers at low temperature, determining – in turn – a greater current.

Respecting to the plot about TAT current ratio (see Figure 4.4), here the trend is well different from a straight line: in fact the Band-to-Band tunneling contributes as an exponential function of the bias (for a demonstration, see the graded case in Figure 4.10).

In the 1D example we made a study on the $B^{Kane}$ coefficient. From that test emerged the variation of the BTBT onset in the $I-V$ characteristic depending on the value of $B^{Kane}$. In our 2D $p-i-n$ photodiode this kind of analysis will converge in an unique – and more physic-based – treatment, which is the focus of the next paragraph.

### 4.2 A new model for BTBT in $p-i-n$ structures

We have just seen the importance of a correct device description through the doping profile definition. If one consider the fact that almost all the tunneling models are conceived for ideally abrupt $pn$ junctions, with constant electric field and parabolic (or triangular) potential barrier approximations, it can be noticed that to apply a generic model to any device could be sometimes an inadequate procedure. With the aim to avoid it, some recent papers proposed a different way to write the Band-to-Band Tunneling Generation rate in $p-i-n$ structures [50, 51]. Such model accounts for a substitution in (4.4) of the term $E^2$ with the expression $E^{(D-1)}_av \cdot E$, in which $E_{av}$ represents the average field in depleted region and $D$ is a fitting parameter to be determined through experimental measurements (a rigorous theoretical justification about the above field splitting can be found in the Vandenberghe work [52]). Thus, the BTBT rate becomes:

$$G_{pin} = A \cdot E^{(D-1)}_av \cdot E \cdot \exp \left( -\frac{B}{E_{av}} \right).$$

(4.5)

The average field in the depleted region depends on the so-called tunneling path length $\ell_{\text{path}}$, i.e. the distance between the two turning points:

$$E_{av} = \frac{E_g}{q \ell_{\text{path}}},$$

(4.6)

where

$$\ell_{\text{path}} = \frac{\epsilon E_g}{q^2 N_A x_p}$$

(4.7)

in which the term $x_p$ is the extension of the depleted region in the $p$-type region (for a mono-dimensional representation see Figure 4.11).

The unknown $x_p$ can be determined through the charge neutrality equation

$$\frac{2 \epsilon (V_{bi} + V_b)}{q} + N_i w_i^2 - N_i^2 \frac{w_i^2}{N_D} = x_p^2 \left( \frac{N_A^2}{N_D} + N_A \right) + x_p \left( -\frac{2 N_A N_i w_i}{N_D} + 2 N_A w_i \right)$$

(4.8)

so

$$x_p = \frac{qw_i N_A (N_i - N_D) + \sqrt{qN_A N_D (qw_i^2 (N_D - N_i) (N_A + N_i) + 2 \varepsilon W \epsilon (N_A + N_D))}}{qN_A (N_A + N_D)}$$

(4.9)
4.2 A new model for BTBT in p-i-n structures

whose low-computational cost version is

\[ x_p = \sqrt{\frac{2 \epsilon N_i \mathcal{E} W}{q N_{A}^2}}, \tag{4.10} \]

and where we used the relation

\[ V_{bi} + V_b = \mathcal{E} W \tag{4.11} \]

in which \( W \) represents the total thickness of the device. Now, inserting \( x_p \) in the (4.7) and then substituting it within the expression of the average electric field we obtain the new Band-to-Band Generation rate for p-i-n diodes. Although this model depends also on a fitting parameter (i.e. the exponent \( D \)), such kind of description is, in principle, more adherent to the real physical features of the device. Supposing the tunneling spatially confined between the points \( x_a \) and \( x_b \) (see Figure 4.11), the BTBT current density along the axis perpendicular to the junctions will be

\[
J_{btbt} = q \int_{x_a}^{x_b} G_{pin} dx
\]

\[
= q A \left( \frac{E_g}{q \ell_{\text{path}}} \right)^{D-1} \frac{q N_{A}}{\epsilon} x_p \int_{x_a}^{x_b} e^{-B q E_g^3 \ell_{\text{path}}} dx
\]

\[
= q^{(3-D)} E_g^{(D-1.5)} \frac{A N_{A} x_p}{\epsilon} e^{-B q E_g^3 \ell_{\text{path}}} \left( x_b - x_a \right) \tag{4.12}
\]

Exploiting the simpler expression of \( x_p \) (see the code in Appendix D), we implement in Sentaurus the Generation rate written in (4.5) in order to test the influence of the
parameter $D$ with respect to the standard Kane formulation. Thus, inserting our values of $N_A$, $N_i$, $N_D$, $w_i$ and $\epsilon$ into (4.10) we found that the average electric field depends on the usual expression of the electric field through

$$E_{av}(E) = \frac{q N_A x_p(E)}{\epsilon} \simeq \delta \sqrt{E}$$

(4.13)

where $\delta = 1.24172 \cdot 10^2$. Then $G_{pin}$ becomes

$$G_{pin} = A \cdot \delta^{(D-1)} E^{\left(\frac{D+1}{2}\right)} \cdot \exp\left[-\frac{B}{\delta \cdot \sqrt{E}}\right].$$

(4.14)

For the first time, this thesis applies the method discussed above to an HgCdTe-based IR photodetector. In Figures 4.12 and 4.13 are shown the dark characteristics for two temperatures and in both doping configurations, comparing the behavior of the standard Kane BTBT model with the new one. The plots also point out how a variation of the $D$ parameter can affect the $I-V$ characteristic.

![Graph](image)

**Figure 4.12.** Comparison between the dark characteristic at $T = 85$ K with the Kane BTBT model and the new one, for $D = 1.5$, $D = 2$ and $D = 3$. 

$E_t = 30$ meV 
$T = 85$ K
4.2 A new model for BTBT in p-i-n structures

Figure 4.13. Comparison between the dark characteristic at $T = 140$ K with the Kane BTBT model and the new one, for $D = 1.5$, $D = 2$ and $D = 3$.

Figure 4.14. Total Generation rate, including SRH, Auger, radiative and BTBT ($G_{pin}$) processes.
As one can see in Figures 4.12 and 4.13, the new BTBT model has the property of changing the tunneling onset during the voltage ramp and also the slope of the characteristic after the bend. This is due to the effect of $D$ both on the pre-factor and on the exponential term in Kane’s formula. Again the current in the abrupt case is greater before the tunneling starting and, after, it becomes smaller than that one of the graded doping profile device. At low temperature the new model tends to anticipate the point where the current rise occurs, dominating with respect to the standard Kane model. If the temperature increases, and keeping constant $D$, the roles are inverted, at least for low values of the parameter. All these considerations perfectly correspond to what we expected by the mathematical standpoint: in fact, since we found $\mathcal{E}_{av} \simeq \delta \sqrt{\mathcal{E}}$, the limit $D \to 3$ corresponds to set

$$G_{pin} \sim A \cdot \delta^2 \mathcal{E}^2 \cdot \exp \left[ -\frac{B}{\delta \cdot \sqrt{\mathcal{E}}} \right].$$

(4.15)

The (4.15) more closely resembles to the Kane’s formula, except for the presence of a square root in the exponential term and of the constant $\delta$, which determines the discrepancies between the two models seen in Figures 4.12 and 4.13, being $\delta$ a positive quantity.

