Master degree thesis
Nanotechnologies for ICTs
Dipartimento di Elettronica e Telecomunicazioni
Politecnico di Torino

Multiphysics simulation and design of silicon photonics Ge-on-Si lateral *pin* waveguide photodetector

Antonio Fronteddu
s287693

* * * * **

Supervisors
Prof. Michele Goano, Supervisor
Mr. Matteo G. C. Alasio, Co-supervisor
Dr. Alberto Tibaldi, Co-supervisor
Prof. Francesco Bertazzi, Co-supervisor
Prof. Giovanni Ghione, Co-supervisor

A.A. 2022/2023, October 16, 2023

Abstract

This work presents a 3D multiphysics model to investigate a lateral Ge-on-Si waveguide photodetector. The simulation is performed using commercial tools: Synopsys TCAD Sentaurus for the electrical part, implementing the drift-diffusion method, and Synopsys RSoft for the light propagation problem using the Finite-Difference-Time-Domain method (FDTD). The first part (chapters 1 and 2) introduces the context and reasons that prompted the start of this work. In the second part (chapter 3), the actual simulation is presented, aimed at reproducing the experimental results from the work of S. Lichke et al. (Nat. Photon 2021). The simulation results for the electro-optic frequency response show a good agreement with the measurements, with a relative mean error of 8 % for frequencies over 200 GHz.

Summary

In the last years, Silicon Photonics has become a promising solution for the ever-growing need for fast data exchange, removing the actual limitation of conventional electrical interconnections in terms of power consumption, bandwidth, and cross-talk, to keep pace with datacom and telecom needs, and with novel technologies and applications like Cloud services and AI training for machine learning. Silicon Photonics is a technology based on silicon for the optical transmission of information that allows complete integration with CMOS technology and silicon production processes. Thanks to this compatibility, this technology is leading toward the production of a single-chip platform composed of multiple devices for signal modulation (modulators) and detection (photodetectors). The capability of complete integration on a single chip is a key factor for scaling and production of more compact circuits, also decreasing the fabrication cost. Silicon has a low absorption coefficient in the bands used for telecom applications (up to the C-Band, centered at a wavelength of 1550 nm), making it a good material to fabricate waveguides and other passive components. On the other hand, silicon cannot be used for signal detection since it doesn't absorb light in the required wavelength range. Germanium, thanks to its capability to be epitaxially grown over silicon despite the significant lattice mismatch between the two materials, and its good optical properties in the wavelength range of interest, has gained increasing popularity in Silicon Photonics, permitting a higher versatility in applications than the use of silicon alone, such as the possibility to integrate photodetectors in the technology.

This work focuses on studying one of the main types of photodetectors used in Silicon Photonics, a lateral Ge-on-Si waveguide photodetector. The reference configuration and technology were inspired by the work of S.Lichke et al.[1], which was also used to compare our results with measured data. The device is simulated by performing a 3D multiphysics electro-optical simulation, employing commercial simulation tools from Synopsys. TCAD Sentaurus suite was used to solve the electrical problem employing a Drift-Diffusion (DD) model, and RSoft FullWAVE for the light propagation problem and the calculation of the optical generation rate, solving Maxwell's equations using the Finite Difference Time Domain (FDTD) method. This work shows the capability of the multiphysics model to predict the device

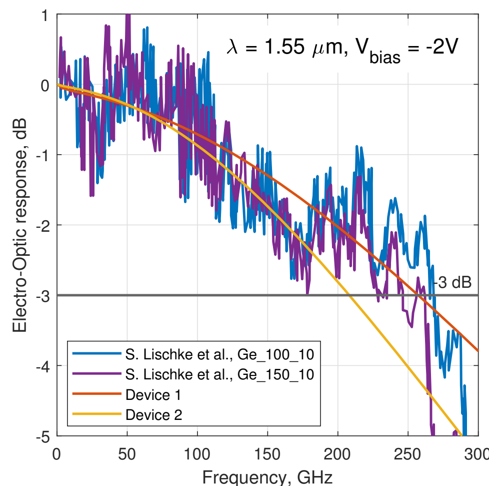


Figure 1: Simulation of the electro-optical frequency response compared to the experimental data. The numbers in the terms "Ge_100_10" and "Ge_150_10" define the thickness and length in nanometers and micrometers respectively of the germanium region for the device represented. *Device 1* and *Device 2* are the respective simulations for the two measurements.

behavior with complex geometries and sub-micron features even for bandwidth higher than 200 GHz, with a relative mean error of 8 %, as shown in figure 1. In this thesis, the critical points of the device are also analyzed in detail, such as the strong dependence of responsivity from material properties such as the absorption coefficient, the effect of the saturation velocity of carriers in the germanium region on the bandwidth (which was further explored outside the thesis scope in a separate work[2]), and the possible optimizations considering variations on the geometry and applied voltage bias.

This work is divided into three chapters and one appendix:

- Chapter 1 introduces the building blocks of Silicon Photonics and the main configuration used in Ge-on-Si waveguide photodetector, the lateral and the vertical configuration. Then, typical figures of merit are analyzed for the two configurations, showing their advantages and disadvantages in terms of production and applications.

- Chapter 2 describes the physics behind the simulation of the device. It starts by describing the electronic transport in semiconductors to introduce the Drift-Diffusion model and the small-signal analysis, then

[1] S. Lischke, A. Peczek, J. S. Morgan, K. Sun, D. Steckler, Y. Yamamoto, F. Korndörfer, C. Mai, S. Marschmeyer, M. Fraschke, A. Krüger, A. Beling, and L. Zimmermann, "Ultra-fast germanium photodiode with 3-dB bandwidth of 265 GHz," *Nature Photon.*, pp. 1–7, 2021.

it describes the Finite Difference Time Domain method for light propagation. It also presents the details of the main physical properties that have a significant impact on the simulation.

- Chapter 3 presents the simulation results, in terms of electro-optical bandwidth, dark current, and responsivity, comparing them with the measurements from the reference paper and highlighting the critical points and possible improvements.
- The Appendix focuses on the technical aspects of the simulation, such as the definition of the mesh and mole fraction profile. It also provides additional theory details about the simulation methods.

[2] M. G. C. Alasio, M. Zhu, A. Fronteddu, A. Cardinale, A. Ballarati, E. Bellotti, G. Ghione, A. Tibaldi, F. Bertazzi, M. Vallone, and M. Goano, "Modeling the electronic transport in FinFET-like lateral Ge-on-Si *pin* waveguide photodetectors for ultra-wide bandwidth applications," in *23rd International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD 2023)*, Torino, Italy, Sep. 2023, pp. 107–108.

Contents

1 Introduction 6
- 1.1 Research context . . . 6
- 1.2 Ge-on-Si Waveguide Photodetector (WPD) . . . 6
  - 1.2.1 Vertical WPD . . . 7
  - 1.2.2 Lateral WPD . . . 7
  - 1.2.3 Waveguide coupling . . . 8
  - 1.2.4 Electro-Optical properties and figures of merit . . . 8
- 1.3 Ultra-fast germanium photodiode . . . 19

2 Methods 23
- 2.1 Electronic transport in semiconductors . . . 23
  - 2.1.1 Carrier densities at equilibrium conditions . . . 23
  - 2.1.2 Carrier densities out of equilibrium conditions . . . 25
  - 2.1.3 Doped semiconductors . . . 26
  - 2.1.4 Degenerate semiconductors . . . 27
- 2.2 Drift-Diffusion transport model . . . 27
  - 2.2.1 Generation and Recombination rates . . . 29
  - 2.2.2 Mobility model . . . 31
- 2.3 Small-signal AC analysis . . . 34
  - 2.3.1 Small-signal Current density . . . 37
  - 2.3.2 Small-signal Transfer function . . . 38
- 2.4 Finite Difference Time Domain method . . . 39
  - 2.4.1 Boundary conditions . . . 42
  - 2.4.2 Simulation stability . . . 42
  - 2.4.3 Stop criteria . . . 43
  - 2.4.4 Optical generation rate . . . 43

3 Results 46
- 3.1 Simulation . . . 46
  - 3.1.1 Optical power propagation . . . 47
- 3.2 Dark current . . . 49
- 3.3 Photocurrent . . . 50
- 3.4 Electro-optic frequency response . . . 53

3.5 Simulation in O-Band 56
3.5.1 Waveguide optimization 57
3.5.2 Effects of the applied voltage bias 58
3.5.3 Effects of the Ge width 62
3.5.4 Effects of the carrier saturation velocity 65
3.5.5 Germanium shape 66

4 Conclusions 71

A Electronic band structure and effective mass approximation 73

B Heterojunction between Ge and Si 75

C Simulation mesh 77

Chapter 1

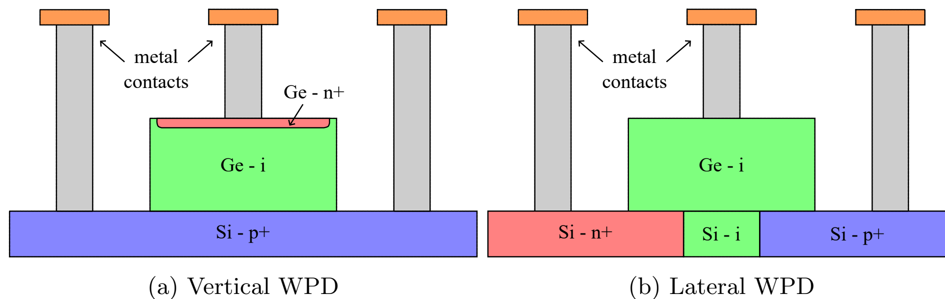
Introduction

1.1 Research context

One of the most active research topics in optoelectronics today is Silicon Photonics. This technology comprehends all electro-optic devices based on silicon, which has the advantage of being fully integrable with CMOS technology and relative production processes. Photodetectors are one example of the application of this technology. Silicon alone, however, doesn't have the ideal characteristics for optic sensing in telecom applications, given its inability to absorb light over 1.1 μm, which excludes the most used bands used for data transmission, which are the O-Band (1.31 μm) and C-Band (1.55 μm), the II and III windows of the optical fibers, corresponding to minimum fiber dispersion and absorption, respectively. The I window which is placed around 0.85 µm, exists only for historical reasons and is no longer used since it was not particularly good from the start, but rather the only one compatible with light sources available at the time (lasers based on GaAs alloys). It is, however, still used for low-cost and non-telecom applications. What permitted the success of silicon-based technology in the optoelectronics field for data transmission, in fact, was not silicon but germanium, which, thanks to its broader absorption profile, includes both O and C bands. More specifically, the interest in the use of germanium on silicon-based devices comes from the possibility of growing epitaxially thin films of germanium over a silicon substrate, even with the large lattice mismatch between the two materials in lattice constant of ~4.2 % [1, 2].

1.2 Ge-on-Si Waveguide Photodetector (WPD)

In this work, we focus on waveguide photodetectors (WPD), a specific class of *pin* photodiode made of silicon with a germanium absorber, connected to a silicon waveguide. The waveguide and the detector can be part of a bigger silicon photonics circuit grown on the same wafer, which can also contain



(a) Vertical WPD (b) Lateral WPD

Figure 1.1: Geometries of vertical and lateral configurations of a Si-on-Ge waveguide photodetector.

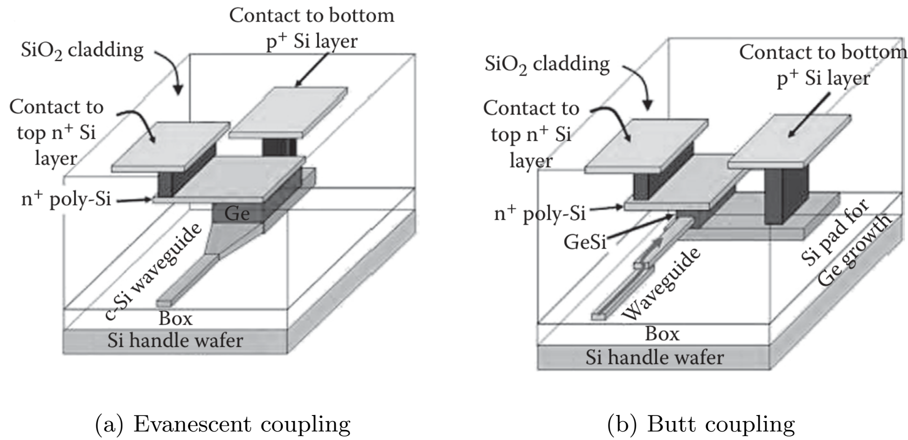
CMOS electronics. There are two main categories of WPD called lateral and vertical depending on the orientation of the *pin* junction, as shown in fig.1.1.

1.2.1 Vertical WPD

In the vertical WPD, a silicon waveguide is connected to a wider p-type doped silicon substrate over which a layer of germanium is epitaxially grown and patterned with etching. Then, the metal contacts are deposited over the silicon substrate and the germanium region. For the latter, before the deposition of the metal contact, implantation of arsenic is performed to create the n-type region of the *pin* junction and to obtain an ohmic contact between germanium and metal. The device is then encapsulated in a silicon oxide layer [3, 4]. The production processes of germanium, however, are not extensively researched and comprehended as much as for silicon, so it is often preferred to avoid too much processing of germanium, such as doping implantation and metal growth [5], which is one the reasons why a lateral configuration can be a good alternative.

1.2.2 Lateral WPD

In the lateral WPD case, there are many more variants in the design with respect to the vertical. Generally, a lateral WPD is composed of a silicon waveguide connected to a silicon substrate, where a layer of germanium is deposited. In this case, the *pin* junction is created horizontally. It could be made by doping the intrinsic germanium in both sides [6], or doping the silicon substrate, creating a double *pin* junction with the central intrinsic region of the substrate and with the germanium deposited over it [7]. Another particularly interesting method, is the one used in [8], where the germanium is patterned in a fin shape and doped silicon is deposited on both sides. This last design will be described in major detail further on since it is the main inspiration for the simulation work on which this thesis is based.



(a) Evanescent coupling (b) Butt coupling

Figure 1.2: Schematics of waveguide coupling methods for vertical Ge-on-Si WPD. Figures taken from [9].

1.2.3 Waveguide coupling

For both vertical and lateral WPD the waveguide was described as connected to the silicon substrate to propagate the light inside the device. This type of configuration is called *evanescent coupling*, that together with *butt coupling* are the most commonly used coupling methods [9]. Evanescent coupling (fig.1.2a) exploits the difference in refractive index between silicon oxide and germanium. The light is confined in the silicon waveguide because of the higher refractive index with respect to the surrounding oxide. When the light reaches the germanium, placed over the silicon, it is no longer well confined as before because of the increase in refractive index. Moreover, the energy gap of Ge is only 0.66 eV, low enough to have absorption for our wavelengths of interest (O-Band and C-Band). The evanescent field on the germanium is thus absorbed by the material gradually transferring the optical power from the waveguide to the detector. Butt coupling (fig.1.2b), instead, is more efficient because it directly couples the optical power into the Ge region, however, it requires a more complex and accurate design and production process to satisfy the mode-matching conditions between the waveguide and germanium region.

1.2.4 Electro-Optical properties and figures of merit[1]

The Ge-on-Si waveguide photodetector is a particular type of *pin* photodetector, which is, in turn, part of a bigger class of devices called photodiodes. A photodetector is a device used to convert a light signal into an electrical signal, by absorbing incident photons on the device with sufficient energy

[1] The formulations in this subsection are partially taken from [10].

($E_{ph} \geq E_g$) to generate electron-hole pairs. This process is called optical generation (see 2.2.1.1). The photogenerated electron-hole pairs are divided by the electric field induced by the reverse bias applied to the diode junction and collected into the contacts, making up the photogenerated current [10]. As for any diode-like device, the reverse bias generates a small current even in dark conditions (i.e. absence of a light source) which is the reverse saturation current that, however, in most cases, represents a negligible noise contribution to the current photogenerated by the input optical power. A *pin* photodiode is a double junction diode composed of a p-type region, an intrinsic region, and a n-type region. The intrinsic section is the depleted region of the device and also where the photons are absorbed and the photocurrent generated.

The total current generated by the photodiode can be modeled, without considering the optical power modulation, as:

$$I_{PD} = I_L - I_0 \left[exp\left(-\frac{V_{PD}}{\eta V_T} \right) - 1 \right] \tag{1.1}$$

where I_L represents the photocurrent, I_0 is the reverse saturation current, V_{PD} is the external reverse voltage bias, η is the ideality factor ($1 \leq \eta \leq 2$) and $V_T = k_B T/q$ is the thermal voltage, which is equal to 26 mV at a temperature $T = 300\,\mathrm{K}$. In reverse bias condition, $V_{PD}/\eta V_T \gg 1$, so we can approximate the expression as:

$$I_{PD} = I_L + I_d \tag{1.2}$$

where $I_d = I_0$ and represents the dark current.

1.2.4.1 Responsivity and Quantum efficiency

The photocurrent I_L can be calculated by integrating the optical generation rate over the device's active region's volume:

$$I_L = q \int_V G^{opt}(\underline{r}, P^{opt}) d\underline{r} \tag{1.3}$$

The optical power is absorbed inside a material with an exponential decay behavior, which, in a simplified one-dimensional view is written as:

$$P^{opt}(x) = P^{opt}(0)\, exp(-\alpha x) \tag{1.4}$$

where α is the absorption coefficient, a material property whose value is dependent on the impinging light wavelength. Differentiating (1.4) with respect to x, and defining a optical power density $\tilde{P}^{opt} = P^{opt}/A$, where A is the illuminated active area:

$$\frac{d\tilde{P}^{opt}}{dx} = -\alpha \tilde{P}^{opt}(x) = (\text{power absorbed in x direction})/V \tag{1.5}$$

Assuming a linear relation between current and optical power, and considering the generation of one electron-hole pair for every absorbed photon, we have:

$$\begin{aligned}\frac{(\text{Power absorbed})/V}{\text{Photon energy}} = \frac{\alpha \tilde{P}^{opt}}{E_{ph}} = \frac{(\text{Energy absorbed})/(t \cdot V)}{\text{Photon energy}} &= \\ = \frac{\text{Number of photons absorbed}}{(t \cdot V)} &= \\ = \frac{\text{Number of e-h pairs generated}}{(t \cdot V)} &= G^{opt}\end{aligned}$$

so the optical generation is equal to:

$$G^{opt} = \frac{\alpha \tilde{P}^{opt}}{E_{ph}} = \frac{\alpha \lambda}{hc} \tilde{P}^{opt} \tag{1.6}$$

The same relation is valid also for the extended three-dimensional case. Now, assuming that the number of collected carriers in the device contacts is equal to the e-h pairs we have:

$$\begin{aligned}&\frac{\text{Number of collected carriers}}{t} = \frac{I_L}{q} = \\ &= V \cdot \frac{\text{Number of e-h pairs generated}}{t \cdot V} = \\ &= V \cdot \frac{\text{Number of photons absorbed}}{t \cdot V} = \\ &= A \int_0^\infty G^{opt}(x) dx = A \int_0^\infty \frac{\alpha \lambda}{hc} \tilde{P}^{opt} dx \\ &= -\frac{A\lambda}{hc} \int_0^\infty \frac{d\tilde{P}^{opt}(x)}{dx} dx \approx \frac{\lambda}{hc} P^{opt}(0)\end{aligned}$$

resulting in:

$$\frac{I_L}{q} = \frac{\lambda}{hc} P^{opt}(0) \tag{1.7}$$

where $P^{opt}(0)$ is the incident power. The expression of the photocurrent then becomes:

$$I_L = \frac{q\lambda}{hc} P^{opt}(0) = \mathfrak{R} P^{opt}(0) \tag{1.8}$$

where $\mathfrak{R}$ is the *responsivity*, whose definition is the ratio between the photocurrent and the incident optical power:

$$\mathfrak{R} \triangleq \frac{I_L}{P^{opt}(0)} = \frac{q\lambda}{hc} = \frac{q}{E_{ph}} \tag{1.9}$$

The second and third passages in (1.9) obtained from the derivation of I_L are a best-case scenario where all incident photons are converted in e-h pairs

and collected to the contacts. The responsivity is dependent on the photon energy. It is zero for $E_{ph} < E_g$, since photons don't have enough energy to generate carriers, and reaches a maximum for $E_{ph} \approx E_g$, then it decreases for higher values of photon energy. The decrease can be explained as a reduction in the efficiency of the generation process: every photon, even with energy much higher than the energy gap generates only one electron-hole pair, so the excess of energy is lost in thermalization processes (phonon emission) instead of being used for the generation of other carriers.

If we take into consideration more realistic cases, it is necessary to introduce the definition of *internal quantum efficiency* and *external quantum efficiency*. The internal quantum efficiency is defined as:

$$\eta_Q = \frac{\text{Generated pairs}}{\text{Photons reaching the active region}} \tag{1.10}$$

whose value is typically $\eta_Q \approx 1$ $(0 \leq \eta_Q \leq 1)$. The internal quantum efficiency considers only the conversion process of photons in carriers in the active region, while the external quantum efficiency takes into consideration the total conversion efficiency of the device treating it as a black box in a similar way to the responsivity, to which is directly related:

$$\eta_\chi = \frac{\text{Collected pairs}}{\text{Incident photons}} = \frac{I_L/q}{P^{opt}/E_{ph}} = \frac{E_{ph}}{q}\mathfrak{R} \tag{1.11}$$

Typically, $\eta_\chi \leq 1$, in the absence of gain[2], moreover, it results to be always lower than η_Q because it considers all loss contributions in the device, including the ones that define the internal quantum efficiency $(\eta_\chi < \eta_Q)$. From the definition of external quantum efficiency, we can rewrite the definition of responsivity including this new term:

$$\mathfrak{R} = \eta_\chi \frac{q}{E_{ph}} = \eta_\chi \frac{q\lambda}{hc} \tag{1.12}$$

If we consider $\eta_\chi = 1$ we obtain again (1.9).

