Frequency Response Optimization of Dual Depletion InGaAs/InP PIN Photodiodes

J. M. Torres PEREIRA and João Paulo N. TORRES*

Instituto de Telecomunicações, Instituto Superior Técnico-Universidade de Lisboa, Lisboa, Portugal

*Corresponding author: João Paulo N. TORRES E-mail: joaotorres@ist.utl.pt

Abstract: The frequency response of a dual depletion p-i-n (PIN) photodiode structure is investigated. It is assumed that the light is incident on the N side and the drift region is located between the N contact and the absorption region. The numerical model takes into account the transit time and the capacitive effects and is applied to photodiodes with non-uniform illumination and linear electric field profile. With an adequate choice of the device’s structural parameters, dual depletion photodiodes can have larger bandwidths than the conventional PIN devices.

Keywords: PIN photodiodes; dual depletion; frequency response; modeling

1. Introduction

The photodiode is one of the most important devices in an optical communication system. The p-i-n (PIN) structure is widely used because it has good frequency response, low noise, and high sensitivity. For long wavelength applications, the materials most commonly used are: InP, for the N and P contact regions, and In0.53Ga0.47As, for the near intrinsic absorption region [1]. The choice of this particular ternary semiconductor is determined by its high absorption coefficient in the 1.0 µm – 1.6 µm wavelength range, because it is lattice matched to InP. Experimental work and theoretical work have shown that there is an upper limit for the PIN photodiode’s bandwidth, which has resulted from the interplay between the transit time and the capacitive effects [2, 3]. Additionally, the bandwidth-quantum efficiency product is nearly constant for a wide range of absorption region lengths, taking values of the order of tens of GHz. The conventional PIN photodiode is therefore not suitable for the last generation communication systems with high bit rate. In order to increase the device’s bandwidth and the bandwidth-quantum efficiency product, the capacitive effects should be reduced. Towards this goal, a new structure has been proposed which is basically a PIN structure with an additional layer of intrinsic InP next to the absorption layer of InGaAs [4]. In fact, the smaller permittivity of InP, compared with that of InGaAs, is responsible for a decrease in the structure’s total capacitance. Some preliminary work has shown that better bandwidths may be obtained with this structure and they depend on the location of the drift layer, relative to the contacts, and on the direction of the incident light [5]. It was reported that the best structures have the drift layer placed next to the N contact and the light should be incident on the N side. The published results assume that the photo
generated carriers move with their drift saturation velocity in the absorption and drift regions. In this work, a very general model is used which enables us to compute the device’s bandwidth for any electric field profile in the absorption and drift region, i.e., for spatial varying carrier’s velocities. The results may be used to optimize the device’s bandwidth regarding its dimensions and bias voltage.

2. Device structure

The structure under investigation is shown in Fig. 1. It consists of two highly doped p⁺ and n⁺ contacts of InP, one region of intrinsic InP (D) with length $\ell_d$, placed next to the n⁺ contact, and the other region of lightly doped In$_{0.53}$Ga$_{0.47}$As (A) with length $\ell_a$. The light, incident on the n⁺ side, being in the wavelength range 1.0 $\mu$m – 1.6 $\mu$m, is only absorbed by the ternary semiconductor because the photon energy is lower than the energy band gap of InP. Under normal operation, the device is reverse biased in such a way that the electric field extends over the absorption and drift regions, A and D layers, respectively, and is responsible for the transport of the optical generated electrons to the n⁺ contact and holes to the p⁺ contact.

3. Modeling

The frequency response of the dual depletion PIN photodiode is computed by considering the transit time and the capacitive effects. The transit time effects are associated with the transport, by the electric field in the absorption and drift regions, of the optical generated electrons and holes. These effects depend mainly on the carrier’s drift velocity and the length of the absorption and drift regions. The capacitive effects are due to the junction capacitance and other parasitic capacitances related to the contact pad and package. By reducing the length of the device, the transit time decreases, and the capacitance increases. Therefore, for short devices, the bandwidth is mainly dependent on the capacitive effects, whereas, for long devices, it is transit time limited. There will be an optimum length where the bandwidth reaches a maximum. Around this maximum, both the capacitive and the transit time effects play important roles.

