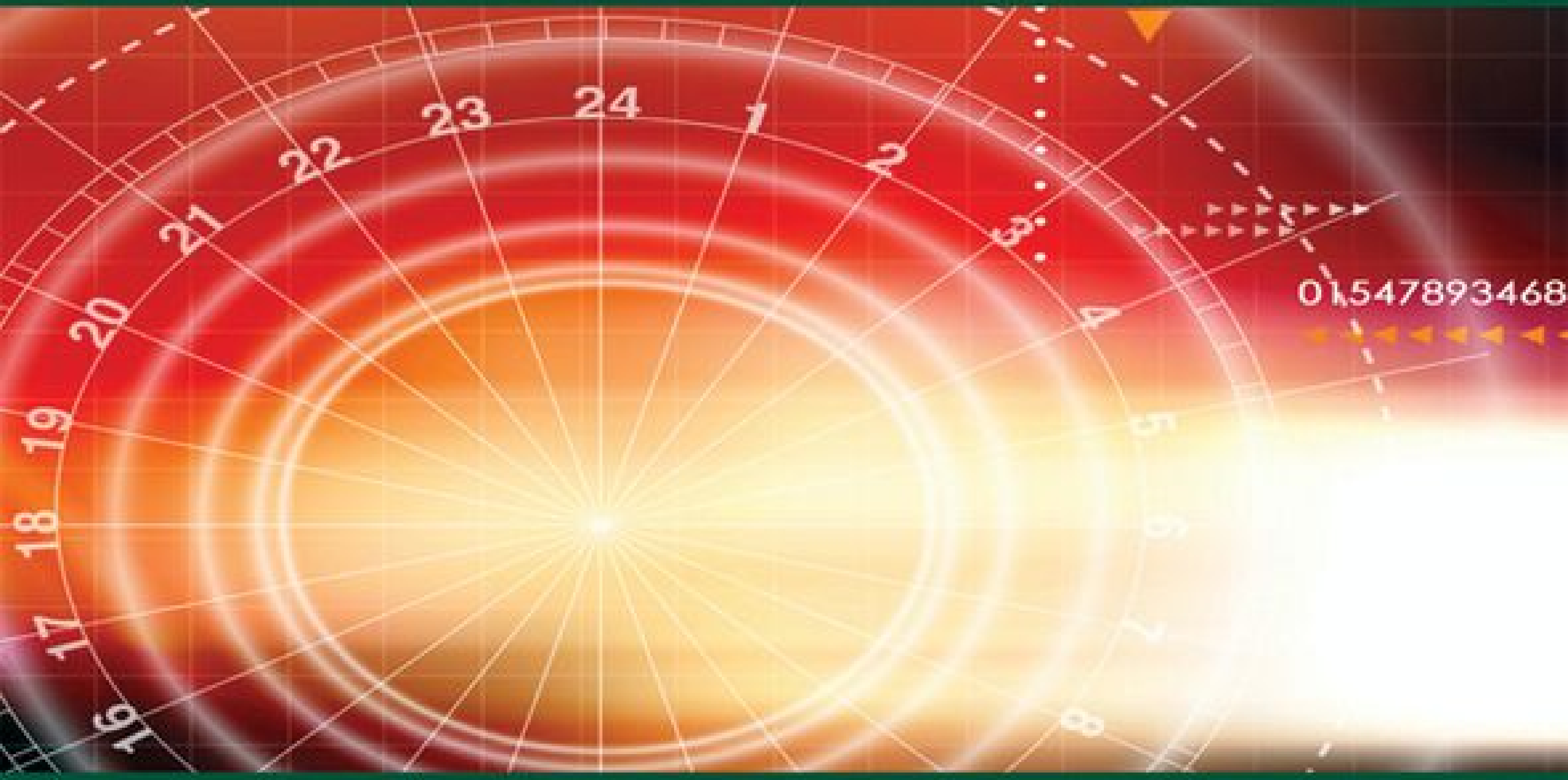


# Optoelectronic Sensors

Edited by
Didier Decoster and Joseph Harari

iSTE

WILEY

# Optoelectronic Sensors

# Optoelectronic Sensors

Edited by
Didier Decoster
Joseph Harari

ISTE

WILEY

First published in France in 2002 by Hermes Science/Lavoisier entitled: *Détecteurs optoélectroniques*

First published in Great Britain and the United States in 2009 by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

---

Library of Congress Cataloging-in-Publication Data

Détecteurs optoélectroniques. English
Optoelectronic sensors / edited by Didier Decoster, Joseph Harari.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84821-078-3
1. Optical detectors. 2. Image sensors. I. Decoster, Didier, 1948- II. Harari, Joseph, 1961- III. Title.
TK8360.O67D4813 2009
681'.25--dc22

2009011542

---

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-078-3

---

Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne.

# Table of Contents

# Preface

Photodetection is found in a large number of professional and mass-market systems. There are numerous applications including: fiber-based and free-space optical telecommunications, galvanic isolation, solar cells, proximity detectors, etc. All of these applications are based on the same process: the transformation of optical power into an electrical signal, with this signal needing to be as large as possible for an optical flux as weak as possible. When information needs to be transmitted quickly, potentially at very high speeds, the photodetector must react very fast. These basic considerations imply that a certain number of performance requirements should be met in order to satisfy the demands of the intended application. It is in this context that various photodetecting structures have been conceived: photoconductors, p-n and p-i-n photodiodes, avalanche photodiodes, phototransistors, Schottky photodiodes, MSM (metal semiconductor-metal) photodetectors. The range of wavelengths relevant to the application also plays an important role. It is often this, through the intermediary of the semiconductor bandgap, which dictates the nature of the material(s) used.

This book gathers together the most detailed and significant contemporary thinking on photodetection for wavelengths from the near-infrared to the ultraviolet[1]. Its content not only gives the reader

---

[1] The very particular nature of photodetection in the mid- and far-infrared requires specific treatment and thus is not covered here.

the grounding to design simple photodetectors with specified performance characteristics, but also discusses the state of the art in photodetection. Chapter 1 begins with an introduction to photodetection and the most well-known photodetector structures. Among these, some are more commonly used than others, or offer more exciting possibilities.

Because of this, each of Chapters 2 to 5 concentrates in great depth on a specific photodetector type. In sequence, the reader will be able to learn about:

– the PIN photodiode which is the photodetector on which the majority of systems are based (Chapter 2);

– the avalanche photodiode, which is a refinement of the PIN photodiode and which, under a sufficiently high bias voltage, allows us to achieve a gain (Chapter 3);

– the phototransistor, which is another means of obtaining gain in photodetection (Chapter 4);

– the MSM photodiode which is well suited to monolithic integrated circuits such as MMICs (Microwave Monolithic Integrated Circuits) (Chapter 5).

Chapter 6, however, is different. It is dedicated to ultraviolet photodetectors. Given the significant role that this wavelength range is likely to take in the near future, notably for environmental applications, this subject deserves a separate chapter. In this case, all types of photodetector are concerned, and it is the material which is the source of improvement, notably with the appearance of AlGaN materials with a very large bandgap.

Finally, Chapter 7 is entirely dedicated to noise, as this concept is absolutely fundamental for the photodetection of low-intensity signals.

The chapters of this book have all been written by specialists and we take this opportunity to thank them sincerely and warmly for their contributions.