Regarding the comparison between the behavior of the two doping configurations, we investigate on the total Generation rate within the device. Figure 4.14 demonstrates that the reason according to which the Band-to-Band Tunneling current in the abrupt case is always smaller lies in the trend of $G_{tot}$ along the device and, especially, within the intrinsic region.

### 4.3 Tunneling Simulations under Illumination

The importance of the parameter $D$ is represented by the degrees of freedom which it offers when simulations are compared with measurements on a real device. Without this important issue of analysis, the study of our HgCdTe-based $p-i-n$ IR photodiode under illumination is intrinsically incomplete. This is the reason which leads us to consider in the last section of this work only the two standard tunneling models previously seen: Hurkx TAT and Kane BTBT. Also the case of abrupt junctions have been neglected, since the aim of the following tests is to provide the reader a brief insight on the behavior of a real device under illumination with respect to the tunneling framework. On the contrary, the abrupt case represents an ideal simplification introducing significant deviances from the more realistic graded doping situation, as widely argued in the previous sections.

In order to simulate the illumination we add to the Sentaurus project a section related to the optical properties of the infrared light beam. The radiation we used consists in a set of 500 rays of unpolarized monochromatic light ($\lambda = 1 \, \mu m$) with variable power densities, and directed perpendicularly to the superficial $p$-type facet. The illuminated window is represented by the linear region ($1.5 \, \mu m$ wide) near the ohmic anodic contact (as a reference, see Figure 3.12).

As shown in Figures 4.15 to 4.17 the photocurrent generated within the device – keeping constant all the optical parameters – changes inversely with the temperature. By increasing the illumination power density from $1 \, W/cm^2$ to $10 \, W/cm^2$ the photocurrent, obviously, increases too. The interesting fact is that this mechanism tends to hide the Band-to-Band tunneling effects on the characteristic: in fact, a higher photocurrent implies a greater electric field (and, hence, bias voltage) at which the BTBT becomes to dominate the curve. So, at a given temperature, the illumination condition favorites the $I-V$ bend shift towards higher $V_b$ values. The effects of the Trap-Assisted Tunneling, instead, at this stage are insignificant being the photocurrent several orders of magnitude greater than
4.3 Tunneling Simulations under Illumination

Figure 4.15. The photocurrent in our p-i-n structure at 85 K and various power densities.

Figure 4.16. The photocurrent in our p-i-n structure at 140 K and various power densities.
4.3 Tunneling Simulations under Illumination

**Figure 4.17.** The photocurrent in our p-i-n structure at 200 K and various power densities.

**Figure 4.18.** Comparison between photocurrents at 85 K with the case of a shallow pi junction.
the tunneling contribution. Nonetheless, TAT mechanism could become crucial in the discrimination between dark and light or at low values of irradiance.

The second test concerns an attempt to optimize the photodetector with respect to the tunneling problem. In the Chapter 2 we seen that a thinner $p$-type region (in those devices where the illuminated surface is $p$-type) should increase the quantum efficiency. In order to demonstrate this property and to analyze the effects on the characteristic we modified the geometry of our $p$-$i$-$n$ diode by reducing the thickness of the upper region, from 2 $\mu$m to a value of 0.5 $\mu$m. Figures 4.18 and 4.19 demonstrate that a shallow junction (dashed curve) improves the photogeneration of carriers, inducing a higher current with respect to the standard geometry (lower gray lines). At the same time it is evident that at low temperature such improvement can lead to a current gain almost comparable to an irradiance doubling (upper gray line of Figure 4.18). Furthermore, at higher temperatures this gain increases up to some units of W/cm$^2$ (see Figure 4.19). Regarding the BTBT, it is also noticeable a slight shift of the curve bending towards the positive bias axis even if, usually, this kind of photodiodes operate within a lower voltage range.

4.4 Conclusions

The physics-based tunneling models represent a powerful investigation tool about the behavior of dark currents in infrared photodiodes. The present study revealed, besides the importance of a correct description of the structure and its features, some interesting phenomena related to the two main tunneling mechanisms: the Trap-Assisted (TAT) and the Band-to-Band one (BTBT). Under a general point of view, the study of two standard (and widely used) tunneling models, i.e. the Hurkx TAT and the Kane BTBT,
on our $\text{Hg}_{0.7054}\text{Cd}_{0.2946}\text{Te}$-based $p$-$i$-$n$ diode demonstrated a prevalence of TAT in the low-bias range, while at higher voltages the BTBT dominates. Furthermore the total dark current with the TAT model in the abrupt profile case is overestimated with respect to the graded one for every bias value. Instead, BTBT model applied to the abrupt junction configuration generates the underestimation of the total dark current above a threshold bias voltage with respect to the graded-doping one.

The Hurkx TAT model shown – as expected – an influence on the reverse dark $I$-$V$ characteristic which decreases with the temperature: in fact, if at low temperature the dark current is evidently enhanced, at 120 K the TAT effects are yet negligible. Instead, the Hurkx TAT Non-local counterpart at 160 K still induces effects. The role of the energy trap – generating higher dark current when $E_t$ decreases – has been also confirmed by our simulations.

By plotting the ratio between the dark current including the Hurkx TAT model and the dark current without any tunneling model – which has been proved to be a linear function of bias – we demonstrated a higher tunneling contribution in the device with gaussian graded profile junctions with respect to the one with abrupt junctions. Such result was unexpected, since the total reverse dark current is always greater in the abrupt case; the reason of this phenomenon lies, probably, in the depleted region narrowing which is present in the graded profile configuration. Furthermore the ratio between the dark currents including, respectively, the Non-local Hurkx TAT and the “standard” Hurkx TAT model shown a decreasing behavior with respect to the voltage bias through a log-like function of the form $\frac{I_{\text{non-local}}}{I_{\text{local}}} \sim -\alpha(T) \cdot \log [\beta(T) \cdot V_b]$, where $\alpha(T)$ and $\beta(T)$ are proper functions of the temperature.