3.1 Transit time effects

Regarding the transit time effects, the frequency response is calculated by combining the coefficients derived from a matrix formulation, following the analytical solution of the continuity equations in the drift and absorption regions [6]. The analytical solutions can be obtained only for constant drift carrier velocities, i.e., when the electric field is constant. In general, this is not the case, and therefore a new approach has been implemented, which enables us to treat situations involving arbitrary electric field profiles [7]. In this case, the particular region is divided into several layers, with each layer being assigned a constant electric field. Thus, it is possible to obtain analytical solutions for each of these layers, and by combining the coefficients of every layer, it is possible to calculate the frequency response of the whole structure. The accuracy of the results depends on the number and size of the layers.

In this work, the electric field is assumed to be constant in the drift region and to vary linearly in the absorption region. The continuity of the displacement vector at the boundary of the two regions establishes the relationship between the corresponding electric fields.

It is also assumed, in our derivation, that diffusion and trapping of carriers are negligible and the light intensity is not high enough to cause
saturation or screening effects.

In the frequency domain, the electron and hole current densities in the \( i \)th layer, \( J_p(x,\omega) \) and \( J_n(x,\omega) \), respectively, obey the following continuity equations [7]:

\[
\frac{i\omega}{v_{in}} J_n(x,\omega) = \frac{d J_n(x,\omega)}{dx} + G_i(x,\omega) \tag{1}
\]

\[
\frac{i\omega}{v_{ip}} J_p(x,\omega) = -\frac{d J_p(x,\omega)}{dx} + G_i(x,\omega) \tag{2}
\]

where \( v_{in} \) and \( v_{ip} \) are the electron and hole drift velocities, respectively, and \( G_i(x,\omega) \) refers to the electron-hole optical generation rate given by

\[
G_i(x,\omega) = \begin{cases}
0 & (0 \leq x \leq \ell_d) \\
q \alpha \phi e^{-\alpha (x-\ell_d)} & (\ell_d \leq x \leq \ell_a + \ell_d)
\end{cases} \tag{3}
\]

The parameter \( \alpha \) is the absorption coefficient, \( q \) is the magnitude of the electron charge, \( \phi \) is the amplitude of the sinusoidal input optical flux component, and \( \ell_d \) is the width of the absorption region.

By solving the continuity (1) and (2), the \( i \)th layer may then be represented by a set of linear response coefficients \( T_i, S_i, R_i \) and \( D_i \). The quantities \( T_i, S_i \) are related to the electron and hole current densities through the equations

\[
\begin{bmatrix}
J_n(x_i) \\
J_p(x_i)
\end{bmatrix} = T_i \begin{bmatrix}
J_n(x_{i-1}) \\
J_p(x_{i-1})
\end{bmatrix} + S_i \tag{4}
\]

with

\[
T_i = \begin{bmatrix}
T_{nn} & T_{np} \\
T_{pn} & T_{pp}
\end{bmatrix} \quad \text{and} \quad S_i = \begin{bmatrix}
S_{np} \\
S_{pp}
\end{bmatrix} \tag{5}
\]

The matrix \( T \) is called the current transfer matrix, and the vector \( S \) represents the contribution to the current due to the optical sources. The quantities \( R_i \) and \( D_i \) are obtained from the equation

\[
p_i(\omega) = S_i \begin{bmatrix}
J_n(x_{i-1}) \\
J_p(x_{i-1})
\end{bmatrix} + D_i \tag{6}
\]

where \( p(\omega) \) is called the partial electrode current given by

\[
p_i(\omega) = \int_{x_i}^{x_{i+1}} \left[ J_n(x,\omega) + J_p(x,\omega) \right] dx \tag{7}
\]

The quantity \( D_i \) is a scalar related to the optical sources whereas \( R^T_i = [R_{pp} \quad R_{np}] \) determines the proportionality between the left-hand terminal currents and \( p(\omega) \).