Didier DECOSTER and Joseph HARARI

# Chapter 1

# Introduction to Semiconductor Photodetectors

## 1.1. Brief overview of semiconductor materials

A semiconductor material is a continuous crystalline medium characterized by an energy band structure corresponding, in the case of an infinite crystal, to a *continuum* of states (which, in practice, means that the characteristic dimensions of the crystal are significantly larger than the lattice parameter of the crystal structure; this applies as long as the crystal dimensions are typically larger than a few dozen nanometers). In general terms, the energy structure of a semiconductor consists of a valence band corresponding to molecular bonding states and a conduction band representing the molecular anti-bonding states. The energy range lying between the top of the valence band and the bottom of the conduction band is known as the forbidden band, or more commonly the bandgap. An electron situated in the valence band is in a ground state and remains localized to a particular atom in the crystal structure, whereas an electron situated in the conduction band exists in an excited state, in a regime where it interacts very weakly with the crystalline structure. What

---

Chapter written by Franck OMNES.

differentiates semiconductors from insulators is essentially the size of the bandgap: we refer to semiconductors where the bandgap of the material is typically less than or equal to 6 eV, and to insulators when the bandgap is more than 6 eV: above this, the solar spectrum arriving on the Earth's surface is unable to produce inter-band transitions of electrons situated in the valence band of the material. Semiconductor materials are mostly divided into two large classes: elemental semiconductors (group IV of the periodic table): silicon, germanium, diamond, etc. and compound semiconductors: IV-IV (SiC), III-V (GaAs, InP, InSb, GaN) and II-VI (CdTe, ZnSe, ZnS, etc.). Impurities can be introduced into the volume of the semiconductor material and can modify its electrical conduction properties, sometimes considerably. An impurity is known as a *donor* when it easily releases a free electron into the conduction band. The characteristic energy level of the impurity is therefore in the bandgap, slightly below the conduction band. For example, in the case of compound semiconductors in group IV of the periodic table such as silicon, the main donor impurities are those which, being from group V of the periodic table (arsenic, phosphorous, etc.), are substituted in place of a silicon atom in the crystal structure: since silicon is tetravalent, these atoms naturally form four covalent bonds with the silicon atoms around them, and also easily give up their surplus electron to the crystal structure. These electrons become free to move, subject to a weak activation energy provided by thermal agitation. In this case we refer to n-type doping. In the case of silicon, a group III element incorporated into the crystal structure of silicon naturally forms three covalent bonds around it, and then completes its own outer-shell electronic structure by capturing an electron from its fourth nearest-neighbor silicon atom, again subject to a weak thermal activation energy. Such an impurity is known as an *acceptor*, and doping with acceptors is known as p-type doping. A hole carrying a positive elementary charge and corresponding to a vacant energy state in the valence band is therefore left in the crystal structure of the silicon. In the case of III-V composites, the donors are mostly atoms from group IV (silicon) substituted in place of group III elements, or group VI elements (S, Se, Te) substituted in place of group V elements, and acceptors are group II (zinc, magnesium) substituted in place of group

III elements. In the case of II-VI composites, the most commonly-encountered donors belong to group VII (chlorine, etc.) substituted in place of group VI elements, and acceptors belong to either group I (lithium, etc.) or to group V (nitrogen, arsenic, phosphorous, etc). In this last case, the group V element is substituted in place of a group VI element in the semiconductor crystal structure, whereas group I acceptors are substituted in place of group II elements. The chemical potential, or Fermi energy, of an *intrinsic* semiconductor (i.e. one free from n and p impurities) is found in the middle of the bandgap of the material. When a moderate n-type doping is added, the Fermi level rises from the middle of the bandgap towards the conduction band, by an increasing amount as the level of doping rises. When the level of n-type doping becomes large, the Fermi level can cross the bottom of the conduction band and be found inside this band (Mott transition). The semiconductor then behaves like a metal and for this reason is called a semi-metal. In this case it is referred to as degenerate. In the case of p-type doping, the semiconductor is said to be degenerate when the Fermi level is below the top of the valence band.

## 1.2. Photodetection with semiconductors: basic phenomena

Photodetection in semiconductors works on the general principle of the creation of electron-hole pairs under the action of light. When a semiconductor material is illuminated by photons of an energy greater than or equal to its bandgap, the absorbed photons promote electrons from the valence band into excited states in the conduction band, where they behave like free electrons able to travel long distances across the crystal structure under the influence of an intrinsic or externally-applied electric field. In addition, the positively-charged holes left in the valence band contribute to electrical conduction by moving from one atomic site to another under the effects of the electric field. In this way the separation of electron-hole pairs generated by the absorption of light gives rise to a *photocurrent*, which refers by definition to the fraction of the photogenerated free charge-carriers collected at the edges of the material by the electrodes of the photodetecting structure, and whose intensity at a given wavelength is an increasing function of the incident light intensity. On this level we can distinguish between two large categories of photodetectors based on the nature of the electric field, which causes the charge separation of photogenerated electron-hold pairs: *photoconductors*, which consist of a simple layer of semiconductor simply with two ohmic contacts, where the electric field leading to the collection of the charge-carriers is provided by applying a bias voltage between the contacts at either end, and *photovoltaic* photodetectors, which use the internal electric field of a p-n or Schottky (metal-semiconductor) junction to achieve the charge separation. This last term covers p-n junction photodetectors (photovoltaic structures consisting of a simple p-n junction, and p-i-n photodetectors which include a thin layer of semiconductor material between the p and n region which is not deliberately doped), as well as all Schottky junction photodetectors (Schottky barrier photodiodes and metal-semiconductor-metal (MSM) photodiodes).

We will now briefly introduce the main physical concepts at the root of the operation of the different semiconductor photodetector families. Here the emphasis is placed on a phenomenological description of the working mechanisms of the devices in question; the corresponding formalism has been deliberately kept to an absolute minimum in the interests of clarity and concision.

## 1.3. Semiconductor devices

Photoconductors represent the simplest conceivable type of photodetector: they consist of a finite-length semiconductor layer with an ohmic contact at each end (Figure 1.1). A fixed voltage of magnitude $V_B$ is applied between the two end contacts, in such a way that a bias current $I_B$ flows through the semiconductor layer, simply following Ohm’s law. The active optical surface is formed from the region between the two collection electrodes. When it is illuminated, the photogenerated changes produced under the effect of the applied electric field lead to a photocurrent $I_{PH}$ which is added to the bias current, effectively increasing the conductivity of the device.

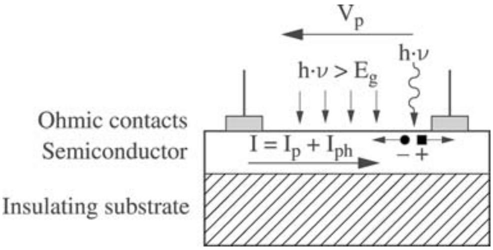


**Figure 1.1.** *Diagram of a photoconducting device*

The main point of interest in a photoconducting device is its increased gain, the response of photoconductors being typically several orders of magnitude greater than that of photovoltaic detectors for a given material. On the other hand, its other operational parameters (bandwidth, UV/visible contrast, infrared sensitivity) are generally below that of other types of photodetectors, which often greatly limits the scope of its potential applications (this is particularly the case for photoconductors based on III-V nitrides, as we will see later on).

## 1.4. p-n junctions and p-i-n structures

In p-n diodes, the metallurgical linkage of a region of a p-type doped semiconductor and a region of n-type doping forms a p-n junction, where the joining of the Fermi levels in equilibrium mostly occurs through a flow of charge between the n and p regions. In equilibrium we therefore find a region with no free charge carriers immediately around the junction, similar to a charged capacitor, where there are, on the n side, positively ionized donors and, on the p side, negatively ionized acceptors (this zone is known as the space charge region (SCR), where ionized donors and acceptors provide fixed charges). The presence of charged donors and acceptors produces an electric field in that region which curves the energy bands and, in equilibrium, forms an energy barrier between the two regions: the bottom of the conduction band and the top of the valence band on the n side are below the corresponding levels on the p side (Figure 1.2).