The Kane Band-to-Band Tunneling implementation generates, as theoretically predicted, a threshold-like dark current which is enhanced with the device temperature decrease. Since the BTBT is an electric field-dependent phenomenon, a low temperature favors the tunneling onset at lower voltage bias with respect to the high temperature case, due to the bandgap widening with $T$. Moreover, also for the BTBT, the tunneling contribution is higher in the device with graded junctions even if, in this case, the result was predictable since after the BTBT threshold in such doping configuration the total dark current is always higher respecting to the abrupt one. The main difference with respect to the TAT model, instead, is represented by the trend of the ratio between the dark current with tunneling and the one without it as a function of the bias voltage, which has an exponential-like behavior instead of the linear trend seen for TAT. Also BTBT model – through the analysis of the total generation rate within the device – seems to confirm the depleted region narrowing in the graded-doping device. Furthermore, we demonstrated that – keeping constant the bias voltage – simulations on an ideally abrupt junction could generate a BTBT effect comparable to that one of a graded-doping device corresponding to a temperature shift of $\Delta T \sim 60$ K. This yields the importance of a correct description of the simulated structure.

Since the standard tunneling models are written for ideally abrupt $pn$ junctions, with the constant electric field and parabolic (or triangular) potential barrier approximation, we test (for the first time on an HgCdTe-based device) a new BTBT model expressly designed for $p$-$i$-$n$ structures. This model starts from the description of the field $E$ within the depleted region as a function of the tunneling path, then splits the term $E^2$ within the standard Kane’s expression into the product $E_{av}^{D-1} \cdot E$ in order to account for the average field. By changing the value of the parameter $D$ we observed how the model diverges from the standard Kane formulation, giving a higher or a smaller reverse dark current accordingly to both $D$ and temperature. Nonetheless, the real effectiveness of such model
requires more investigation, especially in terms of comparison with measurements on real devices, being $D$ a fitting parameter.

Finally, a test with our structure under illumination conditions has been performed. From the simulations emerges the photocurrent dependence on the light intensity (obviously) and also on the device temperature. Furthermore, increasing the IR power density makes the BTBT threshold shifting towards higher bias values, since the photocurrent tends to hide the tunneling effects. A final optimization hint has been given by decreasing the $pi$ junction depth: besides the increase of the quantum efficiency – resulting, as expected, in a higher current which emulates a light intensity increase of some units of $\text{W/cm}^2$, depending on $T$ – we observed again a slight shift of the BTBT threshold bias towards high voltage values.
References


Hg$_{0.7054}$Cd$_{0.2946}$Te Simulation Physical Parameters

Energy Gap

\[
\begin{align*}
E_g(x, 85 \text{ K}) &= -0.25821 + 1.84241 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 100 \text{ K}) &= -0.24993 + 1.82583 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 109 \text{ K}) &= -0.24499 + 1.81599 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 120 \text{ K}) &= -0.23898 + 1.80398 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 140 \text{ K}) &= -0.22811 + 1.78221 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 143 \text{ K}) &= -0.22649 + 1.77899 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 160 \text{ K}) &= -0.21728 + 1.76058 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 180 \text{ K}) &= -0.20648 + 1.73898 x - 0.81 x^2 + 0.832 x^3 \\
E_g(x, 200 \text{ K}) &= -0.19570 + 1.71740 x - 0.81 x^2 + 0.832 x^3 \\
\end{align*}
\]  
(A.1)

\[
\begin{align*}
E_g(0.2946, 85 \text{ K}) &= 0.235537 \text{ eV} \\
E_g(0.2946, 100 \text{ K}) &= 0.238933 \text{ eV} \\
E_g(0.2946, 109 \text{ K}) &= 0.240974 \text{ eV} \\
E_g(0.2946, 120 \text{ K}) &= 0.243446 \text{ eV} \\
E_g(0.2946, 140 \text{ K}) &= 0.247902 \text{ eV} \\
E_g(0.2946, 143 \text{ K}) &= 0.248574 \text{ eV} \\
E_g(0.2946, 160 \text{ K}) &= 0.252360 \text{ eV} \\
E_g(0.2946, 180 \text{ K}) &= 0.256797 \text{ eV} \\
E_g(0.2946, 200 \text{ K}) &= 0.261219 \text{ eV} \\
\end{align*}
\]  
(A.2)
Appendix A. Hg$_{20.7054}$Cd$_{0.2946}$Te Simulation Physical Parameters

Electronic Affinity

\[
\chi(x, 85 \text{ K}) = 4.5074 - 1.67138 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 100 \text{ K}) = 4.5007 - 1.48441 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 109 \text{ K}) = 4.4967 - 1.47641 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 120 \text{ K}) = 4.4918 - 1.46661 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 140 \text{ K}) = 4.4829 - 1.44891 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 143 \text{ K}) = 4.4816 - 1.44631 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 160 \text{ K}) = 4.4741 - 1.43131 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 180 \text{ K}) = 4.4653 - 1.41371 x + 0.65853 x^2 - 0.67642 x^3 \\
\chi(x, 200 \text{ K}) = 4.4566 - 1.39631 x + 0.65853 x^2 - 0.67642 x^3
\]

(A.3)

\[
\chi(0.2946, 85 \text{ K}) = 3.97758 \text{ eV} \\
\chi(0.2946, 100 \text{ K}) = 4.10325 \text{ eV} \\
\chi(0.2946, 109 \text{ K}) = 4.10161 \text{ eV} \\
\chi(0.2946, 120 \text{ K}) = 4.09960 \text{ eV} \\
\chi(0.2946, 140 \text{ K}) = 4.09591 \text{ eV} \\
\chi(0.2946, 143 \text{ K}) = 4.09538 \text{ eV} \\
\chi(0.2946, 160 \text{ K}) = 4.09229 \text{ eV} \\
\chi(0.2946, 180 \text{ K}) = 4.08864 \text{ eV} \\
\chi(0.2946, 200 \text{ K}) = 4.08511 \text{ eV}
\]

(A.4)