1.2.4.2 Photodetector frequency response

A photodetector is used to convert a light signal into an electrical signal. The light contains a high-frequency component that carries the information. The photodetector absorbs the light and generates a photocurrent that must retain the same frequency modulation to transfer the information correctly. This requires a stable behavior of the device in a large frequency range, also called bandwidth. If the optical signal frequency is too high, is possible

[2] there are other types of photodetectors, like avalanche photodiodes, that exploit certain physical phenomena to increase the collected photocurrent, resulting in an external quantum efficiency $\eta_\chi > 1$.

that the detector is not capable of following the same time variation of the light, resulting in deformations in the output electrical signal that could affect the information fidelity. A way to verify if a photodetector is capable of transferring information at a certain frequency correctly is to study the frequency behavior of its responsivity. Assuming that the device works linearly, or is linearized around an operating point (DC component), we can separate the DC component from the AC component (signal):

$$P^{opt}(t) = P^{opt}_{DC} + \delta P^{opt}(t) \tag{1.13a}$$

$$V(t) = V_{DC} + \delta V(t) \tag{1.13b}$$

$$I(t) = I_{DC} + \delta I(t) \tag{1.13c}$$

assuming a sinusoidal modulation of light, we have:

$$\delta P^{opt}(t) = Re\{P^{opt}_0\, e^{i\omega t}\} \tag{1.14a}$$

$$\delta V(t) = Re\{V_0\, e^{i\omega t}\} \tag{1.14b}$$

$$\delta I(t) = Re\{I_0\, e^{i\omega t}\} \tag{1.14c}$$

where ω is the angular modulation frequency of the light signal. The photodetector current is a function of the optical power and the applied voltage. If the signal perturbation is small enough, we can linearize around the operation point, obtaining:

$$I(t) = I_{DC} + \delta I(t) = I_{DC} + \left.\frac{\partial I(t)}{\partial P^{opt}}\right|_{DC} \delta P^{opt}(t) + \left.\frac{\partial I(t)}{\partial V}\right|_{DC} \delta V(t) \tag{1.15}$$

where the second and third terms are the contribution to the signal component of the photocurrent and dark current respectively:

$$\delta I(t) = \left.\frac{\partial I(t)}{\partial P^{opt}}\right|_{DC} \delta P^{opt}(t) + \left.\frac{\partial I(t)}{\partial V}\right|_{DC} \delta V(t) = \delta I_L(t) + \delta I_d(t) \tag{1.16}$$

which can be rewritten as:

$$\begin{aligned} \delta I(t) &= \delta I_L(t) + \delta I_d(t) = \\ &= Re\{\Re(\omega) P^{opt}_0\, e^{i\omega t}\} + Re\{Y_{PD}(\omega) V_0\, e^{i\omega t}\} \end{aligned} \tag{1.17}$$

where $\Re(\omega)$ is the small-signal responsivity, and $Y_{PD}(\omega)$ is the small-signal admittance. Rewriting the current expression is the frequency domain:

$$I(\omega) = Y_{PD}(\omega)V(\omega) + I_L(\omega) \tag{1.18}$$

where,

$$I_L(\omega) = \Re(\omega)P^{opt}(\omega) \tag{1.19}$$

the photocurrent $I_L(\omega)$ is linearly dependent on the optical power $P^{opt}(\omega)$ by means of the complex responsivity $\Re(\omega)$, which can be used to describe the

device frequency response. Considering the small-signal voltage amplitude $V_0 = 0$ (i.e. no modulation of the voltage), the small-signal current $\delta I(\omega)$ becomes equal to the single photogenerated current $\delta I(\omega) = \delta I_L(\omega)$ and it is called *short-circuit photocurrent* (i.e. the photocurrent of the detector whose small-signal load is a short).

The first step is to define a *normalized responsivity* $\mathfrak{r}(\omega)$ with respect to the static case:

$$\frac{I_L(\omega)}{I_L(0)} = \frac{\mathfrak{R}(\omega)}{\mathfrak{R}(0)} \frac{P^{opt}(\omega)}{P^{opt}(0)} = \mathfrak{r}(\omega) \frac{P^{opt}(\omega)}{P^{opt}(0)} \tag{1.20}$$

assuming a constant value of optical power, independently from the modulation frequency, we have:

$$\mathfrak{r}(\omega) = \frac{I_L(\omega)}{I_L(0)} = \frac{\mathfrak{R}(\omega)}{\mathfrak{R}(0)} \rightarrow |\mathfrak{r}(\omega)|_{dB} = 20 \log_{10} \left| \frac{\mathfrak{R}(\omega)}{\mathfrak{R}(0)} \right| \tag{1.21}$$

The resulting frequency response typically is similar to a low-pass filter. The total bandwidth of the device is defined by the frequency value for which the normalized responsivity drops to $-3\,\mathrm{dB}$:

$$\begin{gathered} |\mathfrak{r}(\omega_{3dB})|_{dB} = -3 \rightarrow 20 \log_{10} \left| \frac{\mathfrak{R}(\omega_{3dB})}{\mathfrak{R}(0)} \right| = -3 \\ \mathfrak{R}(f_{3dB}) = \frac{1}{\sqrt{2}} \mathfrak{R}(0) \end{gathered} \tag{1.22}$$

where f_{3dB} is called *intrinsic cut-off frequency*, because it is calculated considering only the short-circuit current, and it is so independent from the detector loading. The intrinsic cut-off frequency, as the name suggests, takes into account only intrinsic phenomena that can influence the final bandwidth of the device, such as transit time and avalanche build-up delay (only in avalanche photodetectors). However, extrinsic contributions, such as load impedance and parasitic contributions in the device contacts can influence greatly the bandwidth. The effect of resistance and capacitive elements both intrinsic and extrinsic is called *RC cut-off*. In *pin* photodiodes *transit time* and *RC cut-off* are the main limiting factors of the device bandwidth.

1.2.4.3 Transit time limited bandwidth in a *pin* PD

To study the limitation in transit time of the bandwidth of a *pin* photodetector, is necessary to start with an analysis of the photocurrent. In a semiconductor device, the current behavior can be described using the *Drift-Diffusion model* (see 2.2). Starting from the continuity equations, assuming a constant electric field, and neglecting the diffusion currents:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \underline{J}_n - U_n \tag{1.23a}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \cdot \underline{J}_p - U_p \tag{1.23b}$$

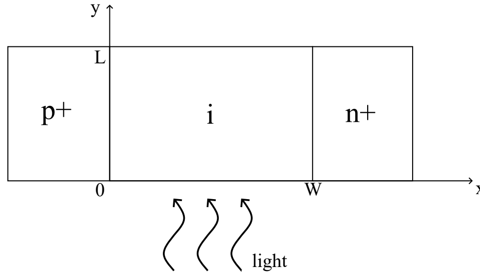


Figure 1.3: *pin* photodiode illuminated from the bottom.

where $U_n = U_p = G^{opt}(x,t)$, and the constitutive relations are:

$$\underline{J}_n = \underline{J}_{n,drift} = qv_{n,sat}n \tag{1.24a}$$

$$\underline{J}_p = \underline{J}_{p,drift} = qv_{p,sat}p \tag{1.24b}$$

Assuming that the device has no variation in the z-direction, we can consider a 2D equivalent geometry, as shown in figure 1.3.

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\left(\frac{\partial J_p}{\partial x} + \frac{\partial J_p}{\partial y}\right) + G_0(y,t) \tag{1.25a}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\left(\frac{\partial J_n}{\partial x} + \frac{\partial J_n}{\partial y}\right) + G_0(y,t) \tag{1.25b}$$

where the derivatives of the current density in function of the y-direction can be considered equal to zero assuming a negligible variation of the current in that direction, which is generally true given the high electric field, due to the applied reverse bias, directed almost exclusively in the x-direction.

The photodetector is illuminated from the bottom of the intrinsic region. The optical power is assumed to be harmonic with angular frequency ω:

$$P^{opt}(t) = P^{opt}(\omega)\, e^{i\omega t} \tag{1.26}$$

substituting the constitutive relations (1.24) in the continuity equations (1.23), and working in the frequency domain, we have:

$$i\omega p(x) = G_0(y) - v_{p,sat}\frac{\partial p(x)}{\partial x} \tag{1.27a}$$

$$i\omega n(x) = G_0(y) + v_{n,sat}\frac{\partial n(x)}{\partial x} \tag{1.27b}$$

The boundary conditions at the junctions p^+i and in^+ are, respectively, $n(0) = 0$ and $p(W) = 0$, since in those positions the total minority carrier densities are zero.

To obtain a solution for the two differential equations (1.27), we can start by considering the homogeneous associate equations, which, for the holes, is:

$$i\omega p'(x) = -v_{p,sat}\frac{\partial p'(x)}{\partial x} \tag{1.28}$$

Since the equation is linear in x, the trial solution is in exponential form:

$$p'(x) = A\, e^{ikx} \tag{1.29}$$

substituting (1.29) in (1.28):

$$\begin{aligned} i\omega A\, e^{ikx} &= -v_{p,sat}\frac{\partial A\, e^{ikx}}{\partial x} \\ i\omega A\, e^{ikx} &= -v_{p,sat}(ik)A\, e^{ikx} \\ \omega &= -v_{p,sat}k \rightarrow k = -\frac{\omega}{v_{p,sat}} \end{aligned}$$

The solution of the complete equation (1.27a) can be expressed as:

$$p(x) = p_1(x) + p_2(x) = A\, e^{-\frac{i\omega x}{v_{p,sat}}} + B \tag{1.30}$$

The values of the constants A and B can be found imposing the boundary conditions defined before:

$$p(W) = 0 = A\, e^{-\frac{i\omega W}{v_{p,sat}}} + B \rightarrow B = -A\, e^{-\frac{i\omega W}{v_{p,sat}}}$$

susbtituing in (1.27a) we can find also the expression for A:

$$\begin{aligned} i\omega p(W) = 0 = G_0(y) - v_{p,sat}\left.\frac{\partial p(x)}{\partial x}\right|_{x=0} \Rightarrow \\ \Rightarrow G_0(y) = v_{p,sat}\left[A\left(-i\frac{\omega}{v_{p,sat}}\right)\, e^{-\frac{i\omega W}{v_{p,sat}}}\right] \\ A = -\frac{G_0(y)}{i\omega}\, e^{\frac{i\omega W}{v_{p,sat}}} \end{aligned}$$

substituting A in B we have:

$$B = \frac{G_0(y)}{i\omega}$$

and finally substituting in the complete trial solution (1.30):

$$\begin{aligned} p(x,y) &= -\frac{G_0(y)}{i\omega}\, e^{-\frac{i\omega(x-W)}{v_{p,sat}}} + \frac{G_0(y)}{i\omega} = \\ &= \frac{G_0(y)}{i\omega}\left(1 - e^{-\frac{i\omega(x-W)}{v_{p,sat}}}\right) \end{aligned} \tag{1.31}$$

repeating the same passages for the electron continuity equation we have:

$$n(x,y) = \frac{G_0(y)}{i\omega}\left(1 - e^{\frac{i\omega x}{v_{n,sat}}}\right) \tag{1.32}$$

substituting both solutions in the current densities:

$$J_p(x,y) = qv_{p,sat}p(x,y) = \frac{qv_{p,sat}G_0(y)}{i\omega}\left[1 - e^{-\frac{i\omega(x-W)}{v_{p,sat}}}\right] \tag{1.33a}$$

$$J_n(x,y) = qv_{n,sat}n(x,y) = \frac{qv_{n,sat}G_0(y)}{i\omega}\left[1 - e^{\frac{i\omega x}{v_{n,sat}}}\right] \tag{1.33b}$$

The total current density is the sum of J_p, J_n and the displacement current density J_{disp}:

$$J_{tot}(\omega) = J_p + J_n + J_{disp} = J_p + J_n + i\omega\epsilon_s E(x,\omega) \tag{1.34}$$

where ϵ_s is the semiconductor electrical permittivity and $E(x,\omega)$ is the electric field.

Integrating both sides of the total current density equation from $x = 0$ to $x = W$:

$$\int_0^W J_{tot}(\omega)dx = WJ_{tot}(\omega) = \int_0^W \left[J_p(x,y) + J_n(x,y) + i\omega\epsilon_s E(x,\omega)\right]dx \Rightarrow$$

$$\Rightarrow J_{tot}(\omega) = \frac{1}{W}\frac{qv_{p,sat}G_0(y)}{i\omega}\left[W - \frac{1 - e^{\frac{i\omega W}{v_{p,sat}}}}{i\frac{\omega}{v_{p,sat}}}\right] +$$

$$+\frac{1}{W}\frac{qv_{n,sat}G_0(y)}{i\omega}\left[W + \frac{1 - e^{\frac{i\omega W}{v_{n,sat}}}}{i\frac{\omega}{v_{n,sat}}}\right] +$$

$$+i\omega\epsilon_s[V(0) - V(W)]$$

introducing the definition of the electron and hole transit times:

$$\tau_{tr,p} = \frac{W}{v_{p,sat}} \qquad \tau_{tr,n} = \frac{W}{v_{n,sat}} \tag{1.35}$$

then substituting in the expression, and dividing and multiplying for W, we have:

$$J_{tot}(\omega) = qWG_0(y)\left[\frac{1}{i\omega\tau_{tr,p}} - \frac{1 - e^{i\omega\tau_{tr,p}}}{\omega^2\tau_{tr,p}^2} + \frac{1}{i\omega\tau_{tr,n}} - \frac{1 - e^{i\omega\tau_{tr,n}}}{\omega^2\tau_{tr,n}^2}\right]$$

$$+i\omega\frac{\epsilon_s}{W}V_A(\omega)$$

integrating also in the y-direction, between $y = 0$ and $y = L$, and considering a distribution of the optical generation rate as

$$G_0(y) = G_0(0)\, e^{-\alpha y} \tag{1.36}$$

it becomes:

$$J_{tot}(\omega) = \frac{qWG_0(y)}{L\alpha}\left(1 - e^{-\alpha L}\right)\left[\frac{1}{i\omega\tau_{tr,p}} - \frac{1 - e^{i\omega\tau_{tr,p}}}{\omega^2\tau_{tr,p}^2} + \right.$$
$$\left. + \frac{1}{i\omega\tau_{tr,n}} - \frac{1 - e^{i\omega\tau_{tr,n}}}{\omega^2\tau_{tr,n}^2}\right] + i\omega\frac{\epsilon_s}{W}V_A(\omega)$$

considering that $G_0(0,\omega)$ is expressed as:

$$G_0(0,\omega) = \eta_Q(1 - R)\frac{\alpha P^{opt}(\omega)}{A_{opt}}\left(\frac{\lambda}{hc}\right) \tag{1.37}$$

where R is the reflectance of the light on the illuminated area $A_{opt} = W \cdot H$, with H being the device height (z-direction). The term $(1 - R)$ then represents the transmitted light inside the device. Substituting the expression in $I_{tot}(\omega) = A \cdot J_{tot}(\omega)$, where $A = H \cdot L$:

$$I_{tot}(\omega) = q\eta_Q(1 - R)P^{opt}(\omega)\left(\frac{\lambda}{hc}\right)\left(1 - e^{-\alpha L}\right)\left[\frac{1}{i\omega\tau_{tr,p}} - \frac{1 - e^{i\omega\tau_{tr,p}}}{\omega^2\tau_{tr,p}^2} + \right.$$
$$\left. + \frac{1}{i\omega\tau_{tr,n}} - \frac{1 - e^{i\omega\tau_{tr,n}}}{\omega^2\tau_{tr,n}^2}\right] + i\omega C V_A(\omega) = -I_L(\omega) + I_{disp}$$

with $C = \frac{\epsilon_s}{W}A$ being the intrinsic device capacitance given by the geometry.

The equation for $\omega \to 0$ reduces to:

$$I_{tot}(0) = -I_L(0) = -q\eta_Q(1 - R)P^{opt}(0)\left(\frac{\lambda}{hc}\right)\left(1 - e^{-\alpha L}\right) \tag{1.38}$$

whose expression can be easily obtained by expanding the exponential terms inside the square brackets in Taylor's series at second order.

Finally, the normalized responsivity can be defined as:

$$\mathfrak{r}(\omega) = \frac{I_L(\omega)}{I_L(0)} = \left[\frac{1}{i\omega\tau_{tr,p}} - \frac{1 - e^{i\omega\tau_{tr,p}}}{\omega^2\tau_{tr,p}^2} + \frac{1}{i\omega\tau_{tr,n}} - \frac{1 - e^{i\omega\tau_{tr,n}}}{\omega^2\tau_{tr,n}^2}\right] \tag{1.39}$$

1.2.4.4 RC limited bandwidth in a *pin* PD

The *pin* photodiode's bandwidth is also limited by capacitive effects, both intrinsic and extrinsic. The RC limit can be evaluated from the equivalent circuit of the photodiode, represented in figure 1.4, where C_j is the junction capacitance, R_D is the diode resistance and I_L is the photocurrent, derived in the previous section (1.39); C_p is the parasitic capacitance and R_s is the parasitic series resistance. These last two elements are extrinsic parameters due to the contacts between the diode and the rest of the circuit, generally an amplifier, here represented as a load resistance $R_L = 50\,\Omega$.

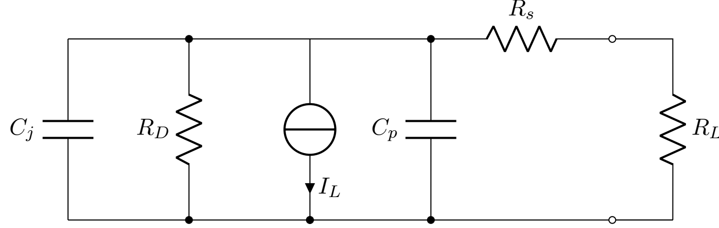


Figure 1.4: Equivalent circuit of a *pin* photodiode.

In general, $R_D \gg R_s, R_L$ so the equivalent resistance of the circuit seen by the current source is $R = R_D||(R_s + R_L) \approx R_s + R_L$, while the equivalent capacitance is $C = C_j + C_p$. The cut-off frequency can be then approximated as:

$$f_{3dB,RC} \approx \frac{1}{2\pi RC} \tag{1.40}$$

Finally, the total cut-off frequency of the device, considering both the RC and transit time limitations can be calculated using the following approximated expression:

$$f_{3dB} \approx \frac{1}{\sqrt{f_{3dB,RC}^{-2} + f_{3dB,tr}^{-2}}} \tag{1.41}$$

1.2.4.5 Bandwidth-efficiency trade-off in a *pin* PD and WPD

The RC and transit time limits are affected by the geometry characteristics of the photodiode. Increasing the thickness of the active region the transit time increases which results in a lower transit time limit for the bandwidth. At the same time, a thicker depleted region results in lower intrinsic capacitance and thus a higher RC limit. The opposite would happen decreasing the thickness. A variation in the area A, instead, would influence only the capacitance, to which it is inversely proportional since it would not change the travel distance of the photogenerated carriers. In brief $f_{3dB,RC} \propto W/A$ while $f_{3dB,tr} \propto 1/W$. Figure 1.5 shows the total cut-off frequency calculated with (1.41) in function of the active region width. Usually, a *pin* photodiode is illuminated laterally through one of the doped regions, meaning that the optical propagation direction is parallel to the carrier's motion. This means that the quantum efficiency is proportional to the width of the intrinsic region $\eta_\chi \propto W$. Increasing the thickness would improve the efficiency but the device would become transit time limited. Decreasing the thickness, instead, would remove the limit of transit time but decrease the device efficiency. Moreover, if the device is too thin it would become also RC-limited. This device limitation is called *bandwidth-efficiency trade-off.* In the previous section, however, *pin* photodiode was illuminated from the bottom, directly in the intrinsic region. In this case, the optical propagation direction

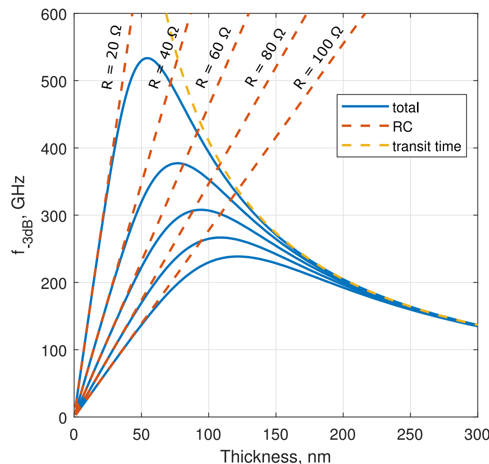


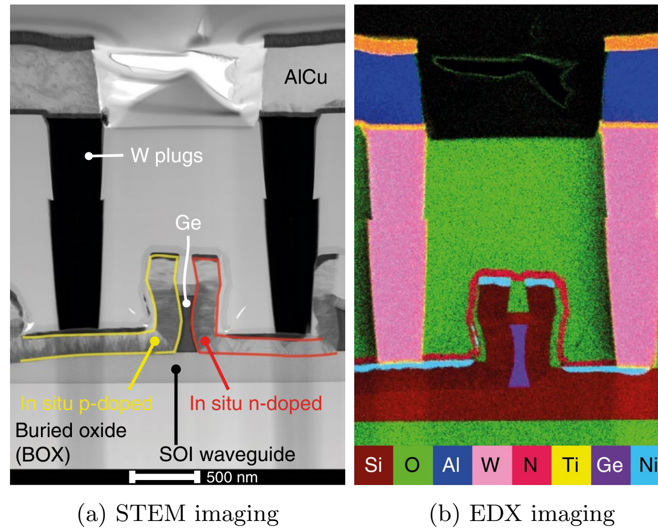
Figure 1.5: *pin* photodiode cut-off frequency in function of the active region width. The represented curves are obtained considering a height $H = 400\,\mathrm{nm}$, a length $L = 10\,\mu\mathrm{m}$ and a saturation velocity for electrons and holes $v_{sat} = 0.743 \times 10^7\,\mathrm{cm\,s^{-1}}$.

and the direction of the carriers are orthogonal, so the quantum efficiency is no longer proportional to the thickness W, but instead is proportional to the device length $\eta_\chi \propto L$. This is the configuration used in the *waveguide pin photodiode* (WPD) where the device is connected to a waveguide directly on the intrinsic region and can be made long as much as needed to reach the highest efficiency possible. The intrinsic region is also made thin to avoid the transit time limit and the RC limit is contained by limiting the height of the device to avoid large values of the junction area ($A = H \cdot L$).

1.3 Ultra-fast germanium photodiode

At the end of 2021 S. Lischke et al. presented a novel lateral waveguide photodetector [8] with evanescent coupling, capable of reaching a $-3\,\mathrm{dB}$ bandwidth of 265 GHz. One of the focus points of this thesis work is to simulate this same device and reproduce the measurement results from this paper. After validating the simulation framework, it will be possible to use it to investigate further the device in ways that are not feasible from simple production and measurement.

The geometry of the device can be observed from the cross-section in figure 1.6. A waveguide is connected to a silicon substrate where germanium is grown epitaxially. Then the germanium is etched in a fin shape, and doped silicon is deposited on both sides to create a horizontal *pin* junction. The silicon is deposited already doped to reduce at minimum the contamination of germanium with doping since it must remain intrinsic. Contacts



(a) STEM imaging (b) EDX imaging

Figure 1.6: Figure (a) shows a Scanning Transmission Electron Microscopy (STEM) image of the cross-section of the device. Figure (b) shows the same cross-section analyzed using Energy-Dispersive X-ray spectroscopy (EDX) to show the material composition. Both figures are taken from [8].

in tungsten are then grown on both doped silicon regions. all the device is then covered with silicon oxide, leaving only the contacts and the end of the silicon waveguide uncovered. The novelty of this device can be found in the unusual geometry of the germanium absorber, characterized by a biconcave shape, which has the advantage of reducing the transit time for photogenerated carriers on the central region of the fin, where the electric field has the peak intensity.