The four linear response coefficients may be used now to compute the corresponding response coefficients of the multilayer structure by following a simple set of rules:

\[
\begin{align*}
T_{i+1,j} &= T_{i+1,i} T_j \\
S_{i+1,j} &= S_{i+1,i} + T_{i+1,i} S_j \\
R^T_{i+1,j} &= R^T_{i+1,i} T_j + R^T_j \\
D_{i+1,j} &= D_{i+1,i} + D_j
\end{align*} \tag{8}
\]

where subscripts \((i), (i+1),\) and \((i+1, i)\) refer to the \( i \)th layer, \((i+1)\)th layer, and the union of the two layers, respectively.

The frequency response is obtained from [6]

\[
I(\omega) = \delta(\omega)/(`\ell_a + \ell_d`) \tag{9}
\]

where the response scalar \( \delta(\omega) \) is given by

\[
\delta(\omega) = D - R_s S_i / T_{nn} \tag{10}
\]

The quantities \( D, R_s, S_i, \) and \( T_{nn} \) for the multilayer structure, are obtained by repeated application of (8).

Solving (1) and (2) using (3) and by relating the current densities as in (4), we are able to obtain \( T \) and \( S \). For the drift region,

\[
T_D = \begin{bmatrix}
e^{-i\omega \ell_d} & 0 \\
0 & e^{i\omega \ell_a}
\end{bmatrix} \quad S_D = 0 \tag{11}
\]

and for the \( i \)th layer of the absorption layer,

\[
T_i = \begin{bmatrix}
e^{-i\omega \ell_i} & 0 \\
0 & e^{i\omega \ell_p}
\end{bmatrix}
\]

\[
S_i = q\alpha \phi \ell_i e^{-\alpha \ell_i} \begin{bmatrix}
f(i\omega \tau_{ip} - \alpha \ell_i) \\
-f(i\omega \tau_{ip} - \alpha \ell_i)
\end{bmatrix} \tag{12}
\]

with \( \tau_{in} = \ell_d / v_{in}, \tau_{ip} = \ell_d / v_{ip}, \tau_{in} = \ell_a / v_{in}, \tau_{ip} = \ell_a / v_{ip}, \) and \( f(\theta) = (1 - e^{-\theta})/\theta \) being \( \tau_{in}, \tau_{ip}, \tau_{en}, \tau_{ip} \) the electron and hole transit time in the drift and \( i \)th layer.

The quantities \( R_i \) and \( D_i \) are obtained from (6) taking into account (7). For the drift region,

\[
R_D = \ell_d \begin{bmatrix}
f(i \omega \tau_{ip}) \\
f(-i \omega \tau_{ip})
\end{bmatrix} \quad D_D = 0 \tag{13}
\]

and for the \( i \)th layer of the absorption layer,
Photonic Sensors

\[ R_j = \ell_i \left[ \frac{f(i\omega \tau_n)}{f(-i\omega \tau_n)} \right] \]

\[ D_j = qa\phi \ell_j^2 e^{\alpha(x_i-x_j)} \times \left[ \frac{f(\alpha \ell_i) - f(-i\omega \tau_n)}{\alpha \ell_i + i\omega \tau_n} \right] \]

(14)

In the drift region, the electron and hole velocities, for each electric field value, are taken or extrapolated from a table obtained from typical InP velocity-field plots [8]. For the InGaAs, the electron and hole drift velocities are related to the electric field by two empirical expressions [9]:

\[ v_n(E) = \left( \mu_n E + \beta v_n E^\gamma \right) \left( 1 + \beta E \right) \]

\[ v_p(E) = v_p \tanh \left( \mu_p E / v_p \right) \]

(15)

where \( \mu_n \) and \( \mu_p \) are the electron and hole mobilities, \( v_n \) and \( v_p \) are the electron and hole saturation velocities, i.e., the drift velocities for high electric fields, and \( \beta \) and \( \gamma \) are adjusting constants, \( \beta = 7.4 \times 10^{-15} (m/V)^{2.5} \); \( \gamma = 2.5 \) (Fig. 2).