Electron and Hole Mobility

\[
\mu_e(x, 85 \text{ K}) = 7.28552 \cdot 10^5 - 5.18639 \cdot 10^6 x + 1.31388 \cdot 10^7 x^2 - 1.16509 \cdot 10^7 x^3 \\
\mu_e(x, 100 \text{ K}) = 4.88289 \cdot 10^5 - 3.35141 \cdot 10^6 x + 8.22026 \cdot 10^7 x^2 - 7.08365 \cdot 10^6 x^3 \\
\mu_e(x, 109 \text{ K}) = 3.94988 \cdot 10^5 - 2.65520 \cdot 10^6 x + 6.39394 \cdot 10^7 x^2 - 5.41957 \cdot 10^6 x^3 \\
\mu_e(x, 120 \text{ K}) = 3.07526 \cdot 10^5 - 2.00303 \cdot 10^6 x + 4.67589 \cdot 10^7 x^2 - 3.84557 \cdot 10^6 x^3 \\
\mu_e(x, 140 \text{ K}) = 2.05216 \cdot 10^5 - 1.26094 \cdot 10^6 x + 2.77119 \cdot 10^7 x^2 - 2.14066 \cdot 10^6 x^3 \\
\mu_e(x, 143 \text{ K}) = 1.93889 \cdot 10^5 - 1.18021 \cdot 10^6 x + 2.56715 \cdot 10^7 x^2 - 1.96047 \cdot 10^6 x^3 \\
\mu_e(x, 160 \text{ K}) = 1.42765 \cdot 10^5 - 8.20778 \cdot 10^6 x + 1.66979 \cdot 10^7 x^2 - 1.17666 \cdot 10^6 x^3 \\
\mu_e(x, 180 \text{ K}) = 1.02441 \cdot 10^5 - 5.44958 \cdot 10^6 x + 9.98315 \cdot 10^7 x^2 - 6.03630 \cdot 10^6 x^3 \\
\mu_e(x, 200 \text{ K}) = 7.52497 \cdot 10^4 - 3.64724 \cdot 10^5 x + 5.72630 \cdot 10^6 x^2 - 2.50859 \cdot 10^5 x^3
\]

(A.5)

\[
\mu_e(0.2946, 85 \text{ K}) = 4.30562 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 100 \text{ K}) = 3.32770 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 109 \text{ K}) = 2.90276 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 120 \text{ K}) = 2.49241 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 140 \text{ K}) = 1.95204 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 143 \text{ K}) = 1.88752 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 160 \text{ K}) = 1.57963 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 180 \text{ K}) = 1.31058 \cdot 10^4 \text{ cm}^2/\text{Vs} \\
\mu_e(0.2946, 200 \text{ K}) = 1.10898 \cdot 10^4 \text{ cm}^2/\text{Vs}
\]

(A.6)
\[ \mu_h(x, T) = 10^{-2} \cdot \mu_e(x, T) \]  

(A.7)

**Low Frequency Dielectric Constant**

\[ \epsilon(x) = 20.5 - 15.5 x + 5.7 x^2 + 1.0162 \cdot 10^{-13} x^3 \]  

(A.8)

\[ \epsilon(0.2946) = 16.4284 \]  

(A.9)

**Auger Coefficients**

\[
\begin{align*}
C_n(x, 85 \text{ K}) & = 2.689 \cdot 10^{-24} - 2.357 \cdot 10^{-23} x + 7.025 \cdot 10^{-22} x^2 - 7.096 \cdot 10^{-23} x^3 \\
C_n(x, 100 \text{ K}) & = 7.546 \cdot 10^{-24} - 6.553 \cdot 10^{-23} x + 1.942 \cdot 10^{-22} x^2 - 1.954 \cdot 10^{-22} x^3 \\
C_n(x, 109 \text{ K}) & = 1.554 \cdot 10^{-24} - 1.343 \cdot 10^{-23} x + 3.966 \cdot 10^{-22} x^2 - 3.982 \cdot 10^{-23} x^3 \\
C_n(x, 120 \text{ K}) & = 4.977 \cdot 10^{-24} - 4.277 \cdot 10^{-23} x + 1.259 \cdot 10^{-22} x^2 - 1.261 \cdot 10^{-22} x^3 \\
C_n(x, 140 \text{ K}) & = 3.442 \cdot 10^{-24} - 2.928 \cdot 10^{-23} x + 8.566 \cdot 10^{-23} x^2 - 8.546 \cdot 10^{-23} x^3 \\
C_n(x, 143 \text{ K}) & = 8.168 \cdot 10^{-25} - 6.931 \cdot 10^{-24} x + 2.028 \cdot 10^{-23} x^2 - 2.022 \cdot 10^{-23} x^3 \\
C_n(x, 160 \text{ K}) & = 2.465 \cdot 10^{-24} - 2.078 \cdot 10^{-23} x + 6.045 \cdot 10^{-23} x^2 - 6.007 \cdot 10^{-23} x^3 \\
C_n(x, 180 \text{ K}) & = 1.816 \cdot 10^{-24} - 1.517 \cdot 10^{-23} x + 4.387 \cdot 10^{-23} x^2 - 4.342 \cdot 10^{-23} x^3 \\
C_n(x, 200 \text{ K}) & = 1.368 \cdot 10^{-24} - 1.132 \cdot 10^{-23} x + 3.256 \cdot 10^{-23} x^2 - 3.211 \cdot 10^{-23} x^3
\end{align*}
\]

(A.10)

\[
\begin{align*}
C_n(0.2946, 85 \text{ K}) & = 2.93324 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 100 \text{ K}) & = 9.89019 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 109 \text{ K}) & = 2.23542 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 120 \text{ K}) & = 7.92801 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 140 \text{ K}) & = 6.43164 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 143 \text{ K}) & = 1.56002 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 160 \text{ K}) & = 5.28532 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 180 \text{ K}) & = 4.39713 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_n(0.2946, 200 \text{ K}) & = 3.70015 \cdot 10^{-26} \text{ cm}^6/\text{s}
\end{align*}
\]

(A.11)

\[
\begin{align*}
C_p(x, 85 \text{ K}) & = 1.673 \cdot 10^{-24} - 2.395 \cdot 10^{-23} x + 7.339 \cdot 10^{-23} x^2 - 7.655 \cdot 10^{-23} x^3 \\
C_p(x, 100 \text{ K}) & = 1.526 \cdot 10^{-24} - 1.326 \cdot 10^{-23} x + 3.930 \cdot 10^{-23} x^2 - 3.955 \cdot 10^{-23} x^3 \\
C_p(x, 109 \text{ K}) & = 1.257 \cdot 10^{-24} - 1.087 \cdot 10^{-23} x + 3.214 \cdot 10^{-23} x^2 - 3.228 \cdot 10^{-23} x^3 \\
C_p(x, 120 \text{ K}) & = 1.09 \cdot 10^{-24} - 8.671 \cdot 10^{-24} x + 2.553 \cdot 10^{-23} x^2 - 2.559 \cdot 10^{-23} x^3 \\
C_p(x, 140 \text{ K}) & = 6.988 \cdot 10^{-25} - 5.951 \cdot 10^{-24} x + 1.742 \cdot 10^{-23} x^2 - 1.739 \cdot 10^{-23} x^3 \\
C_p(x, 143 \text{ K}) & = 6.636 \cdot 10^{-25} - 5.643 \cdot 10^{-24} x + 1.650 \cdot 10^{-23} x^2 - 1.646 \cdot 10^{-23} x^3 \\
C_p(x, 160 \text{ K}) & = 5.017 \cdot 10^{-25} - 4.233 \cdot 10^{-24} x + 1.232 \cdot 10^{-23} x^2 - 1.255 \cdot 10^{-23} x^3 \\
C_p(x, 180 \text{ K}) & = 3.702 \cdot 10^{-25} - 3.096 \cdot 10^{-24} x + 8.961 \cdot 10^{-24} x^2 - 8.877 \cdot 10^{-24} x^3 \\
C_p(x, 200 \text{ K}) & = 2.794 \cdot 10^{-25} - 2.316 \cdot 10^{-24} x + 6.666 \cdot 10^{-24} x^2 - 6.578 \cdot 10^{-24} x^3
\end{align*}
\]

(A.12)
Appendix A.