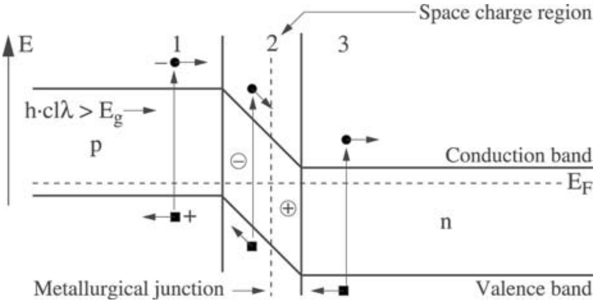


**Figure 1.2.** *Curvature of the energy bands and mechanisms of photocurrent generation in a p-n junction*

The width of the SCR is a decreasing function of the level of doping in the material, while the height of the energy barrier is an increasing function of it. An electron-hole pair produced in this SCR (situation 2 in Figure 1.2) is therefore separated by the effect of the internal electric field of the junction, and so does not recombine. These are the charge carriers which contribute to the photocurrent, to which we can add, to some extent, those generated at a distance from the junction less than or equal to the diffusion length (situations 1 and 3 in Figure 1.2). The band structure of the junction implies that the photocurrent will consist of minority charge carriers. For this reason, the photocurrent flows in the opposite direction to the bias on the diode, where the forward direction is defined as the direction of flow of the majority charge carriers (from the n to the p region in the case of electrons, and vice versa for holes). Moreover, the application of an opposing external electric field ($V_p – V_n < 0$) allows us to increase the height of the energy barrier in the vicinity of the junction, and also increase the spatial extent of the SCR, which significantly improves the efficiency of the separation of electron-hole pairs by increasing the electric field within the junction.

We note that when the doping level is moderate, the width of the SCR is important. This effect is beneficial in the case of p-n junction photodetectors, where in order to increase the photoresponse it is desirable to ensure that the mechanisms of electron-hole pair generation through incident light take place predominately inside the SCR. A simple means of increasing the spatial extent of the SCR is to introduce between the n and p regions a thin layer of intrinsic semiconductor material which is not intentionally doped: the structure is therefore referred to as p-i-n. Such a structure is interesting because it is possible to maintain high levels of doping in the n and p regions without significantly reducing the extent of the SCR, whose width is then largely determined by the thickness of the "i" layer. Additionally, increasing the width of the SCR reduces the capacitance of the structure, which makes p-i-n structures particularly well-suited for high-speed operation.

## 1.5. Avalanche effect in p-i-n structures

When the reverse-bias voltage established at the terminals of a p-i-n structure increases sufficiently that the electric field established in the junction reaches values close to the breakdown field (in structures of micron-scale thickness, this is generally the case when the bias voltage at the terminals reaches a few dozen volts), the photogenerated charge carriers in the SCR (which is effectively the region that is not intentionally doped) are accelerated enough to separate other secondary charge carriers from the atoms in the lattice that they impact in the course of their motion: this is the avalanche effect which results in a multiplication of the charge carriers in the SCR. The gain is therefore greater than 1 for the generation of charge carriers by light, and this gain can even typically reach 10 or 20 under favorable conditions. This effect is exploited in what are called avalanche photodiodes where the levels of n- and p-type doping are generally adjusted to high values above $10^{18}$ $cm^{-3}$ to maximize the intrinsic electric field of the junction.

## 1.6. Schottky junction

A Schottky junction is formed by bringing a metal and a semiconductor into contact. The basic phenomena which lead to the formation of a Schottky junction with an n-type semiconductor are summarized in Figure 1.3.

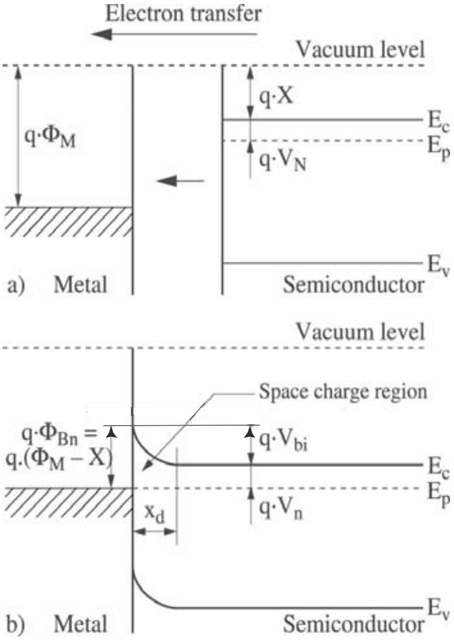


**Figure 1.3.** *Formation of a Schottky junction (in an n-type semiconductor)*

In thermal equilibrium, when the Fermi levels of the metal and the semiconductor are equalized, a transfer of electronic charge occurs from the semiconductor to the metal in the case where the work function $q.\Phi_M$ of the metal ($q$ being the elementary charge) is greater than the electron affinity $X$ of the semiconductor, and a SCR appears at the edge of the semiconductor of width $x_d$ next to the junction, where the only charges present are the positively-ionized donors. A curvature of the energy bands therefore occurs at the junction, which leads to the appearance of an energy barrier between the metal and the semiconductor, called a *Schottky barrier*, whose height is given to first approximation by the expression:

$$q \cdot \Phi_{Bn} = q \cdot \left(\Phi_M - \chi\right) \qquad [1.1]$$

In equilibrium, therefore, we find an intrinsic electric field immediately next to the metal-semiconductor junction which is comparable in form to that found in a p-n junction. Consequently, it is the phenomenon of photogeneration of charge carriers inside and near to the SCR which is responsible for the appearance of a photocurrent, with the electron-hole pairs being separated by the effect of the electric field in the Schottky junction. It is possible, as in the case of the p-n junction, to modify the intensity of the internal electric field in the junction by applying a bias voltage V between the semiconductor and the metal of the Schottky contact (Figure 1.4).

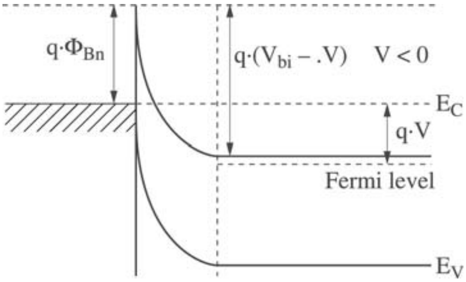


**Figure 1.4.** *Reverse-bias of a Schottky junction (n-type semiconductor material)*

In the case of an n-type semiconductor, the application of a negative voltage between the semiconductor and the metal electrode of the Schottky contact has the effect of reverse-biasing the Schottky junction, which leads to an increase in the height of the effective barrier, along with an increase in the width of the SCR. This last effect is of course favorable for photodetection. Indeed, it follows that the majority charge carriers (electrons) cannot flow towards the Schottky contact, and only the minority carriers (holes) generated by external excitation (in particular photogeneration) can reach the Schottky contact and hence produce an electric current: as in the case of the p-n junction, we therefore find that the current flows in reverse through the Schottky junction, that is, from the semiconductor towards the Schottky contact. The illumination of Schottky photodiodes can occur through the front or rear face (often this second option is chosen in the case where the substrate material is transparent to the light to be detected, as is the case for example with sapphire). In the case of illumination through the front face, we resort to a semi-transparent Schottky contact, characterized by a very small thickness of metal (of the order of 100 Å) selected to ensure sufficient optical transmission: while a thin layer of gold of 100 Å thickness transmits up to 95% of the incident light in the infrared, the percentage transmitted in the ultraviolet is around 30% in the range 300-370 nm. The gain of p-i-n photodiodes (other than the specific case of avalanche photodiodes) and Schottky photodiodes is at most 1, which would be the case if all the photogenerated charge carriers were collected by the electrodes at the ends of the device.