Two different geometries are presented, one with a germanium absorber width of ∼60 nm in the narrowest point of the biconcave shape, and with a calculated mean width of 100 nm, denominated "Ge_100_10"; another, with ∼100 nm in the narrowest point and mean value of 150 nm called "Ge_150_10". The height of both absorbers is 400 nm, while the length is 10 µm to assure the complete absorption of the light.

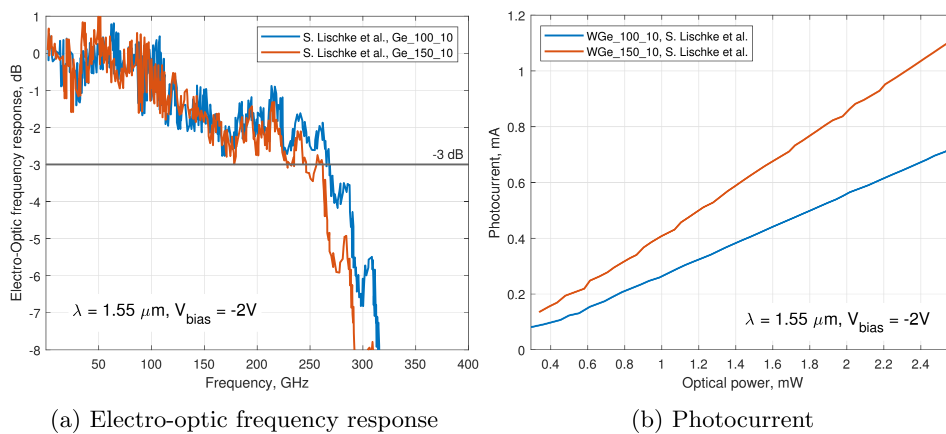
The first device, Ge_100_10, reached a measured −3 dB bandwidth of 265 GHz for a photocurrent of 1 mA. The second device, Ge_150_10, instead, had a −3 dB bandwidth of 240 GHz for the same photocurrent. The difference is due to the higher width of the germanium region in Ge_150_10 since an increased thickness of the absorber, means a longer travel distance for the photogenerated carriers, increasing the transit time. This means that the device Ge_150_10 is transit-time limited. The different thickness, however, influences also the internal responsivity, which is 0.3 A W^{-1} for Ge_100_10, and 0.45 A W^{-1} for Ge_150_10. The smaller volume of the absorber in Ge_100_10 limits the generated photocurrent from the input optical

	Ge_100_10	Ge_150_10
Nominal width (nm)	100	150
f_{-3dB} (GHz)	265	240
Dark current (A)	$\leq 1 \times 10^{-7}$	$\leq 2 \times 10^{-7}$
Photocurrent (A)	1×10^{-3}	1×10^{-3}
Internal responsivity ($\mathrm{A\,W^{-1}}$)	0.3	0.45
Internal quantum efficiency	0.24	0.36
External responsivity ($\mathrm{A\,W^{-1}}$)	0.12	0.2
External quantum efficiency	0.1	0.16
Bandwidth-efficiency product (GHz)	63	86

Table 1.1: Measurements from the paper for devices Ge_100_10 and Ge_150_10 for an applied voltage $V_{bias} = -2\,\mathrm{V}$.

power, reducing the responsivity. There is, then, a trade-off between the two devices, depending on the germanium region geometry. All measurements were performed with an input optical power at a wavelength of 1.55 µm and a bias of −2 V. Table 1.1 summarizes all the important measurements on the two devices from the paper to be compared with the simulation.

The table also shows the external responsivity and external quantum efficiency, but, unfortunately, they are calculated with respect to the input optical power in the optical fiber used to feed the silicon waveguide, so the main contribution of losses in the external values are from the grating coupler, and not from the photodetector. The bandwidth-efficiency product is, in fact, calculated with respect to the internal values, since, otherwise, they would have almost any dependence on the photodetector performances. This also means that we can simulate only the internal responsivity and quantum efficiency and not the external ones. Figure 1.7 shows the measured electro-optic frequency response and photocurrent for the two devices.



(a) Electro-optic frequency response

(b) Photocurrent

Figure 1.7: Figure (a) shows the electro-optic frequency response of the two devices. Figure (b) shows, instead, the photocurrent. The smaller size of the absorber in Ge_100_10 is reflected in the bigger bandwidth, but also in the smaller photocurrent at equal optical power. Both figures are reproduced using the data from [8].

Chapter 2

Methods

2.1 Electronic transport in semiconductors

To study the electronic transport in a semiconductor is necessary to know the distribution in energy of the carriers [11]. The electrons and holes can have only specific values of energy defined by the band structure of the semiconductor. The conductance and valence band in the energy band structure are defined by the respective dispersion relations calculated in the effective mass approximation (see appendix A). In the following section, the definitions of the density of electrons and holes are presented, which are then used for the description and application of the drift-diffusion method in section 2.2.

2.1.1 Carrier densities at equilibrium conditions

Electrons and holes are distributed in energy in the allowed states described by the dispersion relation. Their distribution inside the allowed states, instead, depends on their fermionic nature, so their distributions follow the Fermi-Dirac statistics. The distribution of the states in energy is defined by the Density of States (DOS) which for a 3D (bulk) semiconductor, in the effective mass approximation, is subdivided into two contributions, the density of states of the conduction band $N_c(E)$ and the density of states of the valence band $N_v(E)$:

$$N_c(E) = \frac{4\pi}{h^3}(2m_n^*)^{3/2}\sqrt{E - E_c} \tag{2.1a}$$

$$N_v(E) = \frac{4\pi}{h^3}(2m_p^*)^{3/2}\sqrt{E_v - E} \tag{2.1b}$$

As mentioned before, the electrons and holes are distributed in the allowed states following the Fermi-Dirac statistics, whose formulation is

$$f_n(E) = \frac{1}{1 + exp\left(\frac{E-E_F}{k_BT}\right)} \tag{2.2a}$$

$$f_p(E) = \frac{1}{1 + exp\left(\frac{E_F-E}{k_BT}\right)} \tag{2.2b}$$

for electrons and holes respectively. E_F is the *Fermi level* and it is an index of the total level occupation, representing the point in energy where half of the states are occupied. For a semiconductor in equilibrium conditions, the Fermi level is a constant value. For an intrinsic semiconductor in equilibrium conditions, the Fermi level is placed in the center of the band gap. If the material is doped, the fermi level is shifted from the center, nearer to the conduction band for n-type doping and neared to the valence band for p-type. When the Fermi level is contained inside the band gap, we talk about a *nondegenerate semiconductor*. In this condition is possible to apply the Boltzmann approximation of the Fermi-Dirac statistics:

$$f_n(E) \underset{E \gg E_F}{\approx} exp\left(\frac{E_F - E}{k_BT}\right) \quad f_p(E) \underset{E \ll E_F}{\approx} exp\left(\frac{E - E_F}{k_BT}\right) \tag{2.3}$$

This approximation is valid, however, only if the distance between E and E_F is larger than a few K_BT units. When the Fermi level is near the conduction or valence band, like in the case of a strongly doped semiconductor, the Boltzmann approximation starts to be no more valid and there will be a notable difference with the Fermi distribution. If, instead, we consider a case where the Fermi level is inside one of the two bands, then we talk about a *degenerate semicondutor*, and the Boltzamm approximation cannot be used anymore.

As explained above, the density of states describes the available states for electrons and holes, while the Fermi-Dirac statistic describes how the states are occupied. The product of these two functions gives the total distribution in energy of the carriers. To calculate the electrons and holes densities is sufficient to integrate the product with respect to the energy:

$$n = \int_{E_c}^{\infty} N_c(E) f_n(E) dE = N_c\, exp\left(\frac{E_F - E_c}{k_BT}\right) \tag{2.4a}$$

$$p = \int_{-\infty}^{E_v} N_v(E) f_p(E) dE = N_v\, exp\left(\frac{E_v - E_F}{k_BT}\right) \tag{2.4b}$$

where N_c and N_v are now the *effective densities of states* and are defined as:

$$N_c = 2\frac{(2\pi m_n^* k_BT)^{3/2}}{h^3} \quad N_v = 2\frac{(2\pi m_p^* k_BT)^{3/2}}{h^3} \tag{2.5}$$

The product of n and p, in equilibrium condition, is independent from the fermi level, and is equal to the square of the intrinsic concentration:

$$np = n_i^2 \tag{2.6}$$

this relation is called *mass action law* and is also valid for doped semiconductors.

2.1.2 Carrier densities out of equilibrium conditions

A semiconductor is defined as out of equilibrium when the action of an external cause modifies the electron and hole distributions from the ones at equilibrium. In equilibrium conditions, the carrier population must follow the mass action law, so $np - n_i^2 = 0$. Out of equilibrium, this difference is no more zero, because of generation and recombination mechanisms (see 2.2.1). If the cause of the deviation from the equilibrium condition is maintained stable in time, like in the case of an applied voltage, the system will result to be stationary and can be studied with a similar statistics to the one used for the equilibrium case, called the *quasi-Fermi distribution*. In the quasi-Fermi distribution, the Fermi level is subdivided into two different levels, E_{Fn} and E_{Fp}, which account for the electron and hole populations separately. Aside from this, the expressions are equal to the ones in the equilibrium case:

$$f_n(E, E_{Fn}) = \frac{1}{1 + exp\left(\frac{E - E_{Fn}}{k_B T}\right)} \underset{E \gg E_{Fn}}{\approx} exp\left(\frac{E_{Fn} - E}{k_B T}\right) \tag{2.7a}$$

$$f_p(E, E_{Fp}) = \frac{1}{1 + exp\left(\frac{E_{Fp} - E}{k_B T}\right)} \underset{E \ll E_{Fp}}{\approx} exp\left(\frac{E - E_{Fp}}{k_B T}\right) \tag{2.7b}$$

The expressions of the carrier densities within Boltzmann's approximation become:

$$n = N_c\, exp\left(\frac{E_{Fn} - E_c}{k_B T}\right) \tag{2.8a}$$

$$p = N_v\, exp\left(\frac{E_v - E_{Fp}}{k_B T}\right) \tag{2.8b}$$

the mass action law can also be rewritten with the new definitions of electron and hole densities:

$$np = n_i^2\, exp\left(\frac{E_{Fn} - E_{Fp}}{k_B T}\right) \tag{2.9}$$

in equilibrium conditions $E_{Fn} = E_{Fp} = E_F$ reducing the equation to the original one. From the mass action law we can distinguish two possible cases:

$$np > n_i^2 \quad \text{for } E_{Fn} > E_{Fp} \text{ (Carrier injection)} \tag{2.10a}$$

$$np < n_i^2 \quad \text{for } E_{Fn} < E_{Fp} \text{ (Carrier depletion)} \tag{2.10b}$$

2.1.3 Doped semiconductors

Doping a semiconductor means adding donor or acceptor elements (v or III group elements) to the semiconductor to increase the electrons or holes population respectively. A donor type (n-type) doping is defined by a donor density N_D that is added to the already present n density. Equivalently, for an acceptor type (p-type) there is an acceptor density added to the p density. In the case of silicon, the typical doping elements for n-type doping are Arsenic (As) and Phosphorus (P), while Boron (B) is used in p-type doping. The addition of electrons and holes by means of doping is possible because the doping elements added to the semiconductor have an energy level near the conduction band, for n-type doping, or valence band, for p-type doping, permitting an easy passage of the additional electron of holes to the respective band. The probability of transition, however, depends on the distance between the donor level E_D or acceptor level E_A and the conduction or valence band respectively, which is of a few k_BT units. This means that the temperature influences the transition, and in fact, we have complete ionization of the dopant only for temperatures that reach room temperature.

2.1.3.1 Incomplete ionization model

The total donor or acceptor density that is actually ionized is determined by the *incomplete ionization model*, based on the Fermi-Dirac distribution [12]:

$$N_D^+ = \frac{N_D}{1 + g_D\, exp\left(\frac{E_{Fn}-E_D}{k_BT}\right)}, \text{ for } N_D < N_{D,crit} \tag{2.11a}$$

$$N_A^- = \frac{N_D}{1 + g_A\, exp\left(\frac{E_A-E_{Fp}}{k_BT}\right)}, \text{ for } N_A < N_{A,crit} \tag{2.11b}$$

where g_D and g_A are the degeneracy factors for the impurity levels, while $N_{D,crit}$ and $N_{A,crit}$ are the critic values of doping over which is assumed total ionization. Table 2.1 shows the default parameters value for bulk silicon.

The electron and hole density have to be recalculated with the addition of the doping contribution. Using the expression of the mass action law $np = n_i^2$ and the charge neutrality condition $n + N_A^- = p + N_D^+$, is possible to obtain the equations for n and p:

$$n = \frac{1}{2}\left[N_D^+ - N_A^- + \sqrt{(N_D^+ - N_A^-)^2 + 4n_i^2}\right] \tag{2.12a}$$

$$p = \frac{1}{2}\left[N_A^- - N_D^+ + \sqrt{(N_A^- - N_D^+)^2 + 4n_i^2}\right] \tag{2.12b}$$

Symbol	donor species			
	As	P	Sb	N
E_D (eV)	0.054	0.045	0.039	0.045
g_D	2	2	2	2
$N_{D,crit}$ (cm^{-1})	1×10^{22}	1×10^{22}	1×10^{22}	1×10^{22}
	acceptor species			
	B	Al	In	
E_A (eV)	0.045	0.045	0.160	
g_A	4	4	4	
$N_{A,crit}$ (cm^{-1})	1×10^{22}	1×10^{22}	1×10^{22}	

Table 2.1: Parameters table for incomplete ionization model. The values shown are characteristic of bulk silicon and the default ones used by Synopsys Sentaurus.

2.1.4 Degenerate semiconductors

Until now the Fermi level has always been considered inside the energy gap. Doping can shift the Fermi level to higher or lower energies, but even with very high doping is not possible to shift the level outside the energy gap, since the probability of ionization of the dopant atoms is reduced when E_D is under the Fermi level for electrons or over E_A for holes, until the point where the Fermi levels stops right before crossing the conduction or valence band. However, even if it is not possible in equilibrium conditions, it is out of equilibrium, like for example, the injection of carriers from illuminating the semiconductor. For a degenerate semiconductor, the Boltzmann approximation is not more valid and instead is necessary to use the Fermi-Dirac statistics. This requires the use of the Fermi integral for the calculation of the electron and hole densities:

$$n = \frac{2}{\sqrt{\pi}} N_c \mathcal{F}_{1/2} \left(\frac{E_{Fn} - E_c}{k_B T} \right) \tag{2.13a}$$

$$p = \frac{2}{\sqrt{\pi}} N_v \mathcal{F}_{1/2} \left(\frac{E_v - E_{Fp}}{k_B T} \right) \tag{2.13b}$$

The Fermi-Dirac integral can be solved only by a numerical approach.

2.2 Drift-Diffusion transport model

The model used to describe the electronic transport in the simulation is the Drift-Diffusion model, a simplified macroscopic transport model derived from the semi-classical Boltzmann's transport equation [13]. The model

is valid in steady-state conditions and models the current contribution of electrons and holes in two terms, *drift current* and *diffusion current*:

$$\underline{J}_n = \underline{J}_{n,drift} + \underline{J}_{n,diff} = qn\mu_n\underline{\mathcal{E}} + qD_n\nabla n \tag{2.14a}$$

$$\underline{J}_p = \underline{J}_{p,drift} + \underline{J}_{p,diff} = qp\mu_p\underline{\mathcal{E}} - qD_p\nabla p \tag{2.14b}$$

The drift current represents the current component generated by the applied electric field on the material. It depends, for electrons, on the electron density n, the electron mobility μ_n, and the external electric field $\underline{\mathcal{E}}$. Where, for holes, the first two are substituted with the hole density p and hole mobility μ_p. The diffusion current, instead, is due to the natural behavior of an ensemble of particles with a non-uniform density in space. The particles in the denser region tend to move towards the regions with fewer particles until they reach an equilibrium for the entire domain. This phenomenon is described by Fick's law, which relates the flux of particles in a direction with the concentration gradient through a constant called diffusion coefficient [14]. In the case of carriers the flux is represented by the current densities $\underline{J}_{n,diff}$ and $\underline{J}_{p,diff}$, the concentrations n and p, and the diffusion coefficients D_n and D_p, defined by Einstein's relations:

$$D_n = \frac{k_b T}{q}\mu_n = V_T\mu_n \quad D_p = \frac{k_b T}{q}\mu_p = V_T\mu_p \tag{2.15}$$

where $V_T = k_B T/q \approx 26\,\mathrm{meV}$ is the thermal voltage. From the charge conservation principle, the continuity equations for electrons and holes can be obtained, which define their behavior in function of space and time inside a semiconductor. The equations are defined as the equivalence of the carrier variations in time, with the variation of the carrier fluxes in space, minus the net contribution of generation and recombination phenomena inside the material:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \underline{J}_n - U_n \tag{2.16a}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \underline{J}_p - U_p \tag{2.16b}$$

where $U_n = R_n - G_n$ and $U_p = R_p - G_p$ are the net recombination rates, with R being the recombination and G the generation.

From (2.14) we know that the current density doesn't depend only on the carrier density n and p, but also on the electric field $\underline{\mathcal{E}}$. But, from Gauss' law, we also know that a variation of the electric field in space is due to a change in the charge density distribution ρ, which comprehends the carrier densities and the fixed charges from doping in the material.

$$\nabla \cdot \underline{D} = \nabla \cdot (\epsilon\underline{\mathcal{E}}) = \rho \quad \rho = q(p - n + N_D^+ - N_A^-) \tag{2.17}$$

Putting these two equations together and considering that the electric field is equal to

$$\underline{\mathcal{E}} = -\nabla\phi \tag{2.18}$$

we find Poissons' equation, which relates electrostatic potential directly to the charge density:

$$\nabla \cdot (-\epsilon\nabla\phi) = \rho \tag{2.19}$$

The continuity equations (2.16) with Poisson's equation (2.19) and the definitions of the current densities (2.14) make up the *Drift-Diffusion transport model*, summarized here:

$$\begin{aligned} \frac{\partial n}{\partial t} &= \frac{1}{q}\nabla \cdot \underline{J}_n - U_n \\ \frac{\partial p}{\partial t} &= -\frac{1}{q}\nabla \cdot \underline{J}_p - U_p \\ \nabla \cdot (-\epsilon\nabla\phi) &= \rho \end{aligned} \tag{2.20}$$

this system of equations gives an approximate but effective prediction of the electrical behavior of a semiconductor material or more complex devices [11, 15].

2.2.1 Generation and Recombination rates

The generation and recombination rates are the terms that represent the interaction of the electrons and holes in the conduction band with the other bands and energy levels. There are multiple phenomena that are represented as generation and recombination. They can be subdivided into *intrinsic* processes, further subdivided in *thermal*, *radiative* and *Auger*, present even in an ideal crystal, and *extrinsic* processes, defined by the *Shockley-Read-Hall* (SRH) model and representing the generation and recombination due to defects and impurities in the material. In the following, the processes and relative models implemented in the simulation are presented.

The thermal process model is used to describe recombination and generation in case of the presence of models that consider multiple valence and conduction bands, to describe the thermalization processes, or to approximate all non-radiative phenomena with a single model. In our case, however, we simulate the various processes separately, so the thermal model is not implemented.

2.2.1.1 Radiative process

The radiative process describes the interaction between photons and carriers, and is responsible for the generation of the photocurrent in the photodetector. In the case of recombination, an electron is demoted to the valence band and releases a photon with energy equal to the energy difference between

the two bands ($\sim E_g$). The generation is the opposite, requiring an electron with at least an energy equal to the energy gap. The recombination process is the one responsible for the absorption of the light in the germanium, and the generation of the photocurrent. If seen from the perspective of carriers it is called optical generation rate. The intrinsic dependence on the energy gap makes it a material-dependent process. This dependence is contained in the absorption profile of the material where the process happens. The model implemented is:

$$U^{rad} = r^{opt}(np - n_i^2) \tag{2.21}$$

where r^{opt} is the optical recombination coefficient, expressed in cm^3s^{-1}.

2.2.1.2 Auger process

The Auger process is a three-carrier process. It involves an electron, a hole, and another electron or hole. Depending on the type of third carrier considered, it can be an electron-assisted process or a hole-assisted process. In the case of an electron-assisted recombination process, an electron in the conduction band recombines with a hole, and the difference in energy is given to a third electron in the conduction band in the form of kinetic energy. In the case of the electron-assisted generation process, it is the third electron that gives the energy necessary to promote an electron from the valence band to the conduction band. The equation that implements the Auger net recombination is the following:

$$U^{Au} = (r_{cn}^{Au}n + r_{cp}^{Au}p)(np - n_i^2) \tag{2.22}$$

where r_{cn}^{Au} and r_{cp}^{Au} are the Auger recombination coefficients for electrons and holes whose unit of measure is cm^6s^{-1}.

2.2.1.3 Shockley-Read-Hall process

The Shockley-Read-Hall (SRH) generation and recombination processes model the interaction with trap states due to impurities, including doping, and lattice defects in the material. The trap states have an energy level inside the energy gap that acts as a mid-point between the conduction band and the valence band, assisting thermal generation and recombination, in fact, these phenomena are also called *trap-assisted processes*. The SRH model implemented is the following:

$$U^{SRH} = \frac{np - n_i^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} \tag{2.23}$$

where τ_p and τ_n are the lifetimes of carriers for SRH, which are proportional to the trap density N_t by means of two coefficients of proportionality r_{cp}^{SRH}

and r_{cn}^{SRH}:

$$\tau_p = \frac{1}{r_{cp}^{SRH} N_t} \quad \tau_n = \frac{1}{r_{cn}^{SRH} N_t} \tag{2.24}$$

and n_1 and p_1 are the unitary degeneration factors:

$$n_1 = n_i\, exp\left(\frac{E_{Fi} - E_t}{k_B T}\right) \quad p_1 = n_i\, exp\left(-\frac{E_{Fi} - E_t}{k_B T}\right) \tag{2.25}$$

where E_t is the trap energy with respect to the valence band.

2.2.1.4 Avalanche process

When the electric field overcomes a certain threshold value, it can accelerate carriers to the point where their interaction with the lattice can cause ionization by impact. This is especially true in depletion regions, where the electric field is very high. If the length of the depletion region is greater than the mean free path, then the carriers accelerate enough to cause impact ionization, which can be represented as a generation term:

$$G^{ava} = \frac{1}{q}\Big(\alpha_n |\underline{J}_n| + \alpha_p |\underline{J}_p|\Big) \tag{2.26}$$

where α_n and α_p are the ionization coefficients for electrons and holes, defined as the reciprocal of the carrier mean free path. The ionization coefficients depend on the material properties and can be modelized with different models. This simulation implements the *Okuto-Crowell model*:

$$\alpha(F) \;=\; a \cdot \Big(1 + c(T - T_0)\Big) F^{\gamma}\; exp\left[-\left(\frac{b\,[1 + d(T - T_0)]}{F}\right)^{\delta}\right] \tag{2.27}$$

where T_0 is the room temperature (300 K) and F is the driving field, a term that represents the electric field calculated in a way that improves numerical convergence. The remaining terms are fitting parameters dependent on the material. Table 2.2 contains the typical values for bulk silicon.