Electron Effective Mass

\[
\begin{align*}
C_p(0.2946, 85 K) &= 2.84466 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_p(0.2946, 100 K) &= 1.98792 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_p(0.2946, 109 K) &= 1.79798 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_p(0.2946, 120 K) &= 1.59491 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_p(0.2946, 140 K) &= 1.29503 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_p(0.2946, 143 K) &= 1.25658 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_p(0.2946, 160 K) &= 1.06517 \cdot 10^{-26} \text{ cm}^6/\text{s} \\
C_p(0.2946, 180 K) &= 8.86494 \cdot 10^{-27} \text{ cm}^6/\text{s} \\
C_p(0.2946, 200 K) &= 7.46918 \cdot 10^{-27} \text{ cm}^6/\text{s}
\end{align*}
\]

that must be implemented in the net Auger rate

\[
R_A = (C_n n + C_p p) (np - n_i^2).
\]

Electron Effective Mass

\[
\begin{align*}
\begin{cases}
m_e^*(x, 85 K) = -0.0221098 + 0.172643 x - 0.1822 x^2 + 0.16481 x^3 \\
m_e^*(x, 100 K) = -0.0212823 + 0.169429 x - 0.1766 x^2 + 0.15964 x^3 \\
m_e^*(x, 109 K) = -0.0207934 + 0.167552 x - 0.1734 x^2 + 0.15666 x^3 \\
m_e^*(x, 120 K) = -0.0202020 + 0.165296 x - 0.1696 x^2 + 0.15315 x^3 \\
m_e^*(x, 140 K) = -0.0191470 + 0.161331 x - 0.1630 x^2 + 0.14707 x^3 \\
m_e^*(x, 143 K) = -0.0189899 + 0.160745 x - 0.1620 x^2 + 0.14619 x^3 \\
m_e^*(x, 160 K) = -0.0181085 + 0.157488 x - 0.1568 x^2 + 0.14136 x^3 \\
m_e^*(x, 180 K) = -0.0170901 + 0.153785 x - 0.1509 x^2 + 0.13600 x^3 \\
m_e^*(x, 200 K) = -0.0160848 + 0.150184 x - 0.1453 x^2 + 0.13095 x^3 
\end{cases}
\end{align*}
\]

\[
\begin{align*}
\begin{cases}
m_e^*(0.2946, 85 K) = 0.0171517 m_0 \\
m_e^*(0.2946, 100 K) = 0.0173846 m_0 \\
m_e^*(0.2946, 109 K) = 0.0175227 m_0 \\
m_e^*(0.2946, 120 K) = 0.0176907 m_0 \\
m_e^*(0.2946, 140 K) = 0.0179956 m_0 \\
m_e^*(0.2946, 143 K) = 0.0180409 m_0 \\
m_e^*(0.2946, 160 K) = 0.0182975 m_0 \\
m_e^*(0.2946, 180 K) = 0.0185992 m_0 \\
m_e^*(0.2946, 200 K) = 0.0188990 m_0
\end{cases}
\end{align*}
\]
Radiative Coefficients

\[
C_{\text{rad}}(x, 85 \text{ K}) = 3.228 \cdot 10^{-11} + 7.6204 \cdot 10^{-10} x - 5.4721 \cdot 10^{-10} x^2 + 4.2358 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 100 \text{ K}) = 2.882 \cdot 10^{-11} + 5.9116 \cdot 10^{-10} x - 4.2820 \cdot 10^{-10} x^2 + 3.3234 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 109 \text{ K}) = 2.719 \cdot 10^{-11} + 5.1624 \cdot 10^{-10} x - 3.7580 \cdot 10^{-10} x^2 + 2.9218 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 120 \text{ K}) = 2.551 \cdot 10^{-11} + 4.4342 \cdot 10^{-10} x - 3.2472 \cdot 10^{-10} x^2 + 2.5304 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 140 \text{ K}) = 2.312 \cdot 10^{-11} + 3.4673 \cdot 10^{-10} x - 2.5662 \cdot 10^{-10} x^2 + 2.0087 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 143 \text{ K}) = 2.281 \cdot 10^{-11} + 3.3512 \cdot 10^{-10} x - 2.4841 \cdot 10^{-10} x^2 + 1.9458 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 160 \text{ K}) = 2.129 \cdot 10^{-11} + 2.7948 \cdot 10^{-10} x - 2.0900 \cdot 10^{-10} x^2 + 1.6439 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 180 \text{ K}) = 1.984 \cdot 10^{-11} + 2.3053 \cdot 10^{-10} x - 1.7418 \cdot 10^{-10} x^2 + 1.3771 \cdot 10^{-10} x^3 \\
C_{\text{rad}}(x, 200 \text{ K}) = 1.866 \cdot 10^{-11} + 1.9362 \cdot 10^{-10} x - 1.4782 \cdot 10^{-10} x^2 + 1.1751 \cdot 10^{-10} x^3
\]

that must be implemented in the net recombination rate

\[
U_{\text{rad}} = C_{\text{rad}} \left( n_p - n_i^2 \right).
\]