## 1.7. Metal-semiconductor-metal (MSM) structures

An MSM structure consists of two Schottky electrodes, often interlinked in the form of a comb structure, leaving a free semiconductor surface between the two contacts which forms the active region in which light will be absorbed. A bias voltage can be applied between the two electrodes, in order to break the initial electrical symmetry of the contacts: one of the Schottky junctions is reverse-biased, producing a SCR of increased width, and the other junction is forward-biased.

The absorption of light near the reverse-biased junction creates electron-hole pairs which are separated under the effects of the electric field present in the SCR, thus creating the photocurrent. The other electrode, consisting of a forward-biased (and hence transmissive) Schottky junction, simply acts as a collection electrode. The band diagram of the device under increased bias voltage ($V_B$) is represented schematically in Figure 1.5, in which $L$ is the distance between two adjacent contact fingers, $\Phi_0$ is the height of the Schottky barrier and

$I_{ph}$ is the photocurrent. MSM photodetectors normally use semiconductor materials which are not intentionally doped, are chemically very pure and electrically very resistive. The SCRs associated with Schottky junctions made of these materials are hence of significant width which, for a given bias voltage, allows the electric field of the junction to extend more easily into semiconductor regions some way from the contact. It follows that photogenerated electron-hole pairs are more easily separated and collected by the electrodes at either end.

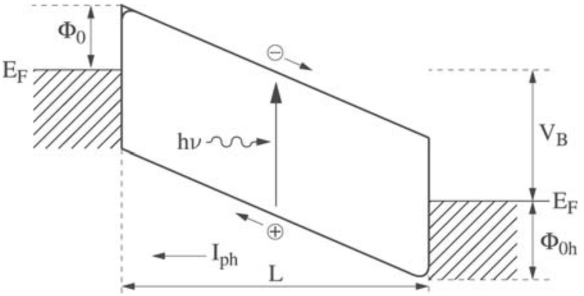


**Figure 1.5.** *Energy band diagram for an MSM structure under electrical bias; effect of illumination*

## 1.8. Operational parameters of photodetectors

The main parameters which define the behavior of an ultraviolet photodetector are respectively the response coefficient, the gain, the quantum efficiency, the bandwidth, the noise equivalent power (NEP) and the detectivity.

### 1.8.1. *Response coefficient, gain and quantum efficiency*

The response coefficient of a photodetector, $R_i$, links the photocurrent $I_{ph}$ to the power of the incident light $P_{opt}$ through the relationship:

$$I_{ph} = R_i \cdot P_{opt} \qquad [1.2]$$

It is important to note in passing that the response coefficient is a quantity independent of the active optical surface of the photodetector structure: indeed, the photocurrent as well as the incident optical power are both, in the ideal case, proportional to the active optical surface. At a given wavelength $\lambda$, the flux $\Phi$ of photons arriving on the semiconductor surface, which is defined as the number of photons reaching the active surface per unit time, is given by:

$$\Phi = P_{opt} \cdot \lambda / \left(h \cdot c\right) \qquad [1.3]$$

where $h$ is the Planck constant and $c$ is the speed of light.

The quantum efficiency $\eta$ is defined as the probability of creating an electron-hole pair from an absorbed photon. Considering that all the incident light is absorbed in the semiconductor material, the rate $G$ of electron-hole pair generation per unit time is thus given by:

$$G = \eta \cdot \Phi = \eta \cdot P_{opt} \cdot \lambda / \left(h \cdot c\right) \qquad [1.4]$$

If we now introduce the gain parameter $g$ which corresponds to the number of charge carriers detected relative to the number of photogenerated electron-hole pairs, then the photocurrent is given by the equation:

$$I_{ph} = q \cdot G \cdot g = q \cdot \eta \cdot P_{opt} \lambda / \left(h \cdot c\right) \cdot g = \left(q \cdot \eta \cdot \lambda / \left(hc\right) \cdot g\right) \cdot P_{opt} \qquad [1.5]$$

where $q$ is the elementary charge (1.602 x $10^{-19}$ C), from which we obtain the expression for the response coefficient of the detector:

$$R_i = q \cdot g \cdot \eta \cdot \lambda / \left(hc\right) \qquad [1.6]$$

### 1.8.2. *Temporal response and bandwidth*

The speed of response of a photodetector may be limited by capacitative effects, by the trapping of charge carriers or by the saturation speed of charge carriers in the semiconductor. These phenomena all lead to a reduction in the response of the photodetector in the high-frequency domain. The cutoff frequency $f_C$ of the photodetector is defined as the frequency of optical signal for which the response coefficient is half that for a continuous optical signal. The temporal response of a photodetector is characterized by the fall time $\tau_f$ (or the rise time $\tau_r$), which is defined as the time needed for the photocurrent to fall from 90% to 10% of its maximum (or to rise from 10% to 90% of it). In the case of a transient exponential response with a time constant $\tau$, the following relationship links the bandwidth $BW$ and the temporal response of the photodetector:

$$BP = 1/\left(2\cdot\pi\cdot\tau\right) = 2.2/\left(2\cdot\pi\tau_m\right) = 2.2/\left(2\cdot\pi\cdot\tau_d\right) \qquad [1.7]$$

### 1.8.3. *Noise equivalent power*

The NEP is defined as the incident optical power for which the signal-to-noise ratio is 1, and hence the photocurrent $I_{ph}$ is equal to the noise current $I_b$. In other words, it is the smallest optical power which can be measured. It follows that the NEP parameter is given by the equation:

$$NEP = I_b / R_i \left(in\ W\right) \qquad [1.8]$$

In the case of white noise, the noise current $I_b$ increases as the square root of the bandwidth of the photodetector device. It follows that it is preferable and customary to use the following expression for the NEP, normalized with respect to the bandwidth $BW$:

$$NEP^* = NEP\cdot\left(BW\right)^{1/2}\left(\text{in } W\cdot Hz^{\ 1/2}\right) \qquad [1.9]$$

In semiconductors, there are five sources of noise:

– shot noise, mainly due to the random nature of the collisions of incident photons;

– thermal noise, due to random collisions of charge carriers with the atoms of the crystal lattice, in permanent vibration due to thermal motion;

– partition noise, caused by the separation of the electric current into two parts flowing across separate electrical contacts;

– generation-recombination noise, caused by the random generation and recombination of charge carriers, either band to band or via trapping levels situated in the bandgap;

– 1/*f* noise, associated with the presence of potential barriers at the level of the electrical contacts. This last type of noise dominates at low frequencies.

### 1.8.4. *Detectivity*

This figure of merit is defined by the equation:

$$D = \left(NEP\right)^{-1} = R_i / I_b \left(in\, W^{-1}\right) \qquad [1.10]$$

In general terms, the photocurrent signal increases in proportion to the active optical area $A_{opt}$, and in addition the noise current increases with the square root of the product of the active optical area with the bandwidth $BW$. It follows that the preferred method of comparing between different photodetectors is to use an expression for the detectivity normalized with respect to these parameters, written:

$$D^* = D \cdot \left(A_{opt} \cdot BP\right)^{1/2} = \left(R_i / I_b\right) \cdot \left(A_{opt} \cdot BP\right)^{1/2} \left(\text{in } W^{-1} \cdot cm.Hz^{1/2}\right) \qquad [1.11]$$

The normalized detectivity is the most important parameter for characterizing a photodetector because it allows direct comparison of the performance of photodetectors using technologies and methods of operation which are at first glance very different. It is clear from the preceding definitions that the determination of the NEP and the detectivity requires measurement of three parameters: the response coefficient, the bandwidth and the noise current of the photodetector device. The measurement of the noise current must be made in darkness. The device is biased using a very stable voltage source, and the entire measurement system must itself have an intrinsic noise level considerably lower than the intrinsic noise of the photodetector device.