2.2.2 Mobility model

The mobility is the proportionality factor that defines the drift velocity of a carrier consequent to the application of an electric field to the material:

$$\underline{v}_{drift,e} = \mu_e \underline{\mathcal{E}} \quad \underline{v}_{drift,h} = \mu_h \underline{\mathcal{E}} \tag{2.28}$$

2.2.2.1 Constant model

The simpler model for mobility would be to use a constant value, but this would not consider the decrease of mobility due to the introduction of doping or the dependence on the electric field which is necessary to correctly

Symbol	Electrons	Holes	Unit
a	0.426	0.243	V^{-1}
b	4.81×10^5	6.53×10^5	$V\,cm^{-1}$
c	3.05×10^{-4}	5.35×10^{-4}	K^{-1}
d	6.86×10^{-4}	5.67×10^{-4}	K^{-1}
γ	1	1	
δ	2	2	

Table 2.2: Parameters table for Okuto-Crowell model. The values shown are characteristic of bulk silicon and the default ones used by Synopsys Sentaurus.

Symbol	Electrons	Holes	Unit
μ_L	1417	470.5	$cm^2\,V^{-1}\,s^{-1}$
ζ	2.5	2.5	

Table 2.3: Parameters table for Constant mobility model. The values shown are characteristic of bulk silicon.

simulate the electronic transport in a highly depleted region like the intrinsic section of a *pin* device, where the electric field is very high resulting, experimentally, in a saturation of the velocity, possible only with a degradation of the mobility. The constant model's only dependence is from the temperature:

$$\mu_{const} = \mu_L \left(\frac{T}{300\,\mathrm{K}} \right)^{-\zeta} \tag{2.29}$$

where μ_L is the value of mobility measured at low-field and room temperature ($T_0 = 300\,\mathrm{K}$), ζ instead is a fitting parameter depending on the material. For this work, all simulations are done at room temperature, so the value of constant mobility stays equal to μ_L. In table 2.3 are shown the typical values used for silicon.

2.2.2.2 Masetti model

The model used to include the dependence from doping concentration is the Masetti model [16]:

$$\mu_{dop} = \mu_{min1}\, exp\left(-\frac{P_c}{N_A + N_D} \right) + \frac{\mu_{const} - \mu_{min2}}{1 + ((N_A + N_D)/C_r)^{\alpha}} - \frac{\mu_1}{1 + (C_s/(N_A + N_D))^{\beta}} \tag{2.30}$$

Symbol	Electrons	Holes	Unit
μ_{min1}	52.2	44.9	$cm^2 V^{-1} s^{-1}$
μ_{min2}	52.2	0	$cm^2 V^{-1} s^{-1}$
μ_1	43.4	29.0	$cm^2 V^{-1} s^{-1}$
P_c	0	9.23×10^{16}	cm^{-3}
C_r	9.68×10^{16}	2.23×10^{17}	cm^{-3}
C_s	3.43×10^{20}	6.10×10^{20}	cm^{-3}
α	0.680	0.719	
β	2.0	2.0	

Table 2.4: Parameters table for Masetti mobility model. The values shown are characteristic of bulk silicon and the default ones used by Synopsys Sentaurus.

where μ_{const} is the value of the constant model and N_A and N_D are respectively the acceptors and donor concentrations, the other terms, instead, are fitting parameters characteristic of the material. In table 2.4, reproduced from [12], are shown the default values used by Synopsys Sentaurus, which are the ones used for bulk silicon.

2.2.2.3 Extended Canali model

The Canali model describes the dependence on the electric field on mobility [17]. The mobility starts from a maximum value that decreases with the increase of the applied electric field. This causes a saturation of the drift velocity (v_{sat}) to values around $1 \times 10^7\,\mathrm{cm\,s^{-1}}$ (for silicon). The model's equation is the following:

$$\mu_{field}(F) = \frac{(\alpha + 1)\mu_{low}}{\alpha + \left[1 + \left(\frac{(\alpha+1)\mu_{low}F}{v_{sat}}\right)^{\beta}\right]^{1/\beta}} \tag{2.31}$$

where μ_{low} refers to the low-field mobility, which corresponds to the constant mobility in the constant model, and β is a fitting parameter dependent on temperature:

$$\beta = \beta_0 \left(\frac{T}{300\,\mathrm{K}}\right)^{\beta_{exp}} \tag{2.32}$$

The driving field F is the model used to consider the electric field dependence. Different models can be used for different applications and to improve numerical convergence. The remaining terms are fitting parameters dependent on the material. Table 2.5 presents the typical values for bulk silicon.

In this work, both the doping dependence model (Masetti model) and the high field saturation model (Canali model) are used for the simulation.

Symbol	Electrons	Holes	Unit
β_0	1.109	1.213	
β_{exp}	0.66	0.17	
α	0	0	

Table 2.5: Parameters table for Extended Canali mobility model. The values shown are characteristic of bulk silicon.

Matthiessen's rule is used to combine multiple mobility models:

$$\frac{1}{\mu} = \frac{1}{\mu_{dop}} + \frac{1}{\mu_{field}} \tag{2.33}$$

2.3 Small-signal AC analysis

Small signal analysis is a method to study the dynamic behavior of a device. It considers a small perturbation of a source term around the static operating point. The system can be then linearized with respect to the perturbation, simplifying the problem from a nonlinear system solvable only using numerical methods, to a linear system that can be solved analytically. The AC analysis starts from the Drift-Diffusion model reported here in a more generalized form [12]:

$$\begin{aligned} \nabla \cdot (\epsilon \nabla \phi + \underline{P}) &= -q\left(p - n + N_D^+ - N_A^-\right) - \rho_{trap} \\ \frac{1}{q}\nabla \cdot \underline{J}_n - U_n &= \frac{\partial n}{\partial t} \\ -\frac{1}{q}\nabla \cdot \underline{J}_p - U_p &= \frac{\partial p}{\partial t} \end{aligned} \tag{2.34}$$

where ρ_{trap} is the density of trap states in the material. The equations can be rewritten as:

$$\begin{aligned} \nabla \cdot (\epsilon \nabla \phi + \underline{P}) + q\left(p - n + N_D^+ - N_A^-\right) + \rho_{trap} &= 0 \\ \frac{1}{q}\nabla \cdot \underline{J}_n - U_n &= \frac{\partial n}{\partial t} \\ -\frac{1}{q}\nabla \cdot \underline{J}_p - U_p &= \frac{\partial p}{\partial t} \end{aligned} \tag{2.35}$$

Discretizing the equations in the three-dimensional mesh we can write:

$$\underline{F}_{\underline{\phi},i}(\underline{\phi}, \underline{n}, \underline{p}) = \underline{0} \tag{2.36a}$$

$$\underline{F}_{\underline{n},i}(\underline{\phi}, \underline{n}, \underline{p}) = \underline{\dot{G}}_{\underline{n},i}(\underline{n}) \tag{2.36b}$$

$$\underline{F}_{\underline{p},i}(\underline{\phi}, \underline{n}, \underline{p}) = \underline{\dot{G}}_{\underline{p},i}(\underline{p}) \tag{2.36c}$$

and

$$\underline{\phi} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_j \\ \vdots \\ \phi_N \end{bmatrix} \quad \underline{n} = \begin{bmatrix} n_1 \\ \vdots \\ n_j \\ \vdots \\ n_N \end{bmatrix} \quad \underline{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_j \\ \vdots \\ p_N \end{bmatrix} \tag{2.37}$$

where $i, j = 1, ..., N$, with N being the total number of mesh points. The system (2.36) is then a $3N \times 3N$ matrix, with 3N equations and 3N unknowns. The mesh points can be subdivided into internal points for $i \in [1, N-m]$, and boundary points for $i \in (N-m, N]$. The boundary points contain the boundary conditions, but, for now, they will be neglected in the formulation and will be added later.

The harmonic source of the perturbation can be a voltage, for *electrical small signal analysis*, or the optical generation rate for the *electro-optical small signal analysis*. In both cases, the term is divided into the static operating point (DC component) and the perturbating harmonic component (AC component):

$$\underline{V}_{tot} = \underline{V}_{DC} + \delta\underline{V} = \underline{V}_{DC} + \underline{V}\ exp(-i\omega) \tag{2.38a}$$

$$\underline{G}^{opt}_{tot} = \underline{G}^{opt}_{DC} + \delta\underline{G}^{opt} = \underline{G}^{opt}_{DC} + \underline{G}^{opt}\ exp(-i\omega) \tag{2.38b}$$

The AC component influences also the electrostatic potential and carriers distributions, which now have a perturbation with the same harmonic behavior:

$$\underline{\zeta}_{tot} = \underline{\zeta}_{DC} + \underline{\zeta}\ exp(-i\omega) \tag{2.39}$$

where $\underline{\zeta}$ represents $\underline{\phi}$, $\underline{n}$, and $\underline{p}$. Given that, for the definition of small signal, the harmonic component is infinitesimal with respect to the static component, we can develop the equations in (2.36) in Taylor's series. Defining, for simplicity, the following vector:

$$\underline{u} = \begin{bmatrix} \underline{\phi} \\ \underline{n} \\ \underline{p} \end{bmatrix} \tag{2.40}$$

The Taylor development of Poisson's equation is:

$$\underline{F}_{\phi}(\underline{u}) = \underline{F}_{\phi}(\underline{u}_{DC} + \delta\underline{u}) = \underline{F}_{\phi}(\underline{u}_{DC}) + \underline{\underline{J}}|_{DC}\ \delta\underline{u} \tag{2.41}$$

The constant term is equal to zero because it corresponds to the static case, whose equation is $F_{\underline{\phi}}(\underline{u}_{DC}) = \underline{0}$, canceling the DC contribution[1]. $\underline{\underline{J}}|_{DC}$, instead, is the Jacobian matrix evaluated in the operating point.

For the continuity equations, the passages are the same, but with a slight complication given by the $\underline{G}$ terms. Considering the electron's continuity

equation (2.36b), it is necessary to rewrite the equation taking all the terms on the left-hand side before linearizing.

$$\underline{F}_{\underline{n},i}(\underline{\phi}, \underline{n}, \underline{p}) - \dot{\underline{G}}_{\underline{n},i}(\underline{n}) = 0 \tag{2.42}$$

which is then redefined as:

$$\underline{\tilde{F}}_{\underline{n},i}(\underline{\phi}, \underline{n}, \underline{p}) = 0 \tag{2.43}$$

so now expanding in the Taylor series at first order:

$$\underline{\tilde{F}}_{\underline{n}}(\underline{u}) = \underline{\tilde{F}}_{\underline{n}}(\underline{u}_{DC} + \delta\underline{u}) = \underline{\tilde{F}}_{\underline{n}}(\underline{u}_{DC}) + \underline{\underline{J}}|_{DC}\, \delta\underline{u} \tag{2.44}$$

The same passages are valid also for the hole continuity equation:

$$\underline{\tilde{F}}_{\underline{p}}(\underline{u}) = \underline{\tilde{F}}_{\underline{p}}(\underline{u}_{DC} + \delta\underline{u}) = \underline{\tilde{F}}_{\underline{p}}(\underline{u}_{DC}) + \underline{\underline{J}}|_{DC}\, \delta\underline{u} \tag{2.45}$$

The static term can be considered equal to zero for the same reason as the Poisson's equation case.

The equations (2.41, 2.44, 2.45) are written with a compact and simple notation, which, anyway, does not clarify the structure of the result obtained. Expanding the notation of the jacobian and the solution vector $\underline{u}$ we have, for a certain mesh node i:

$$F_{\phi,i}(\phi_1 + \delta\phi_1, \dots, n_1 + \delta n_1, \dots, p_1 + \delta p_1, \dots) = 0 \tag{2.46a}$$

$$\tilde{F}_{n,i}(\phi_1 + \delta\phi_1, \dots, n_1 + \delta n_1, \dots, p_1 + \delta p_1, \dots) = 0 \tag{2.46b}$$

$$\tilde{F}_{p,i}(\phi_1 + \delta\phi_1, \dots, n_1 + \delta n_1, \dots, p_1 + \delta p_1, \dots) = 0 \tag{2.46c}$$

linearizing the equations and removing the static terms that, for now, are considered equal to zero:

$$\begin{aligned}
&\frac{\partial F_{\phi_i}}{\partial \phi_1}\delta\phi_1 + \cdots + \frac{\partial F_{\phi_i}}{\partial n_1}\delta n_1 + \cdots + \frac{\partial F_{\phi_i}}{\partial p_1}\delta p_1 + \cdots = 0\\
&\frac{\partial \tilde{F}_{n_i}}{\partial \phi_1}\delta\phi_1 + \cdots + \frac{\partial \tilde{F}_{n_i}}{\partial n_1}\delta n_1 + \cdots + \frac{\partial \tilde{F}_{n_i}}{\partial p_1}\delta p_1 + \cdots = 0\\
&\frac{\partial \tilde{F}_{p_i}}{\partial \phi_1}\delta\phi_1 + \cdots + \frac{\partial \tilde{F}_{p_i}}{\partial n_1}\delta n_1 + \cdots + \frac{\partial \tilde{F}_{p_i}}{\partial p_1}\delta p_1 + \cdots = 0
\end{aligned} \tag{2.47}$$

The 3 equations of the i-th mesh point have become a sum of partial derivatives with respect to the 3N variables. They can be rewritten as a sum of matrices, where every matrix is a 3×3 matrix composed of the terms

[1] Technically the static terms in the boundary could be not zero because of the boundary condition, but, for now, they are not considered in the formulation.

referred to the j-th variables. here we represent the sum of matrices for the i-th mesh node:

$$\sum_j \begin{bmatrix} \frac{\partial F_{\phi_i}}{\partial \phi_j} & \frac{\partial F_{\phi_i}}{\partial n_j} & \frac{\partial F_{\phi_i}}{\partial p_j} \\ \frac{\partial F_{n_i}}{\partial \phi_j} & \frac{\partial F_{n_i}}{\partial n_j} - i\omega \frac{\partial G_{n_i}}{\partial n_j} & \frac{\partial F_{n_i}}{\partial p_j} \\ \frac{\partial F_{p_i}}{\partial \phi_j} & \frac{\partial F_{p_i}}{\partial n_j} & \frac{\partial F_{p_i}}{\partial p_j} - i\omega \frac{\partial G_{p_i}}{\partial p_j} \end{bmatrix}_{DC} \begin{bmatrix} \tilde{\phi}_j \\ \tilde{n}_j \\ \tilde{p}_j \end{bmatrix} = \underline{0} \tag{2.48}$$

where the $-i\omega$ term comes from the temporal derivative of the exponential that defines the harmonic component, and $\tilde{\phi}_j, \tilde{n}_j, \tilde{p}_j$ are the electrostatic potentials and carrier densities normalized. In the case of voltage, normalization is done with respect to the contact where the voltage is applied.

The global AC matrix is obtained by imposing the boundary conditions (BC) and performing the summation in j. The BC commonly used in AC simulation are Newmann boundary conditions and oxide-semiconductor jump conditions[2] and also Dirichlet boundary conditions at the ohmic contacts with $\tilde{n} = \tilde{p} = 0$.

The assembled global matrix with the boundary condition can then be written as:

$$\left[\underline{\underline{J}} + i\underline{\underline{D}}\right] \underline{\tilde{u}} = \underline{B} \tag{2.49}$$

where $\underline{\underline{J}}$ is the jacobian matrix and $\underline{\underline{D}}$ contains the terms $\underline{G}$ present in 2.48. The real vector $\underline{B}$ contains the source term (which could be voltage, optical generation, etc.) and $\underline{\tilde{u}}$ is the solution vector. Considering separately real and imaginary parts, (2.49) can be rewritten as:

$$\begin{bmatrix} \underline{\underline{J}} & -\underline{\underline{D}} \\ \underline{\underline{D}} & \underline{\underline{J}} \end{bmatrix} \begin{bmatrix} \underline{u}_R \\ \underline{u}_I \end{bmatrix} = \begin{bmatrix} \underline{B} \\ \underline{0} \end{bmatrix} \tag{2.50}$$

with $\underline{\tilde{u}} = \underline{u}_R + i\underline{u}_I$ where $\underline{u}_R$ and $\underline{u}_I$ are the real and imaginary parts of the solution matrix.

2.3.1 Small-signal Current density

From the small-signal analysis is also possible to calculate the current density of the device. The displacement, electron, and hole currents can be calculated as:

$$\underline{\tilde{J}}_D = -i\omega\epsilon\nabla\tilde{\phi} \tag{2.51}$$

$$\underline{\tilde{J}}_n = \left.\frac{\partial \underline{J}_n}{\partial \phi}\right|_{DC} \tilde{\phi} + \left.\frac{\partial \underline{J}_n}{\partial n}\right|_{DC} \tilde{n} + \left.\frac{\partial \underline{J}_n}{\partial p}\right|_{DC} \tilde{p} \tag{2.52a}$$

$$\underline{\tilde{J}}_p = \left.\frac{\partial \underline{J}_p}{\partial \phi}\right|_{DC} \tilde{\phi} + \left.\frac{\partial \underline{J}_p}{\partial n}\right|_{DC} \tilde{n} + \left.\frac{\partial \underline{J}_p}{\partial p}\right|_{DC} \tilde{p} \tag{2.52b}$$

[2]Boundary or transition conditions used when the domain is not continuously differentiable [18].

The unit of the current density responses is $\mathrm{A\,cm^{-2}\,V^{-1}}$ in case of small-signal calculated with respect to the voltage, and $\mathrm{A\,cm\,s}$ if it is calculated with respect to optical generation.

2.3.2 Small-signal Transfer function

From the current density calculated in the previous section and the input perturbation, is possible to obtain the frequency-dependent transfer function between circuit nodes of the specified electrical system to which the simulated device is connected.

2.3.2.1 Electrical small-signal Admittance matrix

In the case of electrical small-signal, the transfer function is an admittance matrix represented as:

$$\delta I = Y \delta V \tag{2.53}$$

where Y is the admittance matrix, while δI and δV are the current and voltage in complex values at the selected nodes, respectively. The Y matrix can be then subdivided into real and imaginary parts:

$$Y = A + i\,\omega C \tag{2.54}$$

where A is the conductance matrix, C is the capacitance matrix, and ω is the frequency of the perturbation. This means that, from the small-signal analysis, is possible to obtain the values of resistance and capacitance of the device simulated.

2.3.2.2 Optical small-signal Quantum efficiency matrix

Solving the small-signal with respect to the optical generation rate, or the optical power density, the resulting transfer function obtained corresponds to the quantum efficiency matrix. Generally, the quantum efficiency is defined as the photocurrent I_L divided by the optical power density incident on the device P^{opt}, both normalized with respect to the elementary charge and the photon energy respectively. This means that the quantum efficiency is the responsivity normalized:

$$\eta = \frac{I_L/q}{P^{opt}/\left(\frac{hc}{\lambda}\right)} = \frac{1}{q}\left(\frac{hc}{\lambda}\right)\mathfrak{R} \tag{2.55}$$

For the small signal the quantum efficiency becomes complex because of the dynamic components:

$$H = \frac{1}{q}\left(\frac{hc}{\lambda}\right)\frac{\delta I}{\delta P} \tag{2.56}$$

dividing the current into real and imaginary parts we obtain the two contributions of the quantum efficiency matrix as for the admittance matrix in the electrical small-signal case:

$$H = \frac{1}{q}\left(\frac{hc}{\lambda}\right)\frac{Re\{\delta I\} + i\, Im\{\delta I\}}{\delta P} = \eta + i\,\omega C^{opt} \tag{2.57}$$

where

$$\eta = \frac{1}{q}\left(\frac{hc}{\lambda}\right)\frac{Re\{\delta I\}}{\delta P} \tag{2.58}$$

and

$$C^{opt} = \frac{1}{\omega q}\left(\frac{hc}{\lambda}\right)\frac{Im\{\delta I\}}{\delta P} \tag{2.59}$$

are the expressions of the real and imaginary parts of the quantum efficiency matrix (hence they are also matrices).

All calculations were done considering an optical power perturbation, but it is possible to consider directly a perturbation of the optical generation rate using the relation between the two (see 1.2.4.1):

$$\underline{G}^{opt}(x, y, z) = \frac{\alpha\lambda}{hc}\underline{P}(x, y, z) \tag{2.60}$$

where $\underline{G}^{opt}$ is the distribution of the optical generation rate inside the device due to the distribution of the optical power density $\underline{P}(x, y, z)$, while α is the optical absorption profile.

2.4 Finite Difference Time Domain method

The Finite Difference Time Domain (FDTD) method is a numerical approach that discretizes Maxwell's equations in time and space to study the propagation of electromagnetic waves. The FDTD method gives a rigorous solution of Maxwell's equations without approximations, which is useful when other simulation methods don't provide good quality solutions or have problems with the structure geometry. Synopsys RSoft FullWAVE [19] tool implements this numerical discretization method for calculating the light propagation in the photodetector.

A general formulation can be developed starting from Maxwell's Equations, assuming the absence of current sources ($\underline{J} = 0$) and free charges ($\rho = 0$). We also consider a non-magnetic medium ($\mu = \mu_0$). With these assumptions, Maxwell's equations become:

$$\nabla \times \underline{H} = \frac{\partial \underline{D}}{\partial t} \tag{2.61a}$$

$$\nabla \times \underline{E} = -\mu_0 \frac{\partial \underline{H}}{\partial t} \tag{2.61b}$$

Where $\underline{E}$ and $\underline{H}$ are, respectively, the electric and magnetic fields. $\underline{D}$, instead, is the electric induction, that, together with the magnetic induction $\underline{B}$ is related to the electric and magnetic fields by the constitutive relations:

$$\underline{D} = \epsilon_0 \underline{E} + \underline{P} \tag{2.62a}$$

$$\underline{B} = \mu_0 \underline{H} \tag{2.62b}$$

where $\underline{P} = \epsilon_0 \chi_e \underline{E}$ is the polarization vector. From here a one-dimensional (1D), two-dimensional (2D), or three-dimensional (3D) formulation can be developed. In this work, due to the complicated geometry of the device, is required a fully 3D approach. Equations (2.61) are so expanded using cartesian coordinates in six scalar equations [20]:

$$\frac{\partial D_x}{\partial t} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \tag{2.63a}$$

$$-\frac{\partial D_y}{\partial t} = \frac{\partial H_z}{\partial x} - \frac{\partial H_x}{\partial z} \tag{2.63b}$$

$$\frac{\partial D_z}{\partial t} = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \tag{2.63c}$$

$$-\mu_0 \frac{\partial H_x}{\partial t} = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \tag{2.63d}$$

$$\mu_0 \frac{\partial H_y}{\partial t} = \frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} \tag{2.63e}$$

$$-\mu_0 \frac{\partial H_z}{\partial t} = \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \tag{2.63f}$$

The discretization is done using the Yee algorithm. It solves both electric and magnetic fields in time and space, instead of only one of the two, obtaining a more robust solution than solving the equations for only one field. This makes the solution more accurate for a wider class of structures. Figure 2.1 shows the spatial distribution of field vector components on the cubic unit cell of the Yee discretized space.