Incomplete Ionization Model

According to [31] and for \(N_{D,0} < N_{D,\text{critic}}\), \(N_{A,0} < N_{A,\text{critic}}\):

\[
N_D^{\text{Boltz}} = \frac{N_{D,0}}{g_D e^{\frac{E_D}{kT}} + 1} \\
N_A^{\text{Boltz}} = \frac{N_{A,0}}{g_A e^{\frac{E_A}{kT}} + 1}
\]

or

\[
N_D^{\text{Fermi}} = \frac{N_{D,0}}{\frac{g_D}{\gamma_p} e^{\frac{E_D}{pT}} + 1} \\
N_A^{\text{Fermi}} = \frac{N_{A,0}}{\frac{g_A}{\gamma_p} e^{\frac{E_A}{pT}} + 1}
\]

where \(\gamma_n\) and \(\gamma_p\) are given by equations (2.77) and (2.78),

\[
\begin{align*}
n_1 &= N_C e^{-\frac{E_C - E_{D,0} - \alpha_D}{kT} \sqrt{N_{A,0} + N_{D,0}}} \\
p_1 &= N_C e^{-\frac{E_A - \alpha_A}{kT} \sqrt{N_{A,0} + N_{D,0}} - E_V}
\end{align*}
\]
Appendix A. Hg$_{0.7054}$Cd$_{0.2946}$Te Simulation Physical Parameters

$N_{D,0}$ and $N_{A,0}$ are the substitutional donor/acceptor concentrations, $E_{D,0}$ and $E_{A,0}$ the ionization energies and $g_{D,A}$ the degeneracy factors for the impurity levels. For transient processes the continuity equations account for the terms

$$\begin{align*}
\frac{\partial N_D}{\partial t} &= \sigma_D v_n \left[ \frac{n_1}{g_D} N_{D,0} - \left( n + \frac{n_1}{g_D} N_D \right) \right] \\
\frac{\partial N_A}{\partial t} &= \sigma_A v_p \left[ \frac{p_1}{g_A} N_{A,0} - \left( p + \frac{p_1}{g_A} N_A \right) \right]
\end{align*} \tag{A.23}$$

in which $v_n$ and $v_p$ are the thermal carrier speed and $\sigma_{D,A}$ the cross-sections. The HgCdTe parameters are:

\begin{align*}
E_{D,0} &= -1.89 \times 10^{-3} \text{ [eV]} \tag{A.24} \\
E_{A,0} &= 20.01 \times 10^{-3} \text{ [eV]} \tag{A.25} \\
\alpha_D &= \alpha_A = 0 \text{ [eV cm]} \tag{A.26} \\
g_D &= 2 \tag{A.27} \\
g_A &= 4 \tag{A.28} \\
\sigma_D &= \sigma_A = 10^{-12} \text{ [cm}^2/\text{s]} \tag{A.29} \\
N_{D,\text{critic}} &= N_{A,\text{critic}} = 10^{22} \text{ [cm}^{-3}] \tag{A.30}
\end{align*}
C++ Routine for the 1D Kane BTBT Model

```cpp
#include <iostream>
#include <cmath>
#include <cstdlib>
#include "PMIModels.h"

class Band2Band_MM : public PMI_Recombination {
  protected:
    double pi, q, m0, hbar;
  public:
    Band2Band_MM (const PMI_Environment& env);
    ~Band2Band_MM();
    void Compute_r(const double t, const double n, const double p,
                   const double nie, const double f, double& r);
    void Compute_drdt(const double t, const double n, const double p,
                      const double nie, const double f, double& drdt);
    void Compute_drdn(const double t, const double n, const double p,
                      const double nie, const double f, double& drdn);
    void Compute_drdp(const double t, const double n, const double p,
                      const double nie, const double f, double& drdp);
    void Compute_drdnie(const double t, const double n, const double p,
                        const double nie, const double f, double& drdnie);
    void Compute_drdf(const double t, const double n, const double p,
                      const double nie, const double f, double& drdf);
};
```
Band2Band_MM::
Band2Band_MM (const PMI_Environment& env) :
PMI_Recombination (env)
{};

Band2Band_MM::
~Band2Band_MM ()
{}

t void Band2Band_MM::
Compute_r (const double t, const double n, const double p,
const double nie, const double f, double& r)
{
    double mt, eg, eg300, A, B;
    double pi, q, m0, hbar;
    pi = 3.14159265358979;
    q = 1.602e-19;
    m0 = 9.109e-31;
    hbar = 1.054e-34;
    mt = 0.017 * m0;
    eg = 0.244 * q;
    eg300 = 0.2906 * q;
    A = -(1e-2) * sqrt(2*mt) * pow(q,2) / (4*pow(pi,3)*pow(hbar,2)*sqrt(eg));
    B = (1e-2) * pi * sqrt(mt*pow(eg300,3)) / (2*sqrt(2)*q*hbar);
    r = A * pow(f,2) * exp(-B / f);
}

t void Band2Band_MM::
Compute_drdf (const double t, const double n, const double p,
const double nie, const double f, double& drdf)
{
    double mt, eg, eg300, A, B;
    double pi, q, m0, hbar;
    pi = 3.14159265358979;
    q = 1.602e-19;
    m0 = 9.109e-31;
    hbar = 1.054e-34;
    mt = 0.017 * m0;
    eg = 0.244 * q;
    eg300 = 0.2906 * q;
    A = -(1e-2) * sqrt(2*mt) * pow(q,2) / (4*pow(pi,3)*pow(hbar,2)*sqrt(eg));
\[ B = (1e-2) \times \pi \times \sqrt{m t \times \text{pow}(e g 300, 3)} / (2 \times \sqrt{2} \times q \times h \text{bar}) \]
\[ r = A \times \text{pow}(f, 2) \times \exp(-B / f) \]
\[ \text{if } (r == 0) \{ \]
\[ \text{drdf} = 0.0; \]
\[ \} \text{ else } \{ \]
\[ \text{drdf} = A \times B \times \exp(-B / f) + 2 \times A \times \exp(-B / f) \times f; \]
\[ \} \]

PMI_Recombination* new_PMI_Recombination (const PMI_Environment& env)
{
  return new Band2Band_MM (env);
}
APPENDIX C

C++ Routine for the 2D Kane BTBT Model

#include <iostream>
#include <cmath>
#include <stdlib.h>
#include "PMIModels.h"

class Band2Band_MM : public PMI_Recombination {

protected:
    double pi, q, m0, hbar;

public:
    Band2Band_MM (const PMI_Environment& env);
    ~Band2Band_MM();

    void Compute_t (const double t, const double n, const double p,
                    const double nie, const double f, double& r);

    void Compute_drdt (const double t, const double n, const double p,
                       const double nie, const double f, double& drdt);

    void Compute_drdn (const double t, const double n, const double p,
                       const double nie, const double f, double& drdn);

    void Compute_drdp (const double t, const double n, const double p,
                       const double nie, const double f, double& drdp);

    void Compute_drdnie (const double t, const double n, const double p,
                         const double nie, const double f, double& drdnie);

    void Compute_drdf (const double t, const double n, const double p,
                       const double nie, const double f, double& drdf);
};
Appendix C.  C++ Routine for the 2D Kane BTBT Model