\[ E(x) = \frac{2U_d}{\ell_a^2} \left( x - \ell_d \right) + \left( \frac{U_i - U_d}{\ell_a} \right) \]

(16)

with \( U_d = qN\ell_d^2 / (2\varepsilon_n) \), \( U_d \) is called the punch-through voltage, \( U_i \) is the reverse bias voltage in the absorption region, \( N \) is the residual donor concentration in the absorption region, and \( \varepsilon_n \) is the InGaAs electric permittivity. A typical electric field profile along the drift and the absorption region is shown in Fig. 3.

\[ C_1 = \frac{C_D \times C_A}{C_D + C_A} \] with \( C_D = \varepsilon_D \frac{A}{\ell_d} \); \( C_A = \varepsilon_A \frac{A}{\ell_a} \)

(17)

where \( A \) is the cross sectional area of the device, and \( \varepsilon_D \) and \( \varepsilon_A \) are the electric permittivities of InP and In_{0.53}Ga_{0.47}As, respectively.

\[ I_0 = H_{RC} I_1 \] where \( H_{RC} \) is the transfer function associated with the capacitive effects which, neglecting \( R_d \), may be written as

Fig. 2 Electron and hole drift velocities versus electric field for In_{0.53}Ga_{0.47}As.

3.2 Capacitive effects

The capacitive effects are included in the model by using the small signal equivalent circuit of the PIN photodiode [2], as shown in Fig. 4. It consists of a series resistance \( R_S \), a leakage resistance \( R_L \), a load resistance \( R_L \), a junction capacitance \( C_1 \), and a parasitic capacitance \( C_p \). The junction capacitance takes into account the capacitances due to the drift, \( C_D \), and the absorption, \( C_A \), regions and may be written as

Fig. 3 Electric field profile along the photodiode structure.

Fig. 4 Equivalent circuit model of the PIN photodiode considering the series and leakage resistance.

The current source \( I_1 \) is directly related to the transit time limited frequency response given by (9), and the output current \( I_0 \) may be expressed as

\[ I_0 = H_{RC} I_1 \] where \( H_{RC} \) is the transfer function associated with the capacitive effects which, neglecting \( R_d \), may be written as
\[ H_{RC} = \frac{1}{-\omega^2 R_L R_p C_p C_i + i\omega(C_i(R_L + R_p) + R_L C_p)} + 1. \]  

(17)

The frequency response of the device may then be obtained by the product of (9) and (17) [2].

4. Results and discussion

The simulations use the material parameters from Table 1.

### Table 1 Material parameters at 300 K.

| Parameters                        | Units   | In0.53Ga0.47As | InP |
|-----------------------------------|---------|----------------|-----|
| Absorption coefficient \((\alpha)\) | m\(^{-1}\) | 1.15x10^6      | ----|
| Electron saturation velocity \((v_{ne})\) | m/s    | 6x10^4         | 8.11x10^4|
| Hole saturation velocity \((v_{ph})\) | m/s    | 4.8x10^4       | 8.11x10^4|
| Electron mobility \((\mu_n)\) | m\(^2\)V\(^{-1}\)s\(^{-1}\) | 1.05        | ----|
| Hole mobility \((\mu_p)\) | m\(^2\)V\(^{-1}\)s\(^{-1}\) | 0.042       | ----|
| Electric permittivity \((\varepsilon)\) | F/m    | 14.1 \(\varepsilon_0\) | 12.56 \(\varepsilon_0\) |

The photodiodes are assumed to have an absorption region with a donor concentration \(N = 10^{21}\) m\(^{-3}\). The devices are assigned an area of 5x10^{-10} m\(^2\), a parasitic capacitance of 13x10^{-15} F, and a load resistance of 50\(\Omega\) [2]. In the calculations, the absorption region is divided into 40 layers.