The time is also discretized, and the $\underline{E}$ and $\underline{H}$ components are centered in time using a leapfrog arrangement. $\underline{E}$ components are calculated for a specific time in the Yee discretized space, then these results are used to calculate the $\underline{H}$ components in space but for a time step in between two time steps of the electric field, as shown in a simplified 1D case in figure 2.2.

After the discretization, Maxwell's equation, for the mesh points denoted

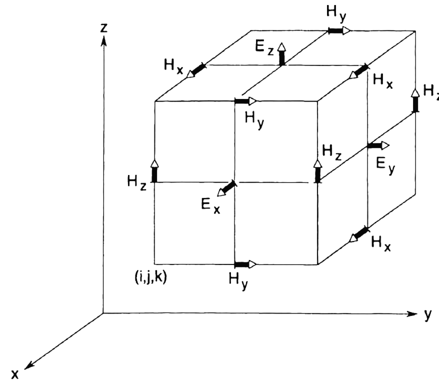


Figure 2.1: Distribution of electric and magnetic field vector components in a cubic unit cell of the Yee discretized space. Figure taken from [21].

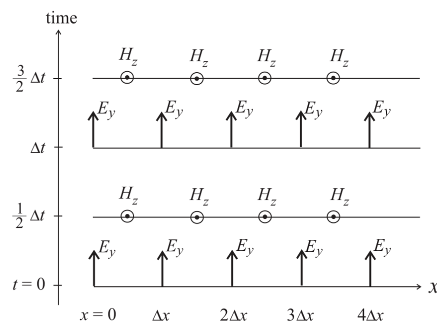


Figure 2.2: Space-time discretization of electric and magnetic field components in the one-dimensional case. The space is discretized with the central difference method, and time with the leapfrog method. Figure taken from [20].

by integers i,j,k and discretized time defined by the integer n, becomes:

$$E^{n+1}_{x(i,j,k)} = H^{n}_{x(i,j,k)} + \frac{\Delta t}{\epsilon \Delta y}\left(H^{n+1/2}_{x(i,j+1,k)} - H^{n+1/2}_{x(i,j,k)}\right) - \frac{\Delta t}{\epsilon \Delta z}\left(H^{n+1/2}_{x(i,j,k+1)} - H^{n+1/2}_{x(i,j,k)}\right) \tag{2.64a}$$

$$H^{n+1/2}_{x(i,j,k)} = H^{n-1/2}_{x(i,j,k)} + \frac{\Delta t}{\mu_0 \Delta z}\left(E^{n}_{x(i,j,k)} - E^{n}_{x(i,j,k-1)}\right) - \frac{\Delta t}{\mu_0 \Delta y}\left(E^{n}_{x(i,j,k)} - E^{n}_{x(i,j-1,k)}\right) \tag{2.64b}$$

where the equations (2.64b) and (2.64a), are the discretized form of two of the six Maxwell's equations, more specifically (2.63b) and (2.63d). All six are solved iteratively in space and time following the Yee algorithm [21].

2.4.1 Boundary conditions

Many types of boundary conditions can be implemented for the spatial grid, such as absorbing boundary conditions, which cancel any wave that reaches the boundaries. One type of absorbing boundary conditions extensively used by RSoft FullWAVE are the perfectly matched layer (PML) boundary conditions, where the electric and magnetic conductivities are calculated so that the wave characteristic impedance is equal inside and outside the calculation area, to avoid reflections at the boundaries. These are also the boundary conditions used in the simulation performed in this work. One last type of boundary conditions used by RSoft are the periodic boundary conditions, used for periodic structures such as distributed Bragg reflectors (DBR).

2.4.2 Simulation stability

There is a condition that must be satisfied to have a stable simulation using the FDTD method called the Courant condition, relating the temporal step size and the dimension of a unit cell of Yee's lattice (spatial grid), which can also be not cubic. The condition can be expressed as:

$$c\,\Delta t < \frac{1}{\sqrt{1/\Delta x^2 + 1/\Delta y^2 + 1/\Delta z^2}} \tag{2.65}$$

where c is the light velocity and the spatial step sizes considered are the smallest ones in the simulation domain. This condition comes from the physical limitation of the maximum phase velocity for an electromagnetic wave, which is the light velocity.

2.4.3 Stop criteria

The simulation start from time $t = 0$ to a predefined stop time t_s. The choice of the stop time defines the total simulation time and the accuracy of the results obtained. The stop criteria can be defined as a tolerance value that must be reached to have convergence. A stop time is defined anyway in case of divergence of the simulation.

2.4.4 Optical generation rate

The FDTD method calculates the electric and magnetic fields, but the objective of the optical simulation is to calculate the spatial profile of the absorbed light that generates carriers in the device. The first step is to obtain an explicit expression for the conductivity of the material. The dispersion relation for a non-magnetic medium ($\mu = \mu_0$) in absence of free charges ($\rho = 0$) can be obtained starting from Maxwell's equations:

$$\nabla \times \underline{H} = \frac{\partial \underline{D}}{\partial t} + \underline{J} \tag{2.66a}$$

$$\nabla \times \underline{E} = -\mu_0 \frac{\partial \underline{H}}{\partial t} \tag{2.66b}$$

which are the same as (2.61) but with the addition of the current term, which takes into account the losses. Applying the curl operator to (2.66b) we have:

$$\nabla \times \nabla \times \underline{E} = -\mu_0 \frac{\partial}{\partial t}(\nabla \times \underline{H}) \tag{2.67}$$

Considering the vector identity $\nabla \times \nabla \times \underline{E} = \nabla \cdot (\nabla \cdot \underline{E}) - \nabla^2 \underline{E}$ for the left hand side, and substituting (2.66a) in the right hand side, we have:

$$\nabla \cdot \underbrace{(\nabla \cdot \underline{E})}_{=0} - \nabla^2 \underline{E} = -\mu_0 \frac{\partial}{\partial t}\left(\frac{\partial \underline{D}}{\partial t} + \underline{J}\right) \tag{2.68}$$

where $\nabla \cdot \underline{E} = \rho = 0$. Now, considering a harmonic field of the type $\underline{E}(\underline{r}, t) = \underline{E}_0\, exp(\underline{kr} - \omega t)$ and substituting, we have:

$$-(i\underline{k})^2 \underline{E} = -\mu_0 \epsilon (-i\omega)^2 \underline{E} - \mu_0 \frac{\partial \underline{J}}{\partial t} \tag{2.69}$$

considering that the current density has the same harmonic behavior of the field, we have $\underline{J} = \underline{E}_0\, exp(-\omega t)$, that substituting, gives:

$$-(i\underline{k})^2 \underline{E} = -\mu_0 \epsilon (-i\omega)^2 \underline{E} - (-i\omega)\mu_0 \underline{J} \tag{2.70a}$$

$$\underline{k}^2 \underline{E} = \mu_0 \epsilon \omega^2 \underline{E} + i\omega \mu_0 \sigma \underline{E} \tag{2.70b}$$

$$\underline{k}^2 = \mu_0 \epsilon \omega^2 + i\omega \mu_0 \sigma \tag{2.70c}$$

where, between the first and second step, the microscopic Ohm's law $\underline{J} = \sigma \underline{E}$ was used, in which σ represents the electrical conductivity. The definition of the dispersion relation is then:

$$\underline{k}^2 = \omega^2 \epsilon \mu_0 \left(1 + i\frac{\sigma}{\omega \epsilon_0}\right) = \underline{k}_0^2 \left(\epsilon_r + i\frac{\sigma}{\omega \epsilon_0}\right), \text{ with } k_0 = \frac{2\pi}{\lambda} \tag{2.71}$$

the term between parenthesis is the complex electrical permittivity of the material, which can be also defined as $\epsilon = \epsilon' + i\epsilon''$, so compared to the expression obtained from Maxwell's equations:

$$\epsilon = \epsilon' + i\epsilon'' = \left(\epsilon_r + i\frac{\sigma}{\omega \epsilon_0}\right) \tag{2.72}$$

so we have:

$$\epsilon' = \epsilon_r \tag{2.73a}$$

$$\epsilon'' = \frac{\sigma}{\omega \epsilon_0} \tag{2.73b}$$

from (2.73b) we can find the expression for the conductivity in function of the imaginary component ϵ'', since the complex refractive index is an input dataset for the simulation.

$$\sigma = \omega \epsilon_0 \epsilon'' = \frac{2\pi c}{\lambda} \epsilon_0 \epsilon'' \tag{2.74}$$

Now, to calculate the spatial distribution of the absorbed power density we can use the continuity equation of the Poynting's vector [22].

$$P^{opt} = \frac{\partial u}{\partial t} = -\nabla \cdot S_{av} \tag{2.75}$$

where P^{opt} is the optical power density, u is the energy density, and S_{av} is the Poynting's vector averaged in time:

$$S_{av} = \frac{1}{2} Re\{\underline{E} \times \underline{H}\} \tag{2.76}$$

Substituting (2.76) in (2.75) we have:

$$P^{opt} = -\nabla \cdot \left(\frac{1}{2} Re\{\underline{E} \times \underline{H}\}\right) \tag{2.77a}$$

$$= -\frac{1}{2} Re\{\nabla \cdot (\underline{E} \times \underline{H})\} \tag{2.77b}$$

To calculate the section $\nabla \cdot (\underline{E} \times \underline{H})$ we can exploit the starting point of the Poynting's theorem's demonstration [23]. Multiplying both sides of (2.66b) for $\underline{H}$, (2.66a) for $\underline{E}$ and subtracting the first to the second we obtain:

$$\underline{H} \cdot (\nabla \times \underline{E}) - \underline{E} \cdot (\nabla \times \underline{H}) = -\underline{H}\frac{\partial \underline{B}}{\partial t} - \underline{E}\frac{\partial \underline{D}}{\partial t} - \underline{E} \cdot \underline{J} \tag{2.78}$$

the left-hand side can be rewritten using the vector identity $\underline{H} \cdot (\nabla \times \underline{E}) - \underline{E} \cdot (\nabla \times \underline{H}) = \nabla \cdot (\underline{E} \times \underline{H})$, obtaining:

$$\nabla \cdot (\underline{E} \times \underline{H}) = -\underline{H}\frac{\partial \underline{B}}{\partial t} - \underline{E}\frac{\partial \underline{D}}{\partial t} - \underline{E} \cdot \underline{J} \tag{2.79}$$

considering a harmonic field we substitute $\partial/\partial t$ with $-i\omega$

$$\nabla \cdot (\underline{E} \times \underline{H}) = i\omega(\underline{HB} + \underline{ED}) - \underline{EJ} \tag{2.80}$$

taking only the real part:

$$Re\{\nabla \cdot (\underline{E} \times \underline{H})\} = -\underline{EJ} = -\sigma|\underline{E}|^2 \tag{2.81}$$

and substituting in (2.77):

$$P^{opt} = \frac{1}{2}\sigma|\underline{E}|^2 \tag{2.82}$$

this expression of the optical power density absorbed can be used to calculate the consequent optical absorption rate G^{opt} as:

$$\underline{G}^{opt}(x, y, z) = \frac{\eta_Q}{E_{ph}}\underline{P}^{opt}(x, y, z) = \frac{\eta_Q \lambda}{hc}\underline{P}^{opt}(x, y, z) \tag{2.83}$$

where η_Q is the internal quantum efficiency, and E_{ph} is the photon energy. In the final expression, the spatial dependence is made explicit to highlight that the simulation calculates a spatial distribution of the optical power density and the optical generation rate. The difference between (2.60) and (2.83) comes from the fact that, in the former, consider a simpler one-dimensional case in a homogeneous material, where the optical power density can be directly calculated from the differential equation (1.5), whose solution was presented immediately before in (1.4), therefore the presence of the absorption coefficient α is given by the derivation of the optical power. This case, instead, presents a generalized approach, where the optical power is given from the electric field distribution calculated directly from Maxwell's equations in three dimensions. The only remaining difference is the presence of the internal quantum efficiency in (2.83), which, however, is assumed to be equal to one.

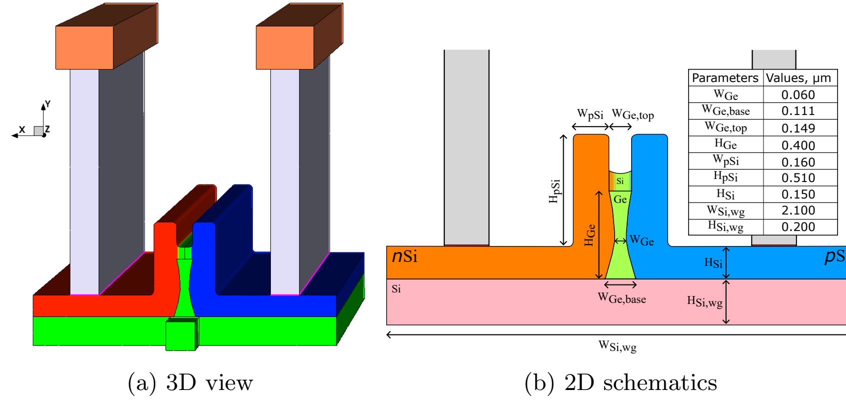
Chapter 3

Results

This chapter presents the results of the simulations performed. It starts by presenting the device geometry, followed by an analysis of the principal figures of merit, such as the dark current, photocurrent, electro-optic frequency response, and responsivity. Their dependence on geometry, material properties, and applied bias is also investigated.

3.1 Simulation

To correctly simulate the device, a 3D multiphysics electro-optical simulation has been performed. The simulation is based on two commercial tools from Synopsys, TCAD Sentaurus and RSoft FullWAVE. TCAD Sentaurus is used to perform the electrical study of the device under equilibrium and bias, by using a Drift-Diffusion model (see ch.2.2), to analyze the main figure of merit of the device, such as the electric field, the carrier densities, the energy band diagram, and the dark current. RSoft FullWAVE, instead, performs a simulation of the light propagation on the device from the waveguide, using a Finite Difference Time Domain method (see ch.2.4). The 3D geometry of the photodetector is reproduced from the data extracted from figure 1.6 and the geometrical description of the device from the paper. The resulting structure is presented in figure 3.1. Two different devices are simulated, named in the reference paper Ge_100_10 and Ge_150_10, and renamed *Device 1* and *Device 2* in the corresponding simulations. Figure 3.1 represents *Device 1*. The device is composed of a short input silicon waveguide with equal height and width of 0.2 µm. It is connected to a silicon substrate of the same height and width of 2.1 µm. Over the substrate, there are the two doped silicon regions p (blue region) and n (red region) with the same level of doping of $1 \times 10^{20}\,\text{cm}^{-3}$. The green color is assigned to the intrinsic semiconductive regions, which are the silicon waveguide, substrate, and the germanium region between the two doped silicon regions. The germanium heigh is 400 nm with a base width of 149 nm. The minimum width is around



Parameters	Values, μm
W_{Ge}	0.060
$W_{Ge,base}$	0.111
$W_{Ge,top}$	0.149
H_{Ge}	0.400
W_{pSi}	0.160
H_{pSi}	0.510
H_{Si}	0.150
$W_{Si,wg}$	2.100
$H_{Si,wg}$	0.200

(a) 3D view

(b) 2D schematics

Figure 3.1: The left figure shows the 3D geometry of the photodetector. The right figure, instead, is a 2D schematic of a cross-section of the device in the xy-plane in a generic point in the z-axis (the structure is invariant in this direction).

the center of the germanium and corresponds to 60 nm. The *Device 2* has the same geometry parameters as *Device 1* aside from the germanium region that is 50 nm thicker, with a minimum width of 110 nm.

Details about the materials, the heterojuction between germanium and silicon, and the simulation mesh of the device are presented in the Appendix.

3.1.1 Optical power propagation

The input optical power reaches the photodiode passing through the waveguide. The only mode propagating in the waveguide is the fundamental mode which focuses the optical power on the center of the waveguide. The light reaches the silicon substrate of the photodiode where it is not more confined, starting to irradiate power in a cone-shape. Part of the power is irradiated directly into the germanium region, while most continue to travel straight. Since there is no more confinement the light traveling in the substrate is absorbed by the germanium as an evanescent field for the entire length of the device. The absorbed light generates photons in the device with a distribution defined by the optical generation rate. Figure 3.2 represents the optical generation rate for the C-Band and 3.3 for the O-Band. The absorption coefficient is higher for germanium in the O-Band, so it presents a higher optical generation rate. The periodic peaks of the generation rate in the absorber are due to the destructive interference caused by the light reflecting back from the upper interface between germanium and silicon oxide, which can be considered beneficial since it helps to better distribute the optical power inside the absorber, avoiding peaks with too high intensity, which would cause screening at higher optical power.

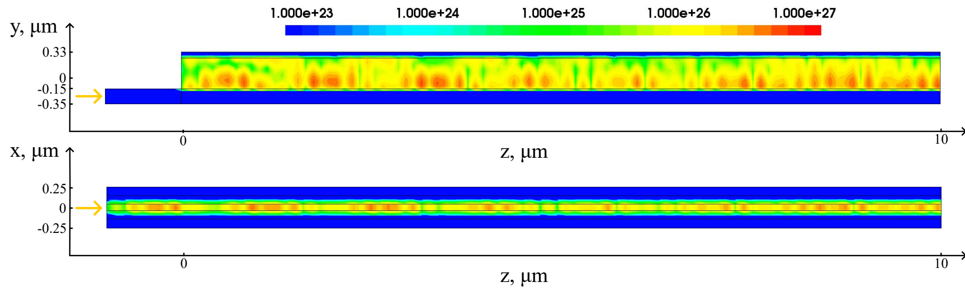


Figure 3.2: Optical generation rate distribution on *Device 1*, considering a $P_{opt} = 200\,\mu\mathrm{W}$ and $\lambda = 1.55\,\mu\mathrm{m}$, in a yz-plane cross-section at $x = 0$ (view from above, upper figure) and in an xz-plane cross-section at $y = 0$ (lateral view, lower figure).

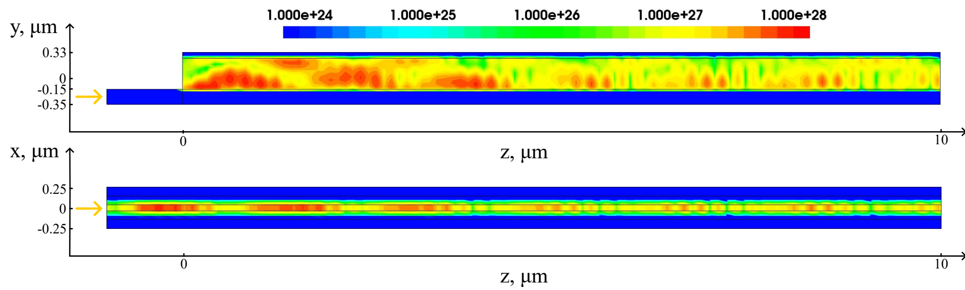


Figure 3.3: Optical generation rate distribution on *Device 1*, considering a $P_{opt} = 200\,\mu\mathrm{W}$ and $\lambda = 1.31\,\mu\mathrm{m}$, in a yz-plane cross-section at $x = 0$ (view from above, upper figure) and in an xz-plane cross-section at $y = 0$ (lateral view, lower figure).

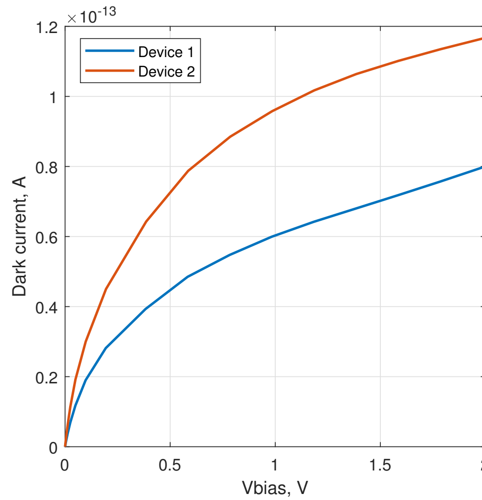


Figure 3.4: Dark current comparison between *Device 1* and *Device 2*.

3.2 Dark current

The first figure of merit analyzed in the simulation is the current generated by the device in dark conditions (i.e. in the absence of a light source) under bias. The current is due to the *pin* junction and can be classified as a noise contribution since it is not caused by an input signal. Fortunately, this type of photodetector has a very low dark current compared to the photocurrent generated, while other classes of optical sensors need specific solutions to reduce this noise source which could be even higher than the photocurrent. As shown in figure 3.4, the dark current simulated in *Device 2* is bigger than the one of *Device 1* by almost half an order of magnitude, with a value of 1.17×10^{-13} A and 8×10^{-14} A respectively.

The reason for this difference can be found in generation processes in the absorber that give a bigger contribution due to the bigger volume, increasing the dark current. In figure 3.5 all the generation and recombination terms meaningful in dark conditions are represented. The germanium region is depleted from carriers because of the *pin* junction and the applied reverse bias, causing the net SRH recombination rate to become negative, generating new carriers that are then removed from the depletion region by the electric field. A small contribution of generation is given also by Auger processes, but it can be considered negligible since it is proportional to the number of carriers in the material. Increasing the width of the germanium region also increases the volume that generates carriers, giving more space to electrons and holes to gain kinetic energy from the acceleration given by the high electric field in the depletion region to the point where it overcomes the threshold limit causing avalanche generation.

However, the measured values for Ge_150_10 and Ge_100_10, as previ-

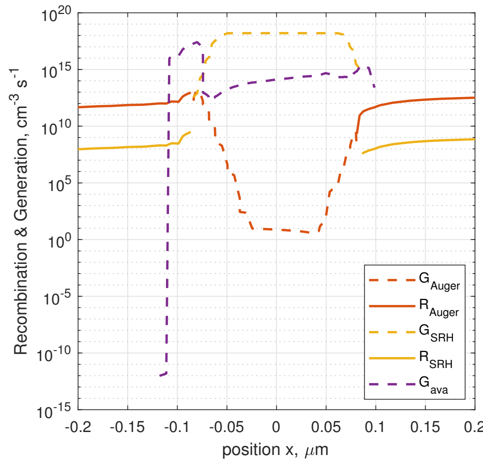


Figure 3.5: Recombination and generation processes represented in absolute value along the x-axis in a cross-section of the device with respect to the z-axis and at 15 nm over the silicon substrate in dark conditions and applied bias of −2 V.

ously presented in table 1.1 are six orders of magnitude higher, even if they display a similar relative difference in value between the two devices. To correctly model the currents would be necessary to know the quantity and distribution of defects on the absorber caused by doping diffused from the silicon regions, impurities, mechanical stress on the interface between silicon and germanium, and other types of defects caused by the production process. This is, however, very difficult to simulate and there are very few measurements and studies about it in literature. Fortunately, as already stated, in this class of photodetectors, the dark current is negligible compared to the photocurrent, so there is no need or interest to further analyze the phenomenon.