# Chapter 2

# PIN Photodiodes for the Visible and Near-Infrared

## 2.1. Introduction

Photodiodes are optoelectronic devices with two electrodes and asymmetric electrical characteristics. They are normally reverse-biased (biased in the non-conducting direction) and used to convert optical fluxes into electrical currents. The term PIN refers to the three doped semiconductor layers in their active part, of types P, intrinsic and N respectively, a structure which aims to optimize the characteristics compared to those of simple PN junction photodiodes.

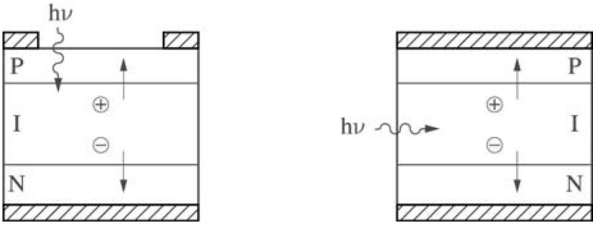


**Figure 2.1.** *Schematic layout of PIN photodiodes: collinear optical flux and charge transport on the left, orthogonal on the right*

---

Chapter written by Baudoin DE CREMOUX.

Figure 2.1 shows the basic processes taking place in a PIN photodiode:

– generation of an electron-hole pair through absorption of a photon, most probably in the I region, either using radiative or non-radiative processes;

– transport under the effects of the electric field present in the I region: of the electron towards the N region and of the hole towards the P region, then to the external circuit through the metallic contacts deposited on the surfaces, which have a window where required.

Thus, the absorption of a photon in the I region leads to the transport of an electron in the external circuit. Figure 2.1 shows the possible arrangements for the optical flux and the electrical current. Collinear geometry is the most commonly used. We use it as the reference type for the analysis of photodiode characteristics. After this, we will consider orthogonal geometry.

The spectral domain of interest, mostly from 400 to 1,700 nm, determines the materials used to fabricate photodiodes. This design is the most mature of the semiconductors, and as a result almost all photodiodes used in this domain have a PIN structure, or their central layer is at least weakly doped compared to the P and N layers.

This chapter is organized in order to first provide the reader with the basic tools to understand the concepts of PIN photodiodes for a given application, then it describes the devices which are currently available or which should be available soon. With this aim in mind, section 2.2 recalls some of the physics of semiconductors, emphasizing the processes taking place in photodiodes, the pertinent characteristics of the materials and the notations used later in the text. Sections 2.3 and 2.4 establish the simple quantitative relations between the structure of devices with typical geometries and their static and then dynamic characteristics. We will stress the compromises required among these characteristics if we are to keep a particular geometry. Section 2.5 then describes the properties of photodiodes with collinear geometry, currently used in a numerous professional and mass-market systems. Finally, section 2.6 describes advanced devices with novel structures which push the boundaries of the previously-discussed compromises. These are for the most part laboratory devices which could find applications in real-world systems.

## 2.2. Physical processes occurring in photodiodes

### 2.2.1. *Electrostatics in PIN diodes: depleted region*

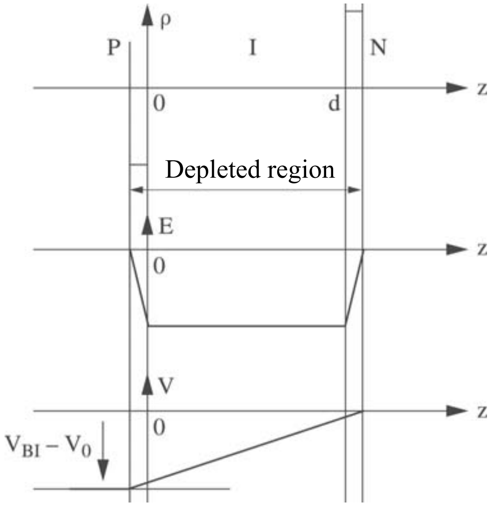


**Figure 2.2.** *From top to bottom, profiles of fixed charge density, electric field and potential across a reverse-biased PIN diode*

In the neighborhood of a PN junction there is a region devoid of free charge carriers (the depleted region) where fixed ionized impurities can be found, negative on the P side and positive of the N side, forming a space charge region (SCR). By integrating the Poisson equation twice in the z direction:

$$\frac{d^2V}{dz^2} = \frac{\rho(z)}{\varepsilon} \qquad [2.1]$$

Where $\rho(z)$ is the charge density and $\varepsilon$ is the dielectric constant, we can obtain the profiles of the electric field $E$ and the potential $V$ across the structure. In the case of an abrupt PN junction, the thickness of the depleted region varies as $\sqrt{(V_{BI}-V)}$ where $V_{BI}$ is the *built-in* or *diffusion* potential present in the absence of external bias applied to the diode, and $V$ is the voltage across the terminals of the diode. As a result, the capacitance of the junction decreases when the reverse-bias voltage increases in absolute value. The case of a PIN diode structure is considered in Figure 2.2, showing in sequence the profiles across the structure of the density of fixed charges, of the electric field and of the potential.

In the realistic approximation that the densities of ionized impurities are much greater than the density of residual impurities in the I region (typically by a factor of 100), we draw the following conclusions:

– the extension of the depleted region into the P and N regions is negligible, and the depleted region is to all intents and purposes coincident with the I region whose thickness, represented by $d$ in the rest of this chapter, is a very important parameter in the definition of a PIN photodiode;

– the capacitance of the junction is independent of the reverse-bias voltage. It is given, as for a conventional capacitor, by:

$$C_J = \varepsilon \frac{S}{d} \tag{2.2}$$

where S is the area of the junction;

– the electric field is constant in the depleted region and, insofar as $V >> V_{BI}$, its value is

$$E \cong V/d \tag{2.3}$$

These approximations will be used throughout the remainder of the chapter unless explicitly stated otherwise.

### 2.2.2. *Mechanisms of electron-hole pair generation*

#### 2.2.2.1. *Radiative generation and optical absorption*

Electronic transitions between permitted levels in the valence band (VB) and levels in the conduction band (CB) can occur under the action of photons transferring their energy to the electrons. Thus, a photon can be absorbed, creating an electron-hole pair. This process of radiative generation, shown schematically in Figure 2.3, is the inverse of the process of recombination which gives rise to stimulated emission of photons in laser diodes.

It follows from this that the only photons absorbed efficiently are those whose energy $h\nu$ is greater than the width of the bandgap $E_G$ of the material under consideration:

$$h\nu \geq E_G = E_C - E_V \qquad [2.4]$$

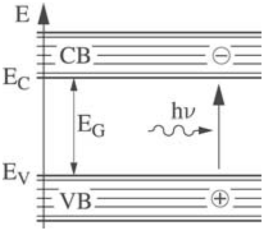


**Figure 2.3.** *Radiative generation*

The material is transparent to photons with lower energies. The parameter $E_G$ is thus a criterion in the choice of materials when designing a photodetector. Optical absorption is characterized by the coefficient of absorption $\alpha$ (cm$^{-1}$), according to the relationship:

$$\alpha = -\frac{dF}{Fdz} \qquad [2.5]$$

where $F$ (cm$^{-2}$s$^{-1}$) is the photon flux propagating in the $z$ direction. The level of radiative generation $G_R$ (cm$^{-3}$s$^{-1}$), the number of electron-hole pairs created per unit volume and time, is therefore:

$$G_R = \alpha F \qquad [2.6]$$

It follows from these definitions that the level of radiative generation in a homogenous material, for a photon flux $F_0$ falling on a surface at $z=0$, can be written:

$$G_R = \alpha F_0 \exp\left(-\alpha z\right) \qquad [2.7]$$

We see that $1/\alpha$ (cm) gives an order of magnitude for the distance required to absorb the radiation. Furthermore, the knowledge of variation of $\alpha$ with photon energy (or with their wavelength) is essential for the choice of materials for use in a photodetector. Figure 2.16 shows graphs for the main semiconductors used in the spectral domain under consideration in this chapter.