The frequency responses are represented in Figs. 5 and 6 and show the expected behavior of a low pass filter. The normalization frequency \(f_0\) is taken as 1 GHz. For a fixed total length of 3\(\mu\)m (Fig.5), the best frequency response is obtained for devices with an absorption region with a length that is about half of the total length. For larger absorption region lengths, the frequency response is determined by the transit time effects, whereas for shorter absorption region lengths, the capacitive effects become dominant. In both cases, a 3-dB bandwidth decrease is observed.

![Fig. 5 Frequency responses for \(\ell_a/\ell_t = 0.5\), \(U_1 = 10\) V, and several values of \(\ell_a/\ell_t\).](image)

![Fig. 6 Frequency responses for \(\ell_a/\ell_t = 0.5\), \(U_1 = 10\) V, and several values of \(\ell_a/\ell_t\).](image)

Fig. 5 Frequency responses for \(\ell_a = 3\) \(\mu\)m, \(U_1 = 10\) V, and several values of \(\ell_a/\ell_t\). \(f_0 = 1\) GHz.

In Fig.6, when the ratio \(\ell_a/\ell_t = 0.5\), the frequency response is better when the total length decreases from 3\(\mu\)m to 1\(\mu\)m which means that, in this case, the frequency response is mainly transit time limited, and for smaller \(\ell_t\), the transit time decreases.

Figure 7 shows the 3-dB device’s bandwidth, \(B\), as a function of the ratio \(\ell_a/\ell_t\), for a bias voltage across the absorption region \(U_1 = 10\) V and several values of the total length \(\ell_t\). This bias voltage value is chosen to ensure that both the drift and the absorption regions are fully depleted. For \(\ell_t \geq 1\) \(\mu\)m, the curves have a maximum for a certain value of \(\ell_a/\ell_t\), which means that the inclusion of a drift region increases the bandwidth compared with the value obtained for the conventional PIN device \((\ell_a/\ell_t \approx 1)\). A observed decrease in bandwidth for devices with \(\ell_t > 1\) \(\mu\)m shows that the transit time effects are dominant. For very short devices, \(\ell_t = 0.5\) \(\mu\)m, there is no visible maximum because the bandwidth is determined by the capacitive effects.

It is important to investigate how the bandwidth
relates to the responsivity and if the optimized double depletion structure is better than the conventional PIN structure.

\[ S = \frac{\lambda (\mu m)}{1.242} \left(1 - e^{-\alpha \ell_e}\right) \text{ (A/W)} \]  

(18)

which shows that, for a fixed wavelength, the responsivity increases with the length of the absorption region. Therefore, there is a compromise between the bandwidth and responsivity.

The plot bandwidth-responsivity is shown in Fig. 8. It can be seen that for the dual depletion structure, with the total length of 1 \( \mu m \), the bandwidth is always higher than the one obtained for the PIN device with the same responsivity. The results show that for an optimized device the dual depletion structure is better than the PIN structure with a responsivity that is less than 0.7 A/W. It can be seen that for the maximum bandwidth of 29 GHz, the responsivity is around 0.47 A/W.

Figure 9 shows the maximum bandwidth versus total length for devices with bias voltages \( U_1 = 6 \text{ V} \) and \( U_1 = 10 \text{ V} \). The bandwidth values in Fig. 9, for \( U_1 = 10 \text{ V} \), correspond to the maximum of the curves shown in Fig. 7, for several values of \( \ell_e \). The absolute bandwidth maximum is seen to be about 29 GHz and is obtained for a device 1 \( \mu m \) long with an absorption region width about half of that length.

By decreasing the voltage from 10 V to 6 V, the bandwidth only changes when \( \ell_e > 1 \mu m \), because below this value, it is mainly dependent on the capacitive effects. The observed higher bandwidth values, for lower voltages, may be explained in terms of the carrier velocity-electric field type of dependence for these materials (Fig. 2). This figure shows that lower electric field values, within a certain range, are responsible for higher electron velocities and therefore lower transit time.