Band2Band_MM::
Band2Band_MM (const PMI_Environment& env) :
PMI_Recombination (env)
{};

Band2Band_MM::
~Band2Band_MM ()
{
}

void Band2Band_MM::
Compute_r (const double t, const double n, const double p,
const double nie, const double f, double& r)
{
double mt, eg, eg300, A, B;
double pi, q, m0, hbar;
const PMIBaseParam* Temp = ReadParameter ("Temp");
double T = *Temp;

pi = 3.14159265358979;
q = 1.602e-19;
m0 = 9.109e-31;
hbar = 1.054e-34;
eg300 = 0.283668 * q;

if (T == 85) {
    eg = 0.235537 * q;
} else if (T == 100) {
    eg = 0.238933 * q;
} else if (T == 109) {
    eg = 0.240974 * q;
} else if (T == 120) {
    eg = 0.243446 * q;
} else if (T == 140) {
    eg = 0.247902 * q;
} else if (T == 143) {
    eg = 0.248574 * q;
} else if (T == 160) {
    eg = 0.252360 * q;
} else if (T == 180) {
    eg = 0.256797 * q;
} else if (T == 200) {
    eg = 0.261219 * q;
}

if (T == 85) {
    mt = 0.0171517 * m0;
} else if (T == 100) {
    mt = 0.0173846 * m0;
} else if (T == 109) {
    mt = 0.0175227 * m0;
} else if (T == 120) {
    mt = 0.0176907 * m0;
} else if (T == 140) {
    mt = 0.0179956 * m0;
} else if (T == 143) {
    mt = 0.0180409 * m0;
} else if (T == 160) {
    mt = 0.0182975 * m0;
} else if (T == 180) {
    mt = 0.0185992 * m0;
}


} else if (T == 200) {
    mt = 0.0188990 * m0;
}

A = -(1e-2) * sqrt(2*mt) * pow(q,2) / (4*pow(pi,3)*pow(hbar,2)*sqrt(eg));
B = (1e-2) * pi * sqrt(mt*pow(eg300,3)) / (2*sqrt(2)*q*hbar);

if (f == 0) {
    r = 0.0;
} else {
    r = A * pow(f,2) * exp(-B / f);
}
}

void Band2Band_MM::
    Compute_drdt (const double t, const double n, const double p,
                   const double nie, const double f, double& drdt)
{
}

void Band2Band_MM::
    Compute_drdn (const double t, const double n, const double p,
                   const double nie, const double f, double& drdn)
{
}

void Band2Band_MM::
    Compute_drdp (const double t, const double n, const double p,
                   const double nie, const double f, double& drdp)
{
}

void Band2Band_MM::
    Compute_drdf (const double t, const double n, const double p,
                   const double nie, const double f, double& drdf)
{
    double mt, eg, eg300, A, B;
    double pi, q, m0, hbar;
    double r;

    const PMIBaseParam* Temp = ReadParameter("Temp");
    double T = *Temp;

    pi = 3.14159265358979;
    q = 1.602e-19;
    m0 = 9.109e-31;
    hbar = 1.054e-34;
    eg300 = 0.283668 * q;

    if (T == 85) {
        eg = 0.235537 * q;
    } else if (T == 100) {
        eg = 0.238933 * q;
    } else if (T == 109) {
        eg = 0.240974 * q;
    } else if (T == 120) {
        eg = 0.243446 * q;
    } else if (T == 140) {
        eg = 0.247902 * q;
    } else if (T == 143) {
Appendix C. C++ Routine for the 2D Kane BTBT Model

\[ e_g = 0.248574 \times q; \]
\[ e_g = 0.252360 \times q; \]
\[ e_g = 0.256797 \times q; \]
\[ e_g = 0.261219 \times q; \]

\[ e_g = 0.248574 \times q; \]
\[ e_g = 0.252360 \times q; \]
\[ e_g = 0.256797 \times q; \]
\[ e_g = 0.261219 \times q; \]

\[ \text{if} \ (T == 85) \{ \]
\[ \text{mt} = 0.0171517 \times m_0; \]
\[ \text{else if} \ (T == 100) \{ \]
\[ \text{mt} = 0.0173846 \times m_0; \]
\[ \text{else if} \ (T == 109) \{ \]
\[ \text{mt} = 0.0175227 \times m_0; \]
\[ \text{else if} \ (T == 120) \{ \]
\[ \text{mt} = 0.0176907 \times m_0; \]
\[ \text{else if} \ (T == 140) \{ \]
\[ \text{mt} = 0.0179956 \times m_0; \]
\[ \text{else if} \ (T == 143) \{ \]
\[ \text{mt} = 0.0180409 \times m_0; \]
\[ \text{else if} \ (T == 160) \{ \]
\[ \text{mt} = 0.0182975 \times m_0; \]
\[ \text{else if} \ (T == 180) \{ \]
\[ \text{mt} = 0.0185992 \times m_0; \]
\[ \text{else if} \ (T == 200) \{ \]
\[ \text{mt} = 0.0188990 \times m_0; \]
\[ \}
\[ A = -(1e-2) \times \sqrt{2 \times \text{mt}} \times \text{pow}(q, 2) / (4 \times \text{pow}(\pi, 3) \times \text{pow}(hbar, 2) \times \text{sqrt}(e_g)); \]
\[ B = (1e-2) \times \pi \times \sqrt{\text{mt} \times \text{pow}(e_g, 3)} / (2 \times \sqrt{2} \times q \times hbar); \]

\[ \text{if} \ (f == 0) \{ \]
\[ \quad r = 0.0; \]
\[ \} \text{ else } \{ \]
\[ \quad r = A \times \text{pow}(f, 2) \times \exp(-B / f); \]
\[ \} \]
\[ \text{if} \ (r == 0) \{ \]
\[ \quad drdf = 0.0; \]
\[ \} \text{ else } \{ \]
\[ \quad drdf = A \times B \times \exp(-B / f) + 2 \times A \times \exp(-B / f) \times f; \]
\[ \}

PMI_Recombination* new_PMI_Recombination (const PMI_Environment& env)
{ return new Band2Band_MM (env); }
}
C++ Routine for the 2D p-i-n BTBT Model

```cpp
#include <iostream>
#include <cmath>
#include <stdlib.h>
#include "PMIModels.h"

class Band2Band_pin_MM : public PMI_Recombination {

protected:
    double pi, q, m0, hbar;

public:
    Band2Band_pin_MM (const PMI_Environment& env);
    ~Band2Band_pin_MM();

    void Compute_r
    (const double t, const double n, const double p,
     const double nie, const double f, double& r);

    void Compute_drdt
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdt);

    void Compute_drdn
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdn);

    void Compute_drdp
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdp);

    void Compute_drdnie
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdnie);

    void Compute_drdf
    (const double t, const double n, const double p,
     const double nie, const double f, double& drdf);
};
```
Band2Band_pin_MM::
Band2Band_pin_MM (const PMI_Environment& env) :
  PMI_Recombination (env)
{;}