The shape of these graphs and the magnitude of $\alpha$ depend on the direct or indirect nature of the band structure of the material. This nature is defined in Figure 2.4 which represents, for the two types of material, the variations in energy of the electrons, with the wavevector $\boldsymbol{k}$ describing their motion in the periodic crystal lattice.

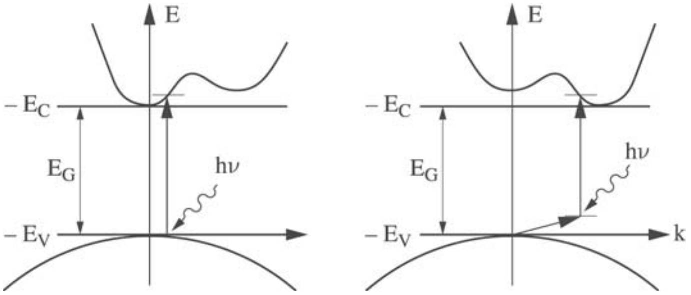


**Figure 2.4.** *Direct bandgap on the left and indirect bandgap on the right*

For materials with a direct bandgap (such as gallium arsenide) the main minimum in the conduction band is found at the same vector ***k*** as the maximum in the valence band. Furthermore, radiative transitions must simultaneously conserve the energy and momentum of the particles involved (the photon and electron). The momentum of the electron is much greater than that of the photon, and as a result radiative transitions are effectively vertical and connect with a high probability those states at the top of the VB and those at the bottom of the CB.

The opposite is true for materials with an indirect bandgap (such as silicon), and radiative transitions require the involvement of the third particle, a phonon, to provide the difference in momentum between the initial and final states. These transitions are much less probable and $\alpha$ grows much less quickly with the energy of the photons above $E_G$.

### 2.2.2.2. *Non-radiative generation*

As well as the radiative processes there are also non-radiative processes of generation-recombination. Thus, in the SRH (*Shockley Read Hall*) process shown in Figure 2.5 in generation mode, the required energy is "borrowed" from the thermal motion of the crystal lattice. This is one of the contributions to the dark current of photodetectors, a detrimental parasitic phenomenon.

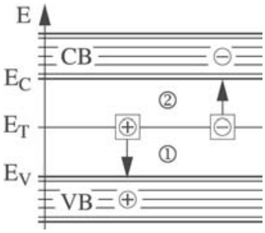


**Figure 2.5.** *Non-radiative SRH generation*

The transition takes place in two parts, via a “deep state” of energy $E_T$, which is close to the middle of the forbidden band and is the result of a crystal defect (impurity, vacancy, etc.). The rate $G_{NR}$ (cm$^{-3}$s$^{-1}$) of SRH generation-recombination can be written [SZE 81, p. 35]:

$$G_{NR} = \frac{\sigma_n \sigma_p v_{th} N_T \left(n_i^2 - np\right)}{\sigma_n \left[n + n_i \exp\left(\frac{E_T - E_I}{kT}\right)\right] + \sigma_p \left[p + n_i \exp\left(-\frac{E_T - E_I}{kT}\right)\right]} \qquad [2.8]$$

with the following notation:

– $\sigma_n$ and $\sigma_p$ capture cross-sections for electrons and holes by the deep state;

– $v_{th}$, thermal velocity $\left(v_{th} = \sqrt{\left(3kT/m^*\right)}\right)$;

– $N_T$, density of deep states;

– $n_i$, intrinsic electron density;

– $n$ and $p$, electron and hole densities;

– $E_I$, middle of the forbidden band;

– $kT$, Boltzmann factor.

In the depleted region of a photodiode we have $np = n_i^2$. If we further assume that $\sigma_n = \sigma_p = \sigma$ and $E_T = E_I$, and define an effective lifetime $\tau_{eff}$:

$$G_{NR} = n_i v_{th} N_T / 2 = n_i / \tau_{eff} \qquad [2.9]$$

We also know that:

$$n_i^2 = N_C N_V \exp\left(-E_G / kT\right) \qquad [2.10]$$

where $N_C$ and $N_V$ are the equivalent densities of state for the CB and VB. We can draw the following conclusions from this:

– $G_{NR}$ increases very rapidly as the ratio $E_G/kT$ decreases;

– $G_{NR}$ is not an intrinsic characteristic of the material, but depends for its purity on $N_T$.

In an electrically neutral region of type p(n), where the electron (hole) deficiency relative to thermodynamic equilibrium is $\delta n(\delta p)$, if we define $\tau n(\tau p)$ as the lifetime of the minority electrons (holes), it follows that:

$$G_{NR} = \delta n \sigma v_{th} N_T = \delta n / \tau_n \qquad [2.11]$$

SRH non-radiative generation is particularly active at the surface of devices, where the density of defects is very high. It can be reduced by a passivation treatment; for silicon the most effective is the deposition of its natural oxide $SiO_2$, silica. Another process of non-radiative generation is shock ionization which occurs in a semiconductor material under a strong electric field. It grows with $E_G$ and the breakdown field is of the order of $E_B = 10^5\ V/cm$ for the materials considered. This effect is one cause of breakdown in diodes, limiting the reverse-bias voltage that can be applied. It is exploited as a gain mechanism in the avalanche photodiodes which are described in Chapter 3, and whose structure is optimized with this in mind.

### 2.2.3. *Transport mechanisms*

The charge carriers, electrons and holes generated by the aforementioned processes, are transported in the device to the two electrical contacts that transfer it to the external circuit – specifically, to the load resistance across the terminals, across which the electrical signal is measured.

### 2.2.3.1. *Transport under electric field and resultant current*

In the depleted region of a photodiode (normally reverse-biased) an elevated electric field is present, and the electrons and holes are driven at a speed which depends on the intensity of the field, as shown in Figure 2.6.

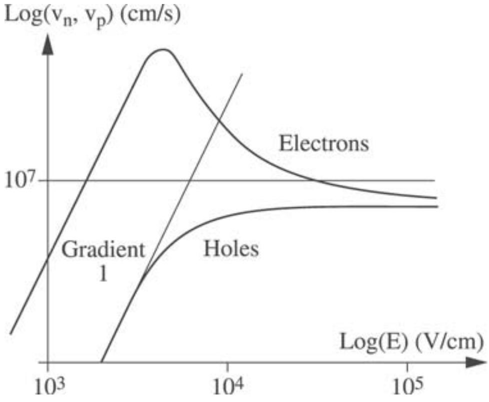


**Figure 2.6.** *Characteristics $v_n(E)$ and $v_p(E)$ in logarithmic coordinates*

For electric fields such that $E < 10^3$ *V/cm*, the velocities are proportional to $E$ (linear regime) and we have:

$$v_{n,p}(E) = \mu_{n,p} E \qquad [2.12]$$

where $\mu_n$ and $\mu_p$ (cm$^2$/Vs) are the respective mobilities of electrons and holes. For electric fields such that $E > 10^4$ V/cm, the velocities are fairly constant and equal to a saturation velocity of the order of $v_S = 10^7$ cm/s. Between these two regions, and for certain materials, the electron velocity passes through a maximum (high mobility regime).