Figure 10 shows the maximum bandwidth versus total length for devices with several cross-sectional areas. It is observed that when the area changes from 1000 \( \mu m^2 \) to 100 \( \mu m^2 \), the optimized bandwidth may change from 22 GHz up to 58 GHz due to a decrease in the device’s junction capacitance [3].
The contour plot of the constant bandwidth, related to $\ell_a$ and $\ell_d$, for $U_1 = 6$ V and $A = 500 \mu m^2$, is shown in Fig. 11. From this figure, it is seen that the maximum bandwidth of about 29 GHz may be obtained for devices approximately 1 $\mu m$ long with a 0.5 $\mu m$ absorption region length. For the same length, if the area of the device is decreased, the capacitance also decreases, and an increase in the device’s bandwidth is expected [3].

These results show a clear improvement compared with those for the conventional PIN [3].

5. Conclusions

The numerical simulation of the frequency response of dual depletion InGaAs/InP PIN photodiodes shows that, for a fixed total length of the device, there is a value of the absorption length that maximizes its bandwidth. This value is seen to be a consequence from the interplay between the transit time in the drift and absorption region and the capacitive effects. For short devices, the maximum bandwidth is seen to be mostly independent of bias voltage because the capacitive effects are dominant. The bandwidth may be optimized by an adequate choice of the absorption and drift region lengths. The optimized bandwidth values range from 22 GHz up to 58 GHz when the cross-sectional area of the device decreases from 1000 $\mu m^2$ to 100 $\mu m^2$, and they are obtained for $\ell_t = 1$ $\mu m$, $\ell_a = 0.5$ $\mu m$ and $\ell_t = 0.5$ $\mu m$, $\ell_a = 0.25$ $\mu m$, respectively. The optimized dual depletion structure is seen to have a larger bandwidth than the conventional PIN structure with the same responsivity.
Open Access  This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

[1] T. P. Pearsall, “Ga₀.₄₇In₀.₅₃As: a ternary semiconductor for photodetector applications,” IEEE Journal of Quantum Electronics, 1980, 16(7): 709–720.

[2] Y. G. Wey, K. Giboney, J. Bowers, M. Rodwell, P. Silvestre, P. Thiagarajan, et al., “110 GHz GaInAs/InP double heterostructure p-i-n photodetectors,” Journal of Lightwave Technology, 1995, 13(7): 1490–1499.

[3] C. M. C. Fernandes and J. M. T. Pereira, “Bandwidth modeling and optimization of PIN photodiodes,” in 2011 IEEE International Conference on Computer as a Tool (EUROCON), Lisbon, pp. 1–4, 2011.

[4] F. J. Effenberger and A. M. Joshi, “Ultrafast, dual-depletion region, InGaAs/InP p-i-n detector,” Journal of Lightwave Technology, 1996, 14(8): 1859–1864.

[5] C. C. Fernandes and J. T. Pereira, “Frequency response analysis of dual depletion PIN photodiodes,” in the 12th Portuguese-Spanish Conference on Electrical Engineering (XIICLEEE), Ponta Delgada, pp. 1–4, 2011.

[6] J. N. Hollenhorst, “Frequency response theory for multilayer photodiodes,” Journal of Lightwave Technology, 1990, 8(4): 531–537.

[7] J. M. T. Pereira, “Modelling the frequency response of p+iInP/n-InGaAs/n+InP photodiodes with an arbitrary electric field profile,” COMPEL – The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 2007, 26(4): 1114–1122.

[8] C. Hammar and B. Vinter, “Calculation of the velocity-field characteristic of n-InP,” Solid State Communication, 1972, 11(5): 751–754.

[9] M. Dentan and B. D. Cremoux, “Numerical simulation of the nonlinear response of a p-i-n photodiode under high illumination,” Journal of Lightwave Technology, 1990, 8(8): 1137–1144.

[10] J. B. Radunovic and D. M. Gvozdie, “Nonstationary and nonlinear response of a p-i-n photodiode made of a two-valley semiconductor,” IEEE Transaction on Electron Devices, 1993, 40(7): 1238–1244.