3.3 Photocurrent

The simulation of the photocurrent requires also the simulation of the light propagation through the waveguide and the device. For this reason, a 3D simulation is essential, especially to simulate the evanescent coupling of the light between the silicon substrate and the germanium absorber. The light propagation problem is solved by RSoft, first calculating the mode in the waveguide and then using FDTD to solve Maxwell's equations and to calculate the optical power distribution inside the device (see 2.4). From the optical power is possible to obtain the optical generation rate of carriers due to the absorption of the photons. The optical generation profile is then used in the 3D TCAD simulation performed with Sentaurus to solve another time

the DD model, obtaining the total current, which is the sum of the dark current and the photocurrent. Since the photocurrent is on the order of mA, while the dark current is only, in the worst case on the order of hundreds of nA, we can consider the latter negligible under illumination.

The photocurrent is generated by the absorption of photons in the germanium absorber, which depends on the optical parameters of the material, more precisely by the absorption coefficient α. The absorption coefficient is, in turn, related to the extinction coefficient k, which is the imaginary part of the refractive index:

$$n(\lambda) = n + ik \tag{3.1a}$$

$$\alpha(\lambda) = \frac{4\pi k(\lambda)}{\lambda} \tag{3.1b}$$

The absorption coefficient is directly related to the absorption process of a photon in a material. To absorb a photon it must have at least an energy equal to the band gap of the material. The transition probability is also influenced by the type of bandgap, direct or indirect, and by the presence of trap states generated by impurities, which could increase the probability of absorption for photons with energy lower than the bandgap since they are placed inside. If the material presents strain, it modifies the energy bands' shape and position, which also can change dramatically the absorption probability. In the device studied in this work, the germanium absorber is grown epitaxially in the silicon substrate, which, because of the lattice mismatch of 4.2 %, is affected by strain. The literature contains much data about bulk germanium but lacks studies and measurements on thin layers, and the effect of strain due to the material where they are grown on. In one of these studies, done by Sorianello et al. [24], high variability in the absorption coefficient between bulk and thin film can be observed. From this, it is possible to deduce that the absorption coefficient is strongly dependent on the production process of the material and possible sources of strain such as interfaces with different materials with different lattice constants, and that also the dimension of the material must be considered. This generates a high variability of the absorption coefficient, which increases even more if we consider pathological cases like the C-Band wavelength (1.55 µm) in the germanium, which is positioned in the absorption threshold, the region with the higher variation, and is also the one used in the measurement of Lischke et al. In literature is possible to find multiple sets of data for the complex refractive index of the germanium. Still, they are almost all from bulk material and they show high variability in the threshold region as can be seen in figure 3.6. All the presented results, if not differently stated, are simulated for an applied voltage of −2 V, as in the reference paper from Lischke et al.

The resulting photocurrent simulated using these sets of data is shown in figure 3.7.

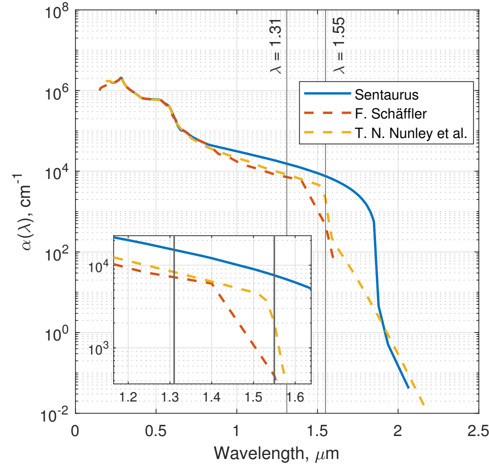
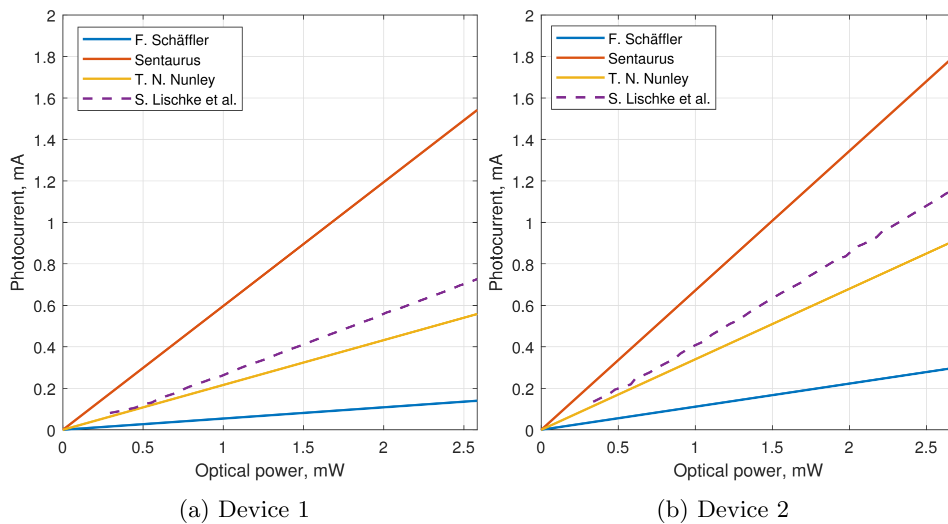


Figure 3.6: Absorption coefficient of bulk germanium for three different sets of data, one from the Sentaurus library and two from literature.



(a) Device 1

(b) Device 2

Figure 3.7: Photocurrent of *Device 1* and *Device 2* for the three sets of data of absorption coefficient presented in figure 3.6 ($\lambda = 1.55\,\mu\text{m}$). The higher volume of the germanium region in *Device 2* results in a higher value of photocurrent and thus of responsivity with respect to *Device 1*.

Data	Responsivity (A/W)	Quantum efficiency
Sentaurus	0.597	0.478
F. Schäffler	0.054	0.043
T. N. Nunley et al.	0.216	0.173
S. Lischke et al.	0.3	0.24

Table 3.1: Responsivity and Quantum efficiency calculated from the photocurrents in figure 3.7 for *Device 1* and $\lambda = 1.55\,\mu m$.

Data	Responsivity (A/W)	Quantum efficiency
Sentaurus	0.672	0.538
F. Schäffler	0.111	0.089
T. N. Nunley et al.	0.340	0.272
S. Lischke et al.	0.45	0.36

Table 3.2: Responsivity and Quantum efficiency calculated from the photocurrents in figure 3.7 for *Device 2* and $\lambda = 1.55\,\mu m$.

As can be observed from the figure, there is high variability in the results depending on the set of data used in the simulation. This can be seen also in table 3.1 and 3.2 which show the calculated responsivity and quantum efficiency in the three cases. It can be concluded that to have a precise prediction of the photocurrent and responsivity it would be necessary to have a dataset for the absorption coefficient specific for the considered case, especially for thin layers of material like in the case described in this work, where the material cannot be considered bulk and the strain is not negligible.

More reliable results could be obtained for the O-Band, where the dependence from strain and dimension is lower. There is no data for comparison from the reference paper to validate the results, but it is still useful to analyze the difference with the C-Band case presented until now. The photocurrent in the O-Band is shown in figure 3.8. As expected, it presents a lower variability between the three sets of data for the absorption coefficient. Table 3.3 and 3.4 show the calculated responsivity and internal quantum efficiency, which are higher than the C-Band case due to the higher absorption.

3.4 Electro-optic frequency response

The electro-optic frequency response is obtained by calculating the small-signal of the current with respect to the optical power density in input (see

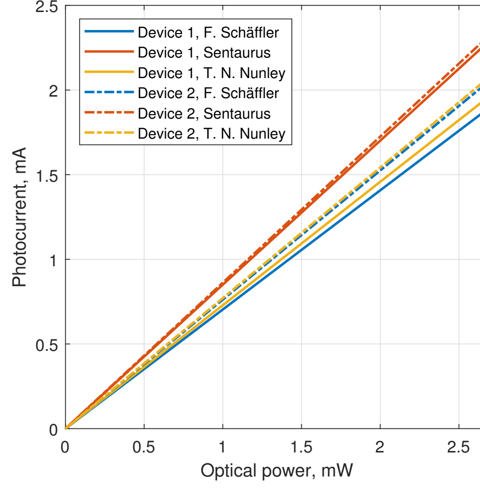


Figure 3.8: Photocurrent of *Device 1* and *Device 2* for the three sets of data of absorption coefficient presented in figure 3.6 ($\lambda = 1.31\,\mu m$).

Data	Responsivity (A/W)	Quantum efficiency
Sentaurus	0.850	0.680
F. Schäffler	0.704	0.563
T. N. Nunley et al.	0.723	0.583

Table 3.3: Responsivity and Quantum efficiency calculated from the photocurrents in figure 3.8 for *Device 1* and $\lambda = 1.31\,\mu m$.

Data	Responsivity (A/W)	Quantum efficiency
Sentaurus	0.862	0.690
F. Schäffler	0.762	0.610
T. N. Nunley et al.	0.771	0.617

Table 3.4: Responsivity and Quantum efficiency calculated from the photocurrents in figure 3.8 for *Device 2* and $\lambda = 1.31\,\mu m$.

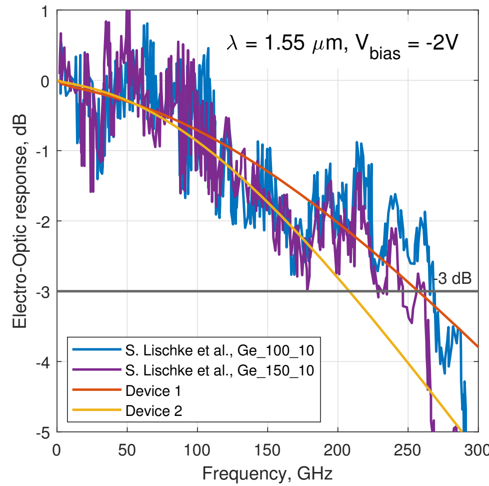


Figure 3.9: Electro-optic frequency response of *Device 1* and *Device 2*.

	f_{-3dB} (GHz)	
	Device 1	Device 2
simulation	257.2	208.1
S. Lischke et al.	265	240

Table 3.5: Comparison simulated and measured cut-off frequency from figure 3.9.

2.3). In figure 3.9 the simulation results for *Device 1* and *Device 2* are compared to the measurements from the paper. The curves of the measurements have a particularly noisy behavior, different from the single pole low-pass filter expected from a photodetector, and also reproduced from the simulation. This is probably due to external effects caused by the measurement instruments. Table 3.5 compares the cut-off frequencies measured and simulated, showing a difference of around 7.8 GHz between the two for *Device 1*, which correspond in a relative error of only around 3 %. For *Device 2* the difference is greater, of 31.9 GHz, corresponding to a relative error of 12 %. The reason for which the model predicts with less accuracy the results for *Device 2* is because the device geometry for the simulation was created using as a reference the section shown in 1.6 with belongs to Ge_100_10 (*Device 1*). In fact, the geometry for *Device 2* is simply the one of *Device 1* with an offset of 50 nm in the germanium absorber width. The complete lack of details about the geometry could be the main reason for this inaccuracy in the results since both the RC limit and the transit time limit of the bandwidth are strongly affected by the geometry.

The results shown in figure 3.9 are calculated using the absorption coefficient dataset from T.N. Nunley et al. However, the dataset chosen does not

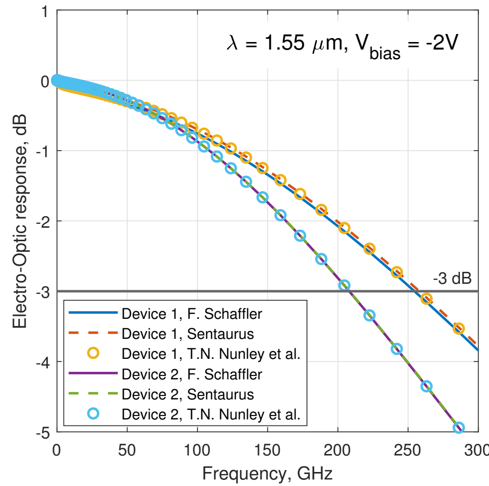


Figure 3.10: Comparison electro-optic frequency response of *Device 1* and *Device 2* for the three dataset of the absorption coefficient. It can be easily seen that there is almost no dependence of the curves on the dataset chosen for the simulation.

have a great impact on the productivity of the simulation, since, as shown in figure 3.10, the difference between two different datasets for the *Device 1* is less than 4 GHz in the worst case, and for *Device 2* is so small that can be considered negligible.

3.5 Simulation in O-Band

Until now, all the results presented were simulated in C-Band to be comparable with the experimental results. However, given all the peculiarities and problems of this specific wavelength in the simulation, in this section, where are presented simulation results with no experimental counterpart, the wavelength used will be 1.31 µm (O-Band), while the dataset used for the absorption coefficient will be the one from F. Schäffler which is plenty validated for this wavelength [25, 26, 27].

Figure 3.11 compares the electro-optical bandwidth calculated for the two wavelengths. The two results are almost equal, with a slightly higher value for the O-Band. More precisely, the O-band presents a cut-off frequency of 261.1 GHz for *Device 1* and 210.2 GHz for *Device 2*, while in C-Band they are equal to 254 GHz and 208 GHz (the values are different from the values in table 3.5 because they are calculated with the absorption coefficient dataset from F. Schäffler).

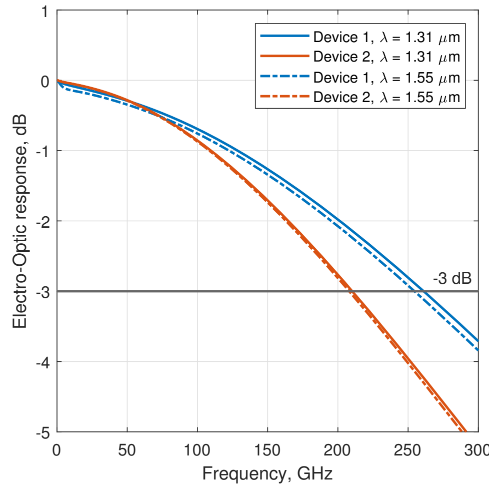
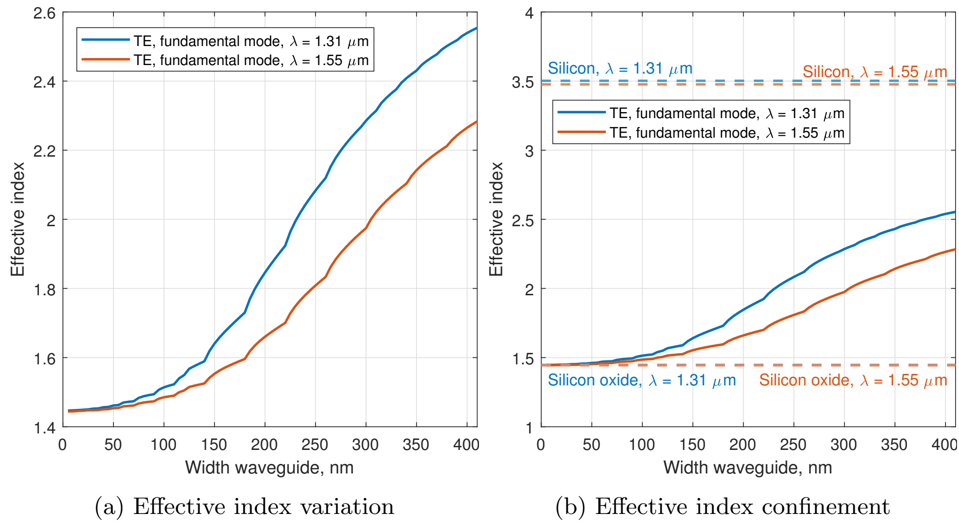


Figure 3.11: Comparison electro-optic frequency response of *Device 1* and *Device 2* between O-Band and C-Band.

3.5.1 Waveguide optimization

To maximize the performance of the device, it is necessary to have good confinement of the optical power inside the input waveguide. If the light is not well confined part of the optical power is lost in the oxide before reaching the photodiode. The waveguide must also be monomodal since the fundamental mode has the best shape to concentrate the light toward the active region. If the waveguide is too large it would not focus the light into the germanium region. Instead, if it is too thin it would not confine the field correctly. An electromagnetic field is confined if the effective index of the propagating mode is contained between the cladding's refractive index and the core's refractive index. An effective index over the core value would mean that the mode cannot propagate, while if it is under the cladding value it is not confined, and the light irradiates outside the core. Figure 3.12a shows the variation of the effective index with respect to the optical waveguide width for the O-Band and C-Band, while figure 3.12b compares the values with the refractive index of the core (silicon), and cladding (silicon oxide), showing that the light is better confined in the O-Band for the same value of thickness since the relative curve is more distant from the refractive index of the cladding.

Only the optical power that reaches the germanium region contributes to the photocurrent, so good confinement of the field is essential to maximize the responsivity of the device. The variation of the responsivity as a function of the optical waveguide width is presented in figure 3.13. The case in O-Band shows a maximum at 220 nm of thickness, and 350 nm for the C-Band. The effective index, however, continues to rise in value even after the peak of responsivity for both cases, which means that there is also another effect



(a) Effective index variation

(b) Effective index confinement

Figure 3.12: Figure (a) shows the variation of the effective index for the O-Band and C-Band with respect to the optical waveguide width for the fundamental mode. Figure (b) compares the curves to the values of the refractive index of the core and cladding to determine is the field is confined or not.

at play, which is supposedly the evanescent coupling with the germanium region, meaning that there is an optimal value of coupling determined by the ratio between the geometrical size of the waveguide and the base of the germanium depending on the wavelength of the light. A more thorough investigation would be however necessary to validate this assumption.

There is no information about the waveguide dimensions in the paper of Lischke et al. Therefore, in the simulations, an arbitrary value of 200 nm is used in both height and width.

Regarding the electro-optical frequency response, as can be expected, there is no particular dependence with respect to the optical waveguide geometry.

3.5.2 Effects of the applied voltage bias

Varying the applied voltage bias to the detector it is possible to observe an increase in responsivity (fig.3.14), especially in the case of *Device 1*.

Normally a photodiode should show almost no dependence on the responsivity with respect to the applied bias, which should be only high enough to ensure that all carriers are collected from the active region before recombining and that the saturation velocity is reached, to maximize the bandwidth. The presence of an increase in responsivity for higher voltages means that the reverse current of the *pin* diode is no more in saturation, but it reached

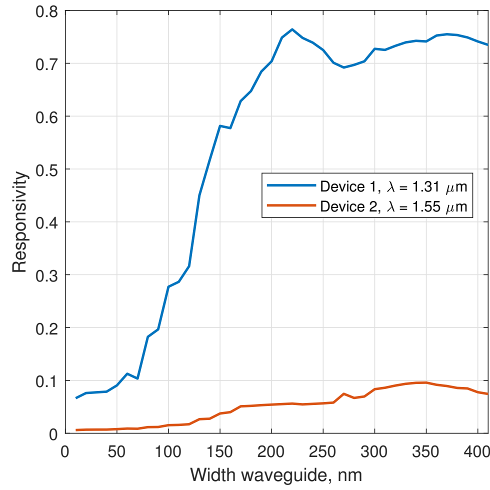


Figure 3.13: Dependence of responsivity with respect to the optical waveguide thickness in *Device 1* for O-Band and C-Band.

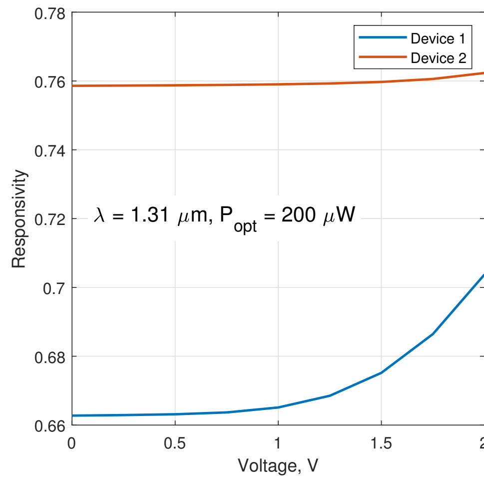
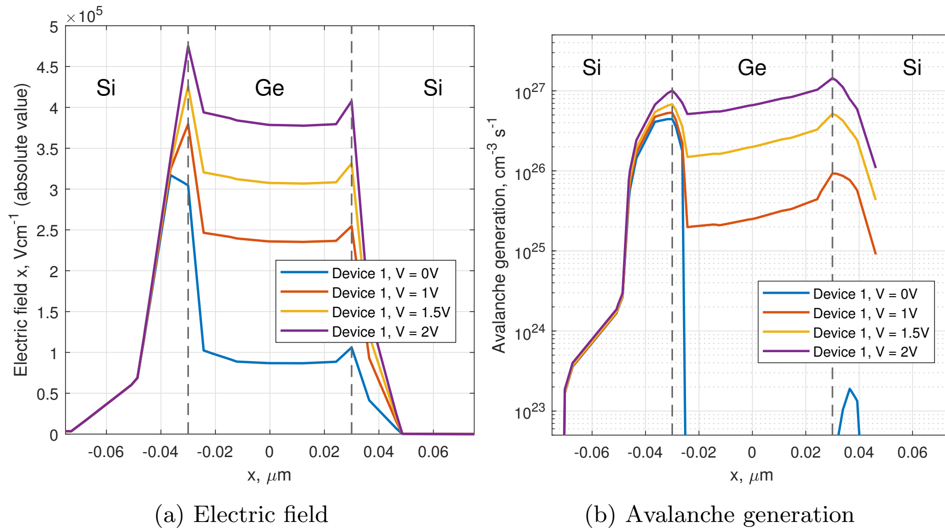


Figure 3.14: Dependence of responsivity with respect to the applied reverse bias.



(a) Electric field (b) Avalanche generation

Figure 3.15: The left figure shows the electric field and avalanche generation along x-direction for a cross-section of the device (*Device 1*) positioned 0.5 µm after the point of connection between the waveguide and the detector (z-direction) at the height for which the germanium absorber thickness is minimum (y-direction).

breakdown. The breakdown is caused by the avalanche generation that happens when the electric field is high enough to accelerate carriers enough to cause impact ionization. Figure 3.15 shows the profile along the x-direction in a cross-section of the device for multiple values of voltage bias in *Device 1*. Figure 3.16, instead, shows the avalanche generation value at the center of the germanium region with respect to the voltage for both devices.

The avalanche generation is much higher in *Device 1* with respect to *Device 2* because the same voltage is applied in a smaller intrinsic region, resulting in a higher value for the electric field, meaning that the thinner device goes in breakdown before the larger one. The presence of the avalanche effect, however, is not positive for the device because, even if it increases the responsivity, it also causes a decrease of the bandwidth due to the time necessary for the avalanche build-up [10] as shown in figure 3.17. In both cases, the bandwidth shows a maximum around 1 V, right before the build-up effect starts to be prevalent.