#### 2.2.3.2. *Transport along the concentration gradient and diffusion current*

In the contiguous electrically neutral regions in the depleted region, electrons and holes can be transported in the absence of an electric field under the effect of their concentration gradients. Their fluxes $F_n$ and $F_p$ can be written:

$$F_{n,p} = D_{n,p} \frac{\partial n, p}{\partial z} \qquad [2.13]$$

where $D_{n,p} = \mu_{n,p} kT/q \left(cm^2/s\right)$ are the diffusion coefficients.

The various recombination processes are also active in these regions, such that the electrons and holes are only efficiently transported over a distance known as the diffusion length and given by:

$$l_{n,p} = \sqrt{D_{n,p} \tau_{n,p}} \qquad [2.14]$$

## 2.3. Static characteristics of PIN photodiodes

### 2.3.1. *I/V characteristics and definition of static parameters*

Figure 2.7 shows I/V curves, linking the current crossing a photodiode to the voltage across its terminals for three illumination conditions. From this we can extract the parameters commonly used to characterize photodiode operation, whose relations to the structure of the device are described in the following sections. The photocurrent $I_P$ produced by a photodiode is proportional to the illumination across a large range of optical power *P*. The coefficient of proportionality is the response coefficient *R* (A/W) (*responsivity*), not to be confused with the *sensitivity,* which measures the minimum optical power detectable by an optical detector, and which depends on the circuit into which the photodiode is inserted and the specifications of the signal to be received.

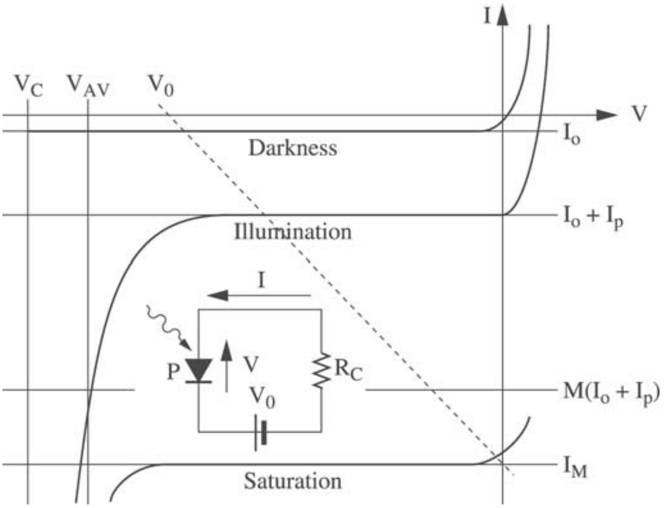


**Figure 2.7.** *I/V characteristics of a photodiode, inserted in the circuit shown and under several illumination conditions*

The external quantum efficiency $\eta_e$ links the generated electron flux to the incident photon flux. Thus, we have:

$$R = \frac{I_p}{P} = \frac{q}{h\nu}\eta_e \qquad [2.15]$$

where $q$ is the charge of an electron and $h\nu$ is the energy of a photon.

The dark current $I_0$, discussed in section 2.3.3, is caused by non-radiative generation and/or the tunnel effect.

The saturation current $I_M$ is the photocurrent above which several mechanisms limit the linearity of the response:

– the intrinsic mechanism, due to the screening of the electric field in the depleted region by the space charge that is produced by the mobile charge carriers created by the illumination (see section 2.3.4);

– the circuit-dependent mechanism, caused by the voltage drop across the load resistance (see the circuit in Figure 2.7).

A gain, characterized by the multiplication coefficient $M$ defined by:

$$M = \frac{I}{I_p + I_0} \qquad [2.16]$$

can be obtained under a high reverse-bias voltage $V_{AV}$ in avalanche photodiodes (see Chapter 3). The breakdown voltage $V_B$ is the highest reverse-bias voltage the diode can support. For $V=V_B$ the dark current grows very rapidly and the power dissipated in the diode can lead to its destruction.

## 2.3.2. *External quantum efficiency*

The external quantum efficiency defined above can be thought of as the product of an internal quantum efficiency $\eta_i$ and an optical efficiency $\eta_o$, the ratio of the photon flux entering the device to the flux of photons incident on its surface. These two ratios, which are both dependent on the wavelength, are considered in the following text.

### 2.3.2.1. *Optical efficiency*

The optical refractive index $n$ of the semiconductors under consideration is in the region of 3, and as a result the incident light rays are refracted and reflected at its surface. At normal incidence, used as the reference situation, the Frensel transmission coefficient determines the optical efficiency:

$$\eta_o = \frac{4n}{(n+1)^2} \qquad [2.17]$$

which is of the order of 70% and depends little on the wavelength and the angle of incidence. It is possible to achieve optical efficiencies close to 100% at a given wavelength $\lambda$, by depositing an antireflection coating on the surface, for example of refractive index $\sqrt{n}$ and thickness $\lambda/\left(4\sqrt{n}\right)$. The optical efficiency therefore depends on $\lambda$ and on the angle of incidence.

### 2.3.2.2. *Internal quantum efficiency*

The electron-hole pairs created by the absorption of photons entering the device may, or may not, give rise to a current flow in the external circuit, depending on where they are generated in the structure. If generation is in the depleted region, the electrons and holes are separated by the electric field, as shown in Figure 2.1, leading to the passage of one electron for each absorbed photon: thus, the internal quantum efficiency is 100%. In the case of generation in the electrically neutral P and N regions, where the electric field is weak, only the electrons (or holes) generated less than one diffusion length away are transported into the depleted region, and the internal quantum efficiency is in most cases weak.

It is therefore desirable to design a photodiode in such a way that the radiation should be entirely absorbed in the depleted region. Figure 2.8 shows a model which provides a good evaluation of the internal quantum efficiency of a PIN photodiode, as a function of the nature of the material on whose surface the radiation falls. The curves 1 and 2 represent the profiles of the radiative generation rate $G_R$ across two structures. The structure for curve 1 is that of a homojunction PIN, where all of the materials used have the same width of bandgap, which is below the energy of the photons to be detected. The radiation is assumed to be incident from the p side, and part of the radiation is absorbed there, where the quantum efficiency is negligible. Another part is absorbed in the i region, and has a quantum efficiency of 100%. Finally, the contribution of the residual radiation entering the n region is also negligible. The quantum efficiency can therefore be written:

$$\eta_i = \exp\left(-\alpha dp\right)\left[1 - \exp\left(-\alpha d\right)\right] \qquad [2.18]$$

where the definitions of $d_P$ and $d$ can be found in Figure 2.8.

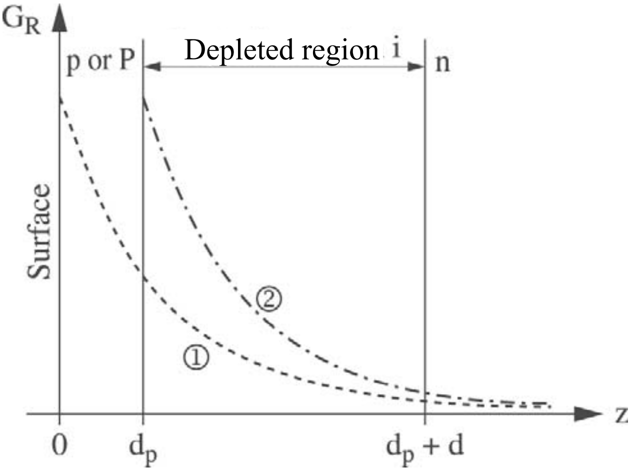


**Figure 2.8.** *Optical generation rate across the structure of two PIN photodiodes: curve 1 shows a homojunction PIN photodiodes; curve 2 shows a heterojunction PIN photodiode*

Curve 2 corresponds to a PIN photodiode, with the capital P indicating that the material in the P region has a bandgap width greater than the energy of the photons to be detected and which is therefore transparent to them. The same reasoning provides us with the quantum efficiency:

$$\eta_i = 1 - \exp\left(-\alpha d\right) \qquad [2.19]$$

### 2.3.3. *Dark current*

The dark current is the sum of several contributions: non-radiative generated currents requiring an input of energy (such as the SRH process) and the tunneling current, which does not require such an input. These two processes are represented in Figure 2.9 for a sharp $p^+n$ homojunction.