Band2Band_pin_MM::
~Band2Band_pin_MM ()
{
}

void Band2Band_pin_MM::
Compute_r (const double t, const double n, const double p,
  const double nie, const double f, double& r)
{
  double mt, eg, eg300, A, B, d;
  double pi, q, m0, hbar;
  const PMIBaseParam* Temp = ReadParameter ("Temp");
  double T = *Temp;

  pi = 3.14159265358979;
  q = 1.602e-19;
  m0 = 9.109e-31;
  hbar = 1.054e-34;
  eg300 = 0.283668 * q;
  d = < >;

  if (T == 85) {
    eg = 0.235537 * q;
  } else if (T == 100) {
    eg = 0.238933 * q;
  } else if (T == 109) {
    eg = 0.240974 * q;
  } else if (T == 120) {
    eg = 0.243446 * q;
  } else if (T == 140) {
    eg = 0.247902 * q;
  } else if (T == 143) {
    eg = 0.248574 * q;
  } else if (T == 160) {
    eg = 0.252360 * q;
  } else if (T == 180) {
    eg = 0.256797 * q;
  } else if (T == 200) {
    eg = 0.261219 * q;
  }

  if (T == 85) {
    mt = 0.0171517 * m0;
  } else if (T == 100) {
    mt = 0.0173846 * m0;
  } else if (T == 109) {
    mt = 0.0175227 * m0;
  } else if (T == 120) {
    mt = 0.0176907 * m0;
  } else if (T == 140) {
    mt = 0.0179956 * m0;
  } else if (T == 143) {
    mt = 0.0180409 * m0;
  } else if (T == 160) {
    mt = 0.0182975 * m0;
  } else if (T == 180) {
  
  
}
mt = 0.0185992 * m0;
} else if (T == 200) {
    mt = 0.0188990 * m0;
}

A = -(1e-2) * sqrt(2*mt) * pow(q,2) / (4*pow(pi,3)*pow(hbar,2)*sqrt(eg));
B = (1e-2) * pi * sqrt(mt*pow(eg300,3)) / (2*sqrt(2)*q*hbar);

if (f == 0) {
    r = 0.0;
} else {
    r = A * pow((1.24172e2),(d-1)) * pow(f,((d+1)*0.5)) * 
        exp( -B / (1.24172e2 * sqrt(f)) );
}

void Band2Band_pin_MM::
    Compute_drdt (const double t, const double n, const double p,
        const double nie, const double f, double& drdt)
{};

void Band2Band_pin_MM::
    Compute_drdf (const double t, const double n, const double p,
        const double nie, const double f, double& drdf)
{};

const PMIBaseParam* Temp = ReadParameter ("Temp");
double T = *Temp;

pi = 3.14159265358979;
q = 1.602e-19;
m0 = 9.109e-31;
hbar = 1.054e-34;
eg300 = 0.283668 * q;
d = < >;

if (T == 85) {
    eg = 0.235537 * q;
} else if (T == 100) {
    eg = 0.238933 * q;
} else if (T == 109) {
    eg = 0.240974 * q;
} else if (T == 120) {
    eg = 0.243446 * q;
} else if (T == 140) {
Appendix D. C++ Routine for the 2D p-i-n BTBT Model

\[
\begin{align*}
eg_e &= 0.247902 \times q; \\
& \text{else if (T == 143) \{} \\
& \quad \neg_e = 0.248574 \times q; \\
& \text{else if (T == 160) \{} \\
& \quad \neg_e = 0.252360 \times q; \\
& \text{else if (T == 180) \{} \\
& \quad \neg_e = 0.256797 \times q; \\
& \text{else if (T == 200) \{} \\
& \quad \neg_e = 0.261219 \times q; \\
& \text{\} }
\end{align*}
\]

\[
\begin{align*}
\text{mt} &= 0.0171517 \times m_0; \\
& \text{else if (T == 100) \{} \\
& \quad \text{mt} = 0.0173846 \times m_0; \\
& \text{else if (T == 109) \{} \\
& \quad \text{mt} = 0.0175227 \times m_0; \\
& \text{else if (T == 120) \{} \\
& \quad \text{mt} = 0.0176907 \times m_0; \\
& \text{else if (T == 140) \{} \\
& \quad \text{mt} = 0.0179956 \times m_0; \\
& \text{else if (T == 143) \{} \\
& \quad \text{mt} = 0.0180409 \times m_0; \\
& \text{else if (T == 160) \{} \\
& \quad \text{mt} = 0.0182975 \times m_0; \\
& \text{else if (T == 180) \{} \\
& \quad \text{mt} = 0.0185992 \times m_0; \\
& \text{else if (T == 200) \{} \\
& \quad \text{mt} = 0.0188990 \times m_0; \\
& \text{\} }
\end{align*}
\]

\[
\begin{align*}
A &= -\left(10^{-2}\right) \times \sqrt{2\times\text{mt}} \times \text{pow}(q,2) / \left(4\times\text{pow}(\pi,3)\times\text{pow}(\hbar,2)\times\sqrt{\neg_e}\right); \\
B &= \left(10^{-2}\right) \times \pi \times \sqrt{\text{mt}\times\text{pow}(\neg_e300,3)} / \left(2\times\sqrt{2}\times q\times\hbar\right); \\
& \text{if (f == 0) \{} \\
& \quad r = 0.0; \\
& \text{else \{} \\
& \quad r = A \times \text{.pow}((1.24172e2),(d-1)) \times \text{pow}(f,(d+1)*0.5)) \times \exp(-B / (1.24172e2 \times \sqrt{f})) ; \\
& \text{\} }
\end{align*}
\]

\[
\begin{align*}
& \text{if (r == 0) \{} \\
& \quad \text{drdf} = 0.0; \\
& \text{else \{} \\
& \quad \text{drdf} = A \times \text{pow}((1.24172e2),d) \times \exp(-B / (1.24172e2 \times \sqrt{f})) \times \text{pow}(f,((d+0.5)-1)) \times \left( (B \times 0.0000324282) + \left(\sqrt{f} \times (0.00402668 \times (1 + d))\right) \right) ; \\
& \text{\} }
\end{align*}
\]

\[
\text{PMI_Recombination* new_PMI_Recombination (const PMI_Environment& env)} \\
\text{\{} \\
\text{return new Band2Band_pin_MM (env);} \\
\text{\}}
\]