The bandwidth-efficiency product summarizes the effect of avalanche generation on the device (figure 3.18), showing, for *Device 1* an increase in the overall performance until around 1 V, before increasing again. However, the second increase is due to the avalanche effect that increases the responsivity more than it reduces the bandwidth, causing an overall increase in the band-efficiency product. In the case of *Device 2*, the curve remains flat after the initial increase because there is no significant avalanche in the simulated

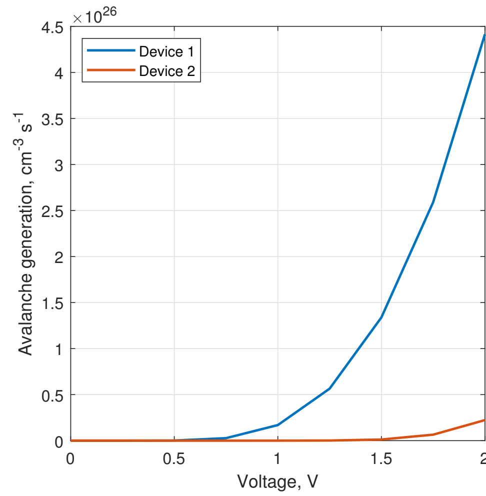
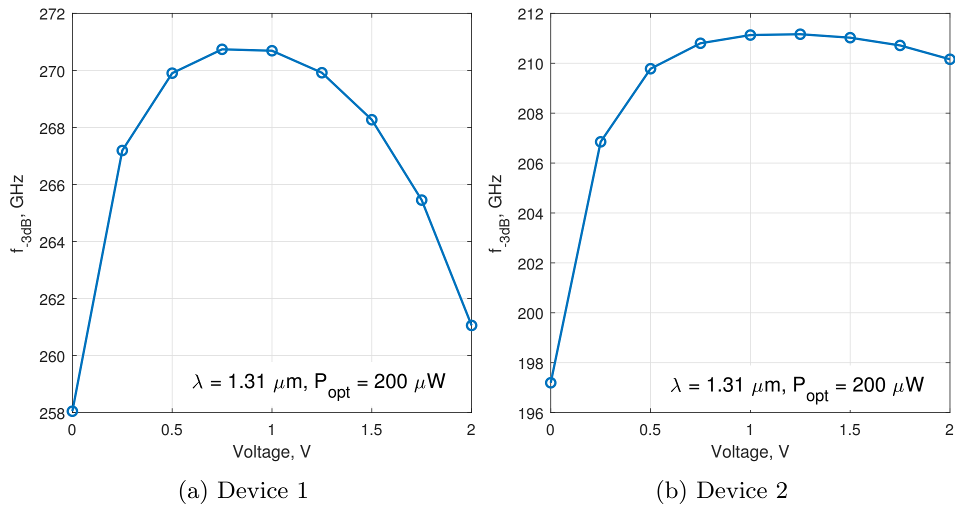


Figure 3.16: Variation of the avalanche generation with respect to the applied voltage for the same cross-section as figure 3.15 at the center of the germanium region ($x = 0$).



(a) Device 1

(b) Device 2

Figure 3.17: Dependence of the cut-off frequency from the applied reverse bias voltage for *Device 1* and *Device 2*.

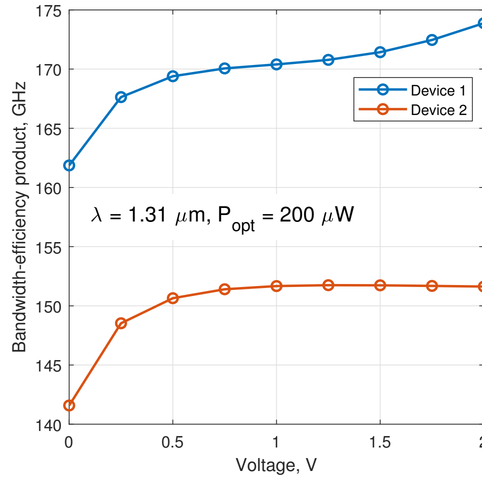


Figure 3.18: Dependence of the Bandwidth-efficiency trade-off from the applied reverse bias.

range of voltages.

Since this device focuses on reaching the highest bandwidth possible, is better to avoid the avalanche effect. This simulation shows that an optimal value to maximize the bandwidth is around 1 V. However, it would not be the same for the measured device, since the simulation applies the voltage directly to the *pin* junction without considering the voltage drop on the contacts, which can be significant. In any case, a slightly lower applied voltage of at least a few tenths of a volt would probably slightly increase the bandwidth in the real device (only for the thinner device equivalent to *Device 1*).

3.5.3 Effects of the Ge width

Variating the germanium region width, the active region of the device, the number of photogenerated carriers changes, causing a consequent variation in the responsivity as shown in figure 3.19. The effective width used in the graph is calculated as a mean value with respect to the concave shape of the germanium. The responsivity shows a noisy and slightly parabolic behavior (the curvature is emphasized by the figure scale), which presents a minimum for a width around 100 nm.

The behavior of the responsivity can be explained by considering the single contributions to the responsivity. From figure 3.20a can be observed that, increasing the thickness of the germanium, the total generation in the volume increases. The blue curve represents the contribution to the generation rate of the germanium only, while the red curve is an integration of the generation rate on the entire device, taking also into consideration the generation in the silicon. The optical generation alone explains only the

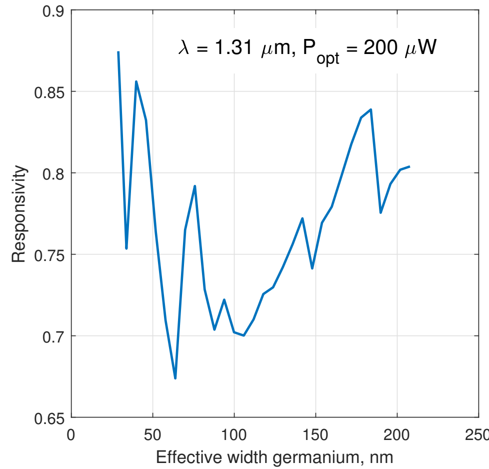
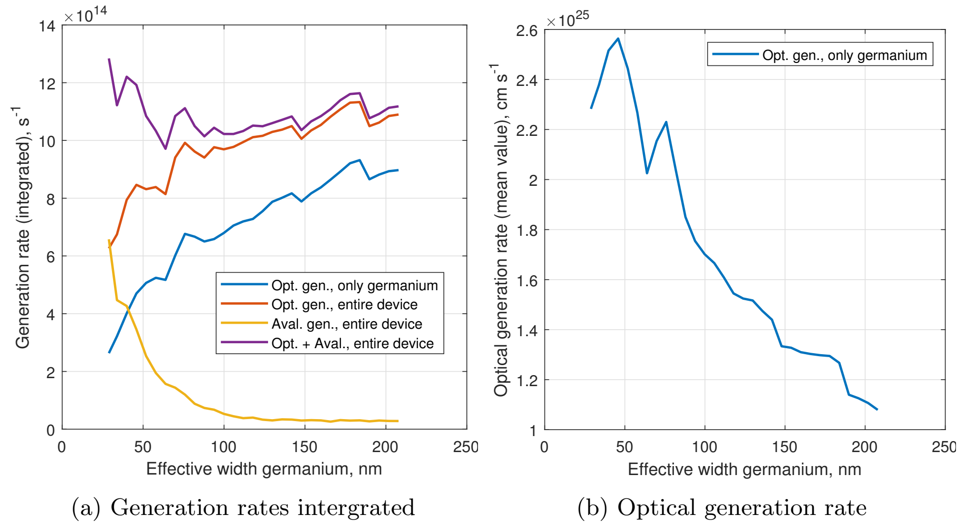


Figure 3.19: Dependence of the responsivity from the effective width of the germanium region.

increase of responsivity at the increase of the absorber volume. The high values of responsivity for low values of germanium width can be explained by the avalanche effect (yellow curve) which is particularly high when the germanium is thin due to the consequent increase in the electric field. Adding up both the optical generation and the avalanche generation we obtain a total generation rate represented by the violet curve. Figure 3.20b shows the optical generation rate integrated on the volume of the germanium region and divided by the same volume. It represents an evaluation of the efficiency of the device to generate photons with the same optical power in input. The decrease in efficiency for higher values of width is due to the lower uniformity of the generation rate compared to smaller germanium regions.

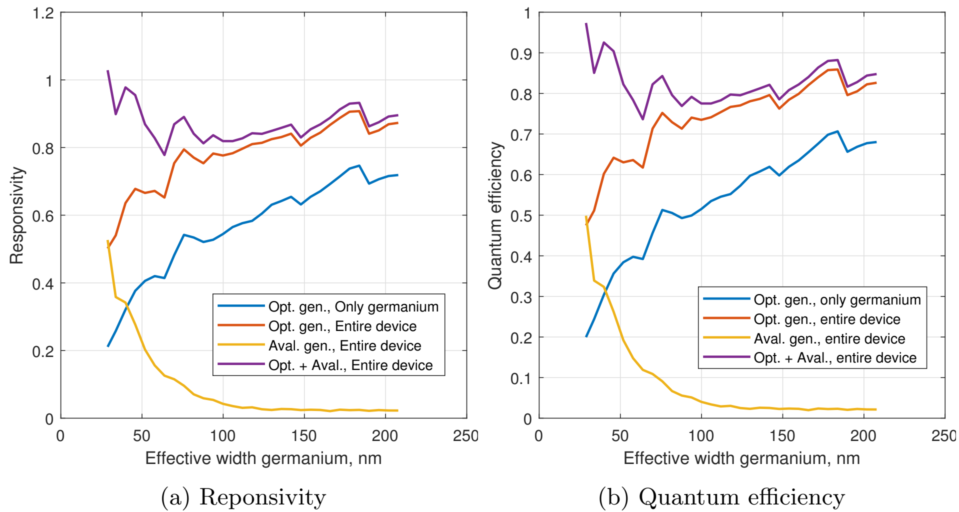
From the sum of the optical and avalanche generation rates is possible to calculate a responsivity and quantum efficiency that take into consideration only these contributions (figure 3.21). In figure 3.22 they are compared with the results obtained by the complete simulation. The difference between the curves is given by the fact that considering only the optical and avalanche generation for the calculation is equivalent to considering the device ideal. That difference is given by the recombination processes that recombine part of the carriers before they are collected by the contacts.

The variation of the germanium width affects also the bandwidth since a wider absorber means a higher transit time (lower transit time limit), but also a smaller capacitance (higher RC limit). Figure 3.19 shows the dependence of the cut-off frequency in function of the germanium width. The behavior is the one expected from the *pin* detector theory (chapter 1.2.4). The only difference is that from the analysis on the responsivity we now know that under around 50 nm of thickness the avalanche generation overcomes



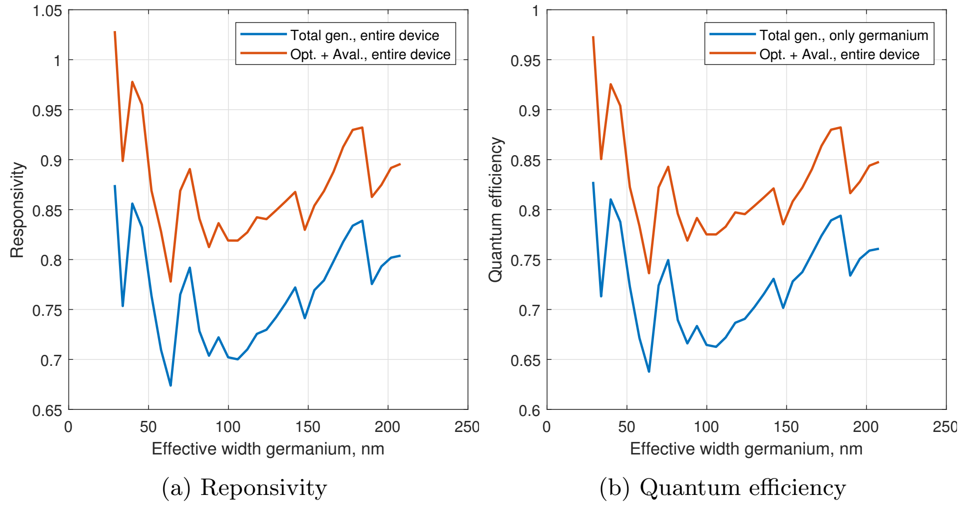
(a) Generation rates intergrated (b) Optical generation rate

Figure 3.20: (a) Generation rates in *Device 1* in function of the germanium region thickness. (b) Optical generation rate in the germanium region of *Device 1* mediated on the volume for different values of the absorber thickness.



(a) Reponsivity (b) Quantum efficiency

Figure 3.21: Responsivity (a) and quantum efficiency (b) in *Device 1* due to the optical and avalanche generation rates.



(a) Reponsivity (b) Quantum efficiency

Figure 3.22: Responsivity (a) and quantum efficiency (b) in *Device 1* due to the optical and avalanche generation rates compared to the total one.

the optical one, meaning that the decrease in bandwidth for smaller values of germanium width is due not only to the capacitive effects but also to the avalanche build-up time.

3.5.4 Effects of the carrier saturation velocity

The saturation velocity is an important parameter of the carriers defined by the transport properties of the material, more specifically the mobility. To have the maximum bandwidth possible it is important that the electric field in the absorber is high enough to permit electrons and holes to reach the saturation velocity. This is one of the reasons for which the device must be in low power injection condition, so that there is no screening of the electric field in the germanium, avoiding a reduction of the electric field, which would also cause the PI characteristic to be no more linear. The value of saturation velocity used for all previous simulations was $0.743 \times 10^7\,\mathrm{cm\,s^{-1}}$. The dependence of the bandwidth with respect to the saturation velocity is shown in figure 3.24 displaying an almost linear behavior. The difference in cut-off frequency between *Device 1* and *Device 2* is due mainly to the transit time, which increases with the increase in the germanium width. The figure makes clear that a precise value of saturation velocity is needed for a correct simulation of the bandwidth, which is however difficult to obtain given the scarcity of information on electrical properties for the germanium in literature. The saturation velocity depends also on the production process of the germanium since it depends on the mobility, which is related to the band structure of the material which, in turn, is affected by material strain and the presence of defects. In devices like the one studied in this work the

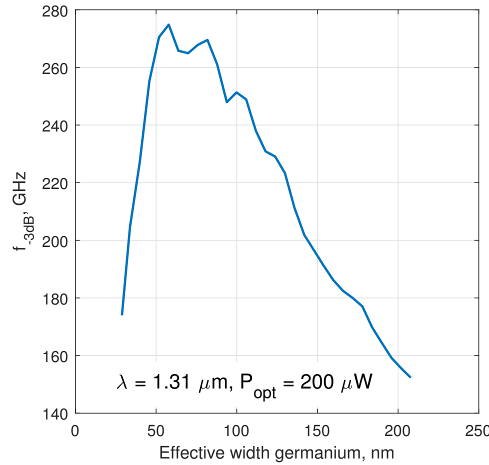


Figure 3.23: Dependence of the cut-off frequency from the effective width of the germanium region.

active region is thin enough that the overshoot in velocity of photogenerated carriers could make up even half of the total path inside the absorber before scattering enough to reach a stable value (saturation velocity). This permits the carriers to travel faster than what would be expected by considering only a constant velocity profile, which would make the saturation velocity an inadequate estimation of the mean velocity of carriers inside the absorber and consequently of the transit time [28].

3.5.5 Germanium shape

The particular biconcave shape of the photodetector's active region was chosen by Liscke et al. because, as they stated in their paper, thanks to its smaller thickness in the center, where there is supposedly the maximum optical power density, it improves the transit time limit by decreasing the travel distance in the germanium for these photogenerated carriers. To verify if this statement is valid, or at least reproducible in the simulation, two different geometries for the germanium absorber are compared. One is the already discussed biconcave shape in *Device 1* and *Device 2* and the other is a rectangular shape with an effective width calculated as a mean value of the curvature in the biconcave germanium. The calculated values are 88 nm for *Device 1* and 139 nm for *Device 2*. A comparison between the two geometries is shown in figure 3.25.

The simulation results for the photocurrent are shown in figure 3.26a. *Device 1* shows a slightly higher photocurrent for the rectangular shape, while for *Device 2* is the opposite. For the electro-optical frequency response, presented in figure 3.26b, the rectangular shape has better bandwidth in both devices, especially in *Device 1*, while for *Device 2* it is almost

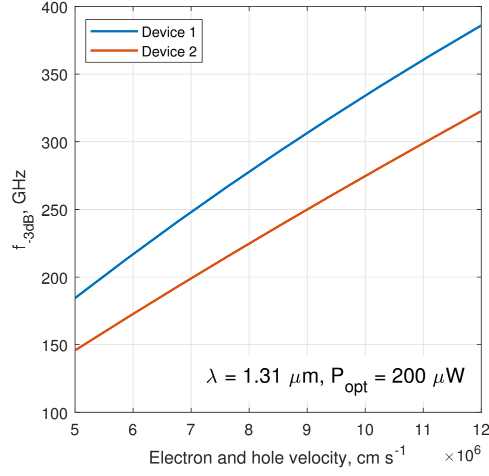
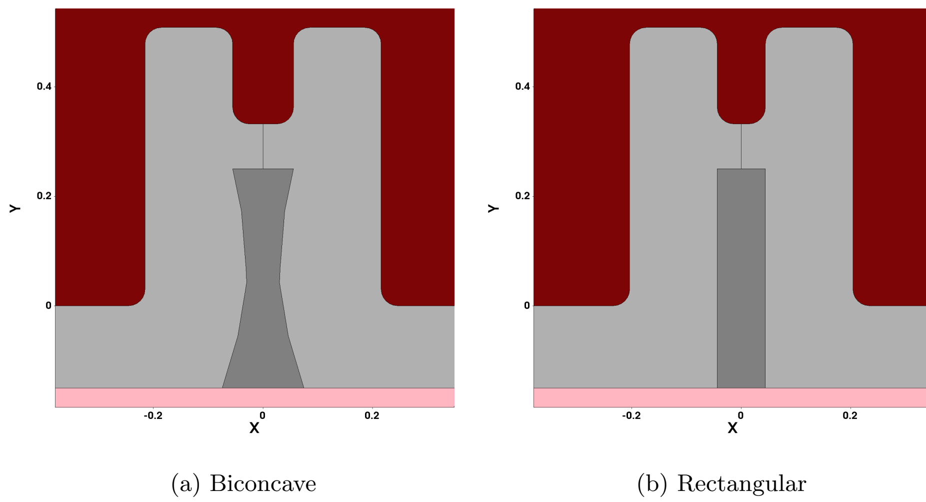


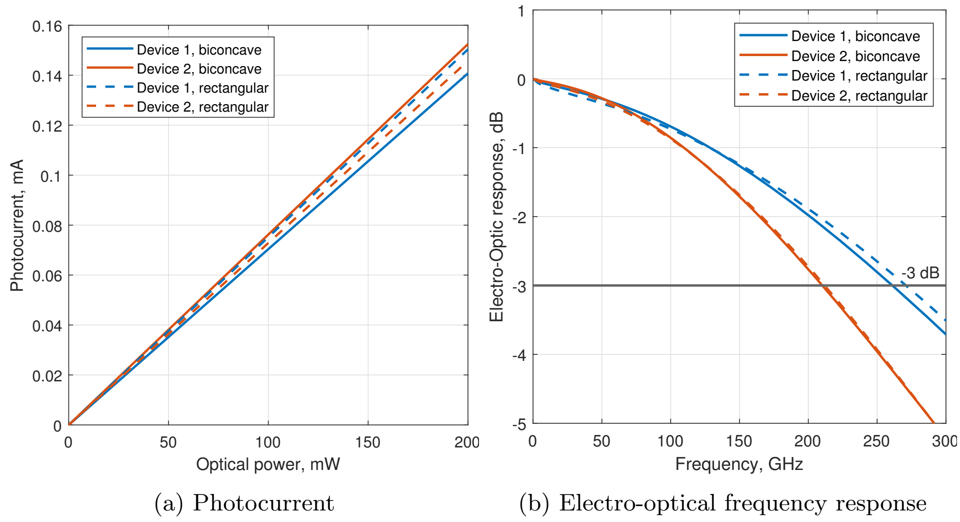
Figure 3.24: Dependence of the cut-off frequency from the saturation velocity of electrons and holes. The same value of saturation velocity was used for both carriers.



(a) Biconcave

(b) Rectangular

Figure 3.25: Comparison between the biconcave and rectangular geometries for the germanium absorber. The left figure is taken from the simulation of *Device 1*, while the figure on the right is obtained with effective width from the same device.



(a) Photocurrent (b) Electro-optical frequency response

Figure 3.26: Photocurrent and Electro-optical frequency response comparison between the rectangular and biconcave shape for the germanium region.

Device	Responsivity (A/W)	Cut-off frequency (GHz)
1, biconcave	0.704	261.1
1, rectangular	0.752	270.9
2, biconcave	0.762	210.2
2, rectangular	0.729	211.8

Table 3.6: Responsivity and cut-off frequency comparison between the rectangular and biconcave shape for the germanium region.

equal. A numerical comparison is presented in table 3.6 in the form of responsivity and cut-off frequency. The reason why the rectangular-shaped germanium has a better responsivity in *Device 1* with respect to *Device 2* is probably because of the presence of avalanche generation which starts to give a little increment in the photocurrent without impacting particularly the bandwidth that remains higher than the biconcave counterpart for *Device 1*. This is probably due to the transit time limit contributions of the top and bottom sections in the biconcave case that have more weight in limiting the bandwidth than the avalanche effect in the rectangular case.

From the simulations, there appears to be no particular advantage in the biconcave shape with respect to the rectangular one. Apparently, it is also slightly detrimental. Moreover, the mention of the optical power density being more concentrated in the thinner region of the biconcave shape seems to be not completely true. In figure 3.27 the distribution of the optical generation rate along the z-direction in the center of the absorber is

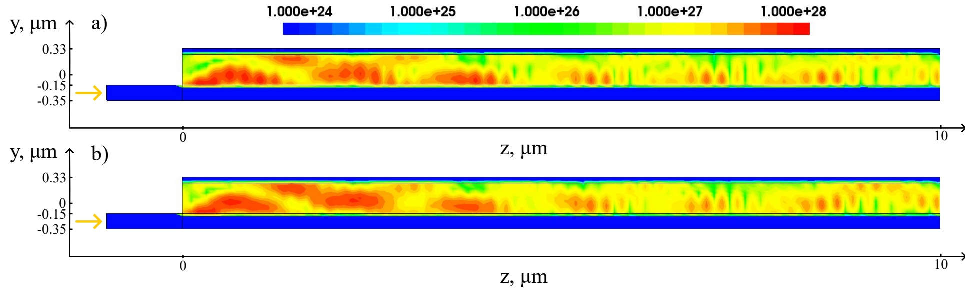
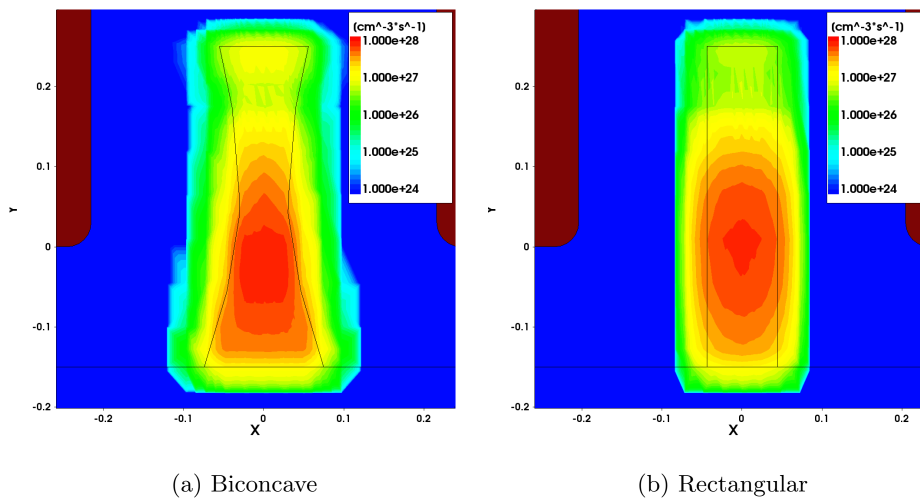


Figure 3.27: Distribution of the optical generation rate in *Device 1* in the yz-plane for $x = 0$. a) case with the biconcave germanium. b) case with the rectangular germanium.

presented and shows that, for both geometries, in the initial few micrometers where the majority of the photogeneration happens, there is no specific preferential position in the vertical (y) direction of the absorber. Figure 3.28 shows the distribution in the xy-plane cross-section of the device for $z = 0.6\,\mu\text{m}$ which is the position showing the higher concentration of optical generation in the center of the device, which is still relatively limited to the lower section of the germanium, especially for the biconcave shape that seems to limit the propagation of the light over the geometric bottleneck in the center.

Even if none of the simulation results seems to be in favor of the novel shape proposed by Lischke et al., it is also true that the effective width used for the comparison could be not adequate since it is possibly more complicated to produce a thinner rectangular germanium region than using the equivalent biconcave germanium. Therefore, further studies with more information about the real samples would be necessary to determine if this shape can improve or not the device's performance.



(a) Biconcave

(b) Rectangular

Figure 3.28: Distribution of the optical generation rate in *Device 1* in the xy-plane for $z = 0.6\,\mu\text{m}$. a) case with the biconcave germanium. b) case with the rectangular germanium.

Chapter 4

Conclusions

In conclusion, this work presented a 3D multiphysics simulation framework based on Drift-Diffusion and Finite-Difference-Time-Domain methods for lateral waveguide *pin* photodetectors (LWPD) capable of predicting accurately the device bandwidth even in sophisticated devices and bandwidth higher than 200 GHz, with a mean error of 8 %. This demonstrated the simulation framework's capability to simulate LWPD devices with complex geometries and high bandwidth with good reliability.

A simulation gives also much more versatility in studying the device performance with respect to direct measurements on a produced sample, as shown in the results section (chapter 3), permitting analysis and optimization of the device by modifying both the geometry and material's electrical and optical properties. Many critical points were highlighted in this work, such as the need for exhaustive and precise data on the material's properties to correctly simulate the device's dark current and photocurrent (responsivity). The dark current is particularly affected by the presence of defects and strain caused by the heterojunction with silicon, the diffusion of doping on the germanium, and the production process. The photocurrent depends mainly on the absorption coefficient, which lacks precise data in the literature for the C-Band ($\lambda = 1.55$ microm), and on the avalanche effect that gives a not negligible contribution in devices with a thin absorption region such as the one studied in this work. The electro-optical bandwidth, instead, is mainly affected by the geometry and the carrier saturation velocity, but thanks to the smaller number of not controllable parameters, it is much more reliable in simulation results as has been demonstrated in this work.

Many aspects were investigated, such as the effects on the device of the applied voltage bias, the variation of the germanium region geometry, differences in the responsivity for different sets of data for the absorption coefficient of the germanium, and variation in the carrier transport parameters, such as the saturation velocity.

A more extensive study on the saturation velocity effects in the LWPD's

electro-optical bandwidth was presented in the 23rd International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD) based on the same simulation framework developed in this thesis work [28].

Appendix A

Electronic band structure and effective mass approximation

The electronic band structure of a material is defined by its atomic structure and composition. A semiconductor is a crystalline material, implying a periodicity in its atomic structure and the consequent definition of a unit cell in the direct space[1]. The periodic arrangement of atoms also implies a periodicity in the total Coulomb potential obtained from the superposition of all contributions from the single atoms. The motion of electrons and holes in the material is ruled by quantum mechanics and *Schrödinger's equation* where the periodic potential is considered. From the solution of the equation, it is possible to obtain the energy dispersion relation $E(\underline{k})$ that defines the allowed energy states for the carriers in the function of the wavevector $\underline{k}$, a term related to the linear momentum $\underline{p}$ from the relation $\underline{k} = \hbar\underline{p}$. The band structure of the semiconductor is so obtained by representing the dispersion relation in the reciprocal space[2]. Thanks to the periodicity, the band structure can be limited to the *Brillouin First Zone*, which is the equivalent in the reciprocal space of the unit cell. The band structure is composed of multiple bands subdivided into two groups, the *conduction bands* and the *valence bands*, divided by a region with no allowed energy states, called *band gap*. This region is specifically defined as the distance between the minimum of the conduction band and the maximum of the valence band. The complete band structure of a material is very complicated and often it is not necessary to consider it in its entirety to study the electron transport of a semiconductor. In fact, the higher contribution to the transport is given

[1]Direct space, or direct lattice, is the real coordinate space $\mathbb{R}^n$ where the dimensions of the material's lattice structure are defined.

[2]Reciprocal space, or reciprocal lattice, is the $\mathbb{K}^n$ space generated using as a base the wavevector $\underline{k}$ components.

by the energy states near the band gap, which are positioned in the minimum and maximum of the conduction and valence band respectively. This means that it is possible to approximate the entire band structure considering only a fitting in these two regions. The maximum and minimum are approximated with a parable using a second-order Taylor expansion. This approximation is known as *effective mass approximation*. The two parables are used to define a parameter that represents the transport properties of the material, and that will characterize the motion of the respective carriers, called *effective mass*. The effective mass can be defined starting from the definition of momentum for a free electron $\underline{p} = m_n \underline{v}_n$, where $\underline{v}_n$ is the speed of the electron, and m_n is the rest mass. The energy of a free electron is defined as:

$$E(\underline{k}) = \frac{\hbar^2 k^2}{2m_n} \tag{A.1}$$

The idea behind the effective mass approximation is to treat electrons and holes as if they were in free space (*free electron gas*), but with a corrected value of rest mass that accounts for the properties of the material and its periodic potential. This is possible using the definition of electron velocity in a crystal lattice periodic potential (whose demonstration can be found in [29]):

$$v_n(\underline{k}) = \frac{1}{\hbar} \nabla_{\underline{k}} E_c(\underline{k}) \tag{A.2}$$

so, putting together the definition of crystal momentum with the definition of linear momentum $\underline{p} = \hbar \underline{k} = m_n \underline{v}_n$, and considering the new definition for electron velocity written above:

$$\hbar \underline{k} = m_n \underline{v}_n \tag{A.3a}$$

$$\hbar \underline{k} = \frac{m_n^*}{\hbar} \nabla_{\underline{k}} E_c(\underline{k}) \tag{A.3b}$$

$$(m_n^*)^{-1} = \frac{1}{\hbar^2} \nabla_{\underline{k}} E_c(\underline{k}) \tag{A.3c}$$

where m_n becomes m_n^* since now we are considering no more a free electron but an electron interacting with the periodic potential. The equation (A.3c) is the definition of electron effective mass. An equivalent reasoning can be applied to find the hole effective mass, so the definitions for the electron and hole effective masses are:

$$(m_n^*)^{-1} = \frac{1}{\hbar^2} \nabla_{\underline{k}} E_c(\underline{k}) \quad (m_p^*)^{-1} = \frac{1}{\hbar^2} \nabla_{\underline{k}} E_v(\underline{k}) \tag{A.4}$$

where $E_c(\underline{k})$ and $E_v(\underline{k})$ are the parabolic approximations of the dispersion relation for the conduction band and the valence band respectively:

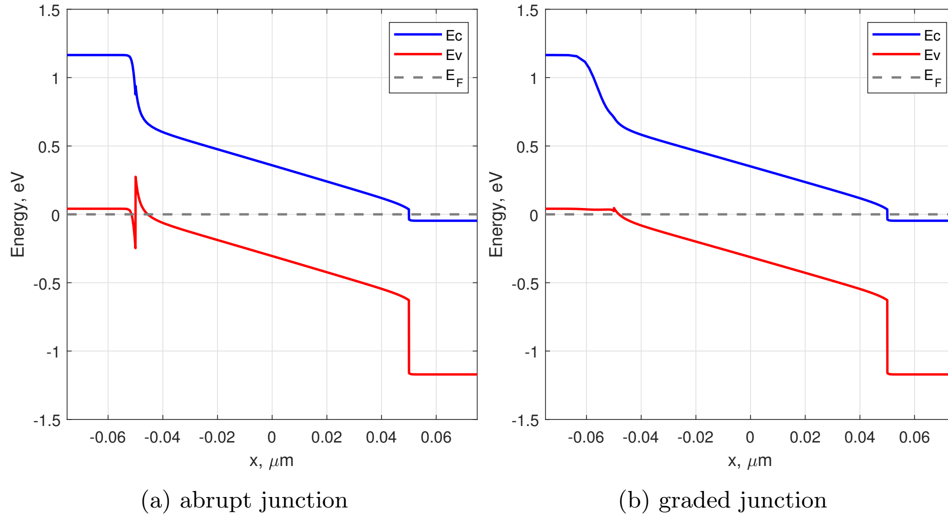
$$E_n - E_c \approx \frac{\hbar^2 k^2}{2m_n^*} \quad E_v - E_p \approx \frac{\hbar^2 k^2}{2m_p^*} \tag{A.5}$$

where E_n is the electron energy and E_p the hole energy.

Appendix B

Heterojunction between Ge and Si

Given the difference in band gap between germanium and silicon, a discontinuity is generated in the interface for both the conduction and valence bands (Type II heterojunction). However, while the discontinuity in the conduction bands gives no problems, the one in the valence band generates a barrier hindering the convergence of the simulation out of equilibrium (figure B.1a). Fortunately, due to the multiple heating processes needed during the production, there is a bit of diffusion of atoms between the two materials that relaxes slightly this discontinuity, removing at least partially the barrier. In the simulation, the doped silicon regions and germanium region are simulated as a silicon-germanium alloy with a molar fraction profile set to zero in the silicon and to one on the germanium. The transition in the interface behaves like an Error function to have a smooth variation between the two materials as can be observed in figure B.2. This removes completely the barriers in the valence band solving the convergence problems for the Drift-Diffusion simulation (figure B.1b).



(a) abrupt junction

(b) graded junction

Figure B.1: Band diagram of the device considering a germanium width of 100 nm centered in $x = 0$. The left figure shows the case with an abrupt junction and the consequent discontinuity. In the right figure, the graded junction solves the problem, removing the barrier in the valence band.

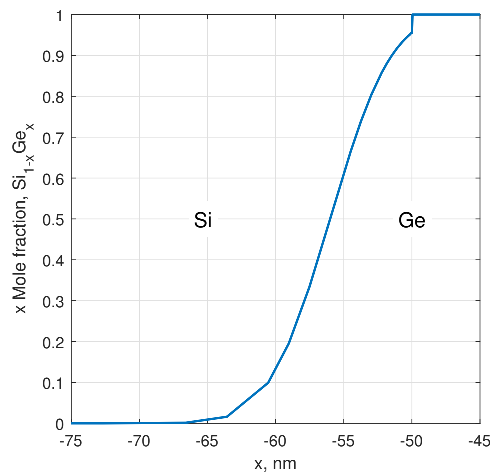


Figure B.2: Molar fraction profile in the interface between silicon and germanium.

Appendix C

Simulation mesh

A simulation mesh is a spatial map of points in which the simulation is performed. every point is associated with a spatial domain. In one dimension the associated domains are segments, in two dimensions are triangles, and in three dimensions are tetrahedra. Not all regions of the device need to be mapped with the same density of points, but instead, it is better to have a non-uniform mesh to reduce the computational weight and to focalize the simulation only where there are strong variations in the various physical quantities. This is also required to make the simulation converge. Moreover, these are also often the regions where the phenomena we are interested in observing happen. For this project, the complicated shape of the germanium region and the presence of a mole fraction profile add an ulterior complication to the mesh definition. A cross-section of the device (figure C.1) shows the distribution of the mesh points that increase in density near the junction to reproduce correctly the molar fraction profile. A denser mesh is maintained for the entire region of germanium since it is the focal part of the photodiode. A particularly dense mesh can be seen as well in the small silicon region over the germanium. It is implemented to represent correctly the small Error function profiles of both p and n doping introduced to simulate the slight diffusion of doping in the intrinsic silicon due to the thermal processes used during the production. The optical silicon waveguide, the silicon substrate, and the metal contacts have a constant rough mesh to decrease the simulation weight since they are not the focus point of the simulation. In the z-direction, the mesh is even more lax since there is no variation in the electrical properties.

The FDTD method, used to simulate the optical propagation of light, uses a different mesh with respect to the electrical simulation. The Yee's algorithm used in FDTD discretizes the electric and magnetic fields in space in a cubic grid, which means that the most natural choice is a mesh that is also cubic. Differently from the mesh for the electrical problem, the optical mesh cannot be easily refined in the different regions of the device, so it

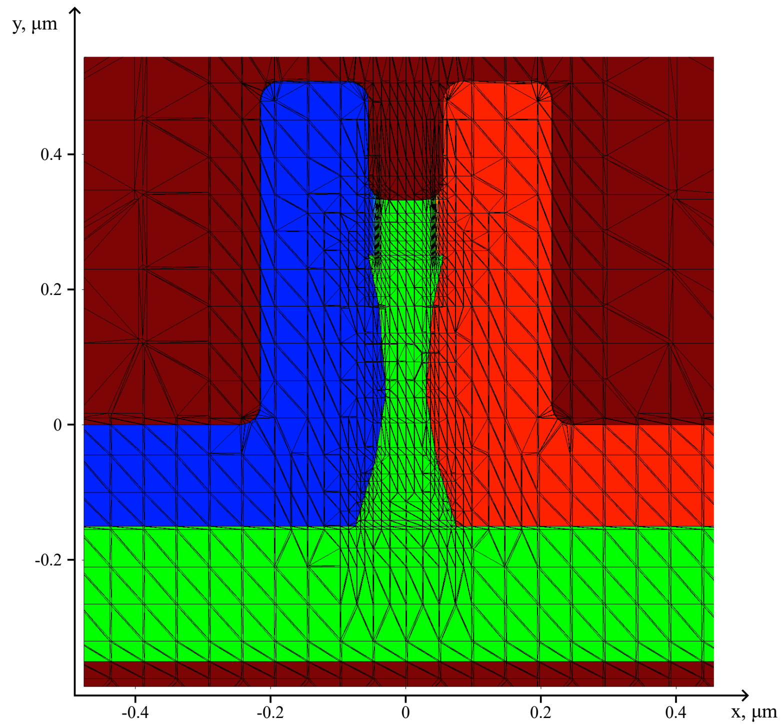


Figure C.1: Simulation mesh of *Device 1* in the *pin* junction for a cross-section in the xy-plane.

remains equal in size in the entire simulation domain.

Bibliography

[1] J. Michel, J. Liu, and L. C. Kimerling, "High-performance Ge-on-Si photodetectors," *Nature Photon.*, vol. 4, no. 8, pp. 527–534, 2010.

[2] L. Virot, D. Benedikovic, B. Szelag, C. Alonso-Ramos, B. Karakus, J.-M. Hartmann, X. Le Roux, P. Crozat, E. Cassan, D. Marris-Morini *et al.*, "Integrated waveguide PIN photodiodes exploiting lateral Si/Ge/Si heterojunction," *Opt. Express*, vol. 25, no. 16, pp. 19 487–19 496, Aug. 2017.

[3] T. Yin, R. Cohen, M. M. Morse, G. Sarid, Y. Chetrit, D. Rubin, and M. J. Paniccia, "31 GHz Ge *n-i-p* waveguide photodetectors on silicon-on-insulator substrate," *Opt. Express*, vol. 15, no. 21, pp. 13 965–13 971, Oct. 2007.

[4] D. Zhu, J. Zheng, Y. Qamar, O. Martynov, F. Rezaie, and E. Preisler, "A high performance Ge PIN photodiode compatible with high volume silicon photonics production processes," in *2018 IEEE 15th International Conference on Group IV Photonics (GFP)*, 2018.

[5] Y. Zhang, S. Yang, Y. Yang, M. Gould, N. Ophir, A. E.-J. Lim, G.-Q. Lo, P. Magill, K. Bergman, T. Baehr-Jones *et al.*, "A high-responsivity photodetector absent metal-germanium direct contact," *Opt. Express*, vol. 22, no. 9, pp. 11 367–11 375, May 2014.

[6] L. Virot, L. Vivien, A. Polzer, D. Marris-Morini, J. Osmond, J. M. Hartmann, P. Crozat, E. Cassan, C. Baudot, C. Kopp *et al.*, "40 Gbit/s germanium waveguide photodetector on silicon," in *Silicon Photonics and Photonic Integrated Circuits III*, vol. 8431, 2012, p. 84310A.

[7] H. Chen, P. Verheyen, P. D. Heyn, G. Lepage, J. D. Coster, S. Balakrishnan, P. Absil, W. Yao, L. Shen, G. Roelkens, and J. V. Campenhout, "-1 V bias 67 GHz bandwidth Si-contacted germanium waveguide p-i-n photodetector for optical links at 56 Gbps and beyond," *Opt. Express*, vol. 24, no. 5, pp. 4622–4631, Mar. 2016.

[8] S. Lischke, A. Peczek, J. S. Morgan, K. Sun, D. Steckler, Y. Yamamoto, F. Korndörfer, C. Mai, S. Marschmeyer, M. Fraschke, A. Krüger,

A. Beling, and L. Zimmermann, "Ultra-fast germanium photodiode with 3-dB bandwidth of 265 GHz," *Nature Photon.*, pp. 1–7, 2021.

[9] J. Michel, S. J. Koester, J. Liu, X. Wang, M. Geis, Steven J. Spector, M. E. Grein, Jung U. Yoon, T. M. Lyszczarz, and N.-N. Feng, "Photodetectors," in *Handbook of Silicon Photonics*, L. Vivien and L. Pavesi, Eds. Boca Raton, FL: CRC Press, 2013, ch. 10, pp. 479–552.

[10] G. Ghione, *Semiconductor Devices for High-Speed Optoelectronics.* Cambridge, U.K.: Cambridge University Press, 2009.

[11] ——, *Dispositivi per la microelettronica.* Milano: McGraw-Hill, 1998.

[12] *Sentaurus Device User Guide. Version N-2017.09*, Synopsys, Inc., Mountain View, CA, Sep. 2017.

[13] A. Jüngel, *Transport Equations for Semiconductors*, 1st ed. Heidelberg: Springer Berlin, 2019.

[14] A. Fick, "Ueber diffusion," *Ann. Phys.*, vol. 170, no. 1, pp. 59–86, 1855.

[15] F. Bonani, S. D. Guerrieri, G. Masera, and G. Piccinini, *Dispositivi e tecnologie elettroniche*, 1st ed. Torino: CLUT, 2007.

[16] G. Masetti, M. Severi, and S. Solmi, "Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon," *IEEE Trans. Electron Devices*, vol. 30, no. 7, pp. 764–769, Jul. 1983.

[17] C. Canali, G. Majni, R. Minder, and G. Ottaviani, "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature," *IEEE Trans. Electron Devices*, vol. 22, no. 11, pp. 1045–1047, Aug. 1975.

[18] K. Hutter and K. Jöhnk, *Continuum Methods of Physical Modeling: Continuum Mechanics, Dimensional Analysis, Turbulence.* Springer Berlin Heidelberg, 2004.

[19] *RSoft FullWAVE User Guide, v2019.09*, Synopsys, Inc., Optical Solutions Group, Ossining, NY, 2019.

[20] M. S. Wartak, *Computational Photonics: An Introduction with MATLAB.* Cambridge University Press, 2013.

[21] A. Taflove and S. C. Hagness, Eds., *Computational electrodynamics: the finite-difference time-domain method*, 3rd ed. Norwood, MA: Artech House, 2005.

[22] D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. Cambridge University Press, 2017.

[23] J. A. Kong, *Electromagnetic Wave Theory.* Cambridge: EMW Publishing, 2008.

[24] V. Sorianello, L. Colace, N. Armani, F. Rossi, C. Ferrari, L. Lazzarini, and G. Assanto, "Low-temperature germanium thin films on silicon," *Opt. Mater. Express*, vol. 1, no. 5, pp. 856–865, Sep. 2011.

[25] M. G. C. Alasio, M. Vallone, A. Tibaldi, S. Namnabat, D. Adams, P. Gothoskar, F. Forghieri, G. Masini, F. Bertazzi, G. Ghione, M. Gioannini, and M. Goano, "Towards 200 Gbit/s low-bias-operation vertical Ge-on-Si waveguide photodetectors," *Optica Open*, Jul. 2023.

[26] M. G. C. Alasio, P. Franco, A. Tibaldi, F. Bertazzi, S. Namnabat, D. Adams, P. Gothoskar, G. Masini, F. Forghieri, G. Ghione, and M. Goano, "3D multiphysics transient modeling of vertical Ge-on-Si *pin* waveguide photodetectors," in *22nd International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD 2022)*, online, Sep. 2022, pp. 5–6.

[27] M. G. C. Alasio, M. Goano, A. Tibaldi, F. Bertazzi, S. Namnabat, D. Adams, P. Gothoskar, F. Forghieri, G. Ghione, and M. Vallone, "Ge-on-Si waveguide photodetectors: multiphysics modeling and experimental validation," in *21st International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD 2021)*, online, Sep. 2021, pp. 37–38.

[28] M. G. C. Alasio, M. Zhu, A. Fronteddu, A. Cardinale, A. Ballarati, E. Bellotti, G. Ghione, A. Tibaldi, F. Bertazzi, M. Vallone, and M. Goano, "Modeling the electronic transport in FinFET-like lateral Ge-on-Si *pin* waveguide photodetectors for ultra-wide bandwidth applications," in *23rd International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD 2023)*, Torino, Italy, Sep. 2023, pp. 107–108.

[29] N. W. Ashcroft and N. D. Mermin, *Solid State Physics.* Philadelphia: Saunders, 1976.