### 2.3.3.1. *Generation currents*

According to equation [2.9], the dark current contribution of generation in the depleted region is:

$$I_0 = SdG_{NR} = Sdn_i \big/ \tau_{eff} \qquad [2.20]$$

Taking account of equation [2.11] – of the fact that the shortage of electrons (holes) in the p(n) material near the depleted region is $\delta n, p = n_i^2 \big/ N_{AD}$, and the fact that the thickness of the p(n) material is $L_{n,p}$ (see section 2.2.3.2) – the generation contribution in the neutral p and n regions can be written:

$$I_0 = Sqn_i^2 \left( L_n \big/ N_A \tau_n + L_p \big/ N_D \tau_p \right) \qquad [2.21]$$

This is the traditional expression for the reverse current of a diode according to Shockley.

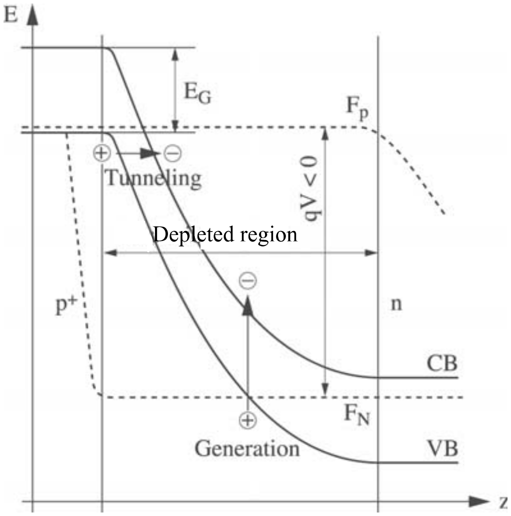


**Figure 2.9.** *Profile of the VB and the CB of a reverse-biased p+n junction. The generation process (shown here in the depleted region) requires an input of energy; the tunnel process occurs at constant energy*

Thus, we see from equations [2.20] and [2.21] that the generation current is primarily a function of $n_i$ and varies very rapidly with $E_G/kT$.

### 2.3.3.2. *Tunneling current*

Using the tunneling effect electrons can also be transported at a constant energy across the triangular potential barrier present at the junction, as long as it is narrow enough. A calculation of the current gives the following expression [STI 82]:

$$I_T = S\frac{\sqrt{2m^* q^3 EV}}{4\pi^3 \hbar^2 E_G^{1/2}} \exp\left(-\frac{\pi\sqrt{2m^* E_G^{3/2}}}{4q\hbar E}\right) \qquad [2.22]$$

where $E=V/d$ is the electric field in the depleted region, which is constant for a PIN diode. From equation [2.22] we determine that, for a given material, $I_T$ increases very rapidly with the voltage $V$ across the terminals of the diode, mostly as a result of the increase in $E$. Additionally, $I_T$ increases when $E_G$ decreases, such that the contribution of the tunneling current to the dark current is most significant for photodiodes made of materials with a small bandgap.

## 2.3.4. ***Breakdown voltage***

### 2.3.4.1. *Zener breakdown via the tunneling effect*

The very rapid variation of the tunneling dark current with the reverse-bias voltage is the first process of breakdown. This mainly applies to materials with a narrow depleted region and a small bandgap; the two parameters which reduce the width and the height of the potential barrier at the PI junction. The tunneling effect can be the dominant one in the materials used for detection of the longest wavelengths considered in this chapter (see section 2.5.3.2), if the doping level of the i zone is too high [TAK 80].

#### 2.3.4.2. *Avalanche breakdown*

We can define a critical electric field $E_C$, of the order of 2 x $10^5$ V/cm for the lightly-doped materials considered in this chapter, where the multiplication coefficient becomes infinite and leads to breakdown. In the case of PIN photodiodes, which are not designed to take advantage of avalanche gain, breakdown can arise locally, most often around the edge of the junction where the point effect increases the electric field. The breakdown voltage is therefore approximated below by:

$$V_C \cong E_C d \qquad [2.23]$$

### 2.3.5. *Saturation current*

#### 2.3.5.1. *Extrinsic limitation by the voltage drop across the load resistance*

In order to have a linear response, the photodiode must remain reverse-biased ($V$<0) so that the carriers are separated by the electric field in the depleted region. Thus, it can be seen in the simple bias diagram of Figure 2.7 that for a bias voltage $V_0$, the voltage drop in the load resistance $R_L$ can cancel out the voltage $V$ across the diode terminals for a limiting current $I_M$ such that:

$$I_M = V_0 / R_L \qquad [2.24]$$

#### 2.3.5.2. *Intrinsic limitation by the space charge of mobile charge carriers*

It is especially important to take into account the contribution of mobile charge carriers to the electric field $E$ in the depleted region for high levels of incident optical power. Specifically, $E$ must satisfy the following conditions:

$$0 \leq |E| \leq E_B \qquad [2.25]$$

in order to avoid breakdown of the junction.

We will determine the maximum value $I_M$ of the photocurrent that simultaneously satisfies these conditions for a heterojunction PIN structure, with a coefficient of absorption in the i region such that $\alpha d \gg 1$, and all the illumination is absorbed close to the Pi junction. We further assume that the mobile electrons in the i region are transported with the saturation velocity $v_s$, independent of the electric field.

Figure 2.10 shows the density profiles of the fixed and mobile charges and the electric field profile across the structure. Given the hypotheses set out in this section, it follows that the charge density due to electrons is $\rho = J/v_s$, where $J$ is the current density, the electric field variation in the depleted region is $\delta E = \rho d/\varepsilon$ and the maximum current $I_M$, obtained when $\delta E = E_B$, can be written:

$$I_M = C_J v_s E_B \qquad [2.26]$$

where $C_J$ is the capacitance of the junction.

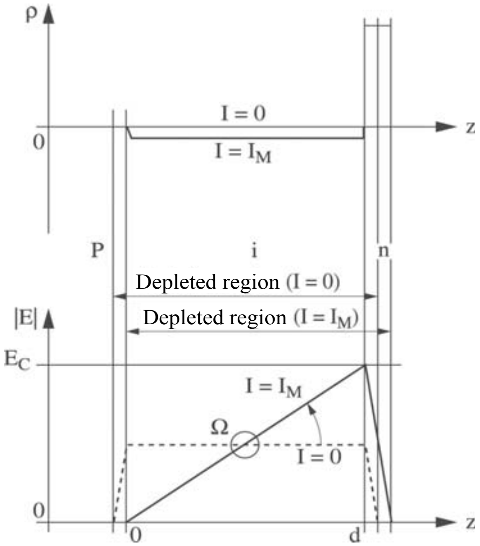


**Figure 2.10.** *Space charge density and electric field, for I=0 (dotted line) and for I=$I_M$ (solid line). For I=$I_M$, the electric field varies from 0 to the breakdown field $E_B$*

Note that, for this value of current, the voltage across the terminals of the junction is

$$V = E_B d/2 = V_B/2 \qquad [2.27]$$

We can easily show that, regardless of the components in the bias circuit, the part of the curve representing the variation of the electric field in the depletion region (marked Ω) rotates around a fixed point as the photocurrent varies. Its position depends on the values of $V_0$ and $R_L$; Figure 2.10 provides an example where $R_L = 0$ and the voltage across the diode terminals is constant and equal to the bias voltage $V_0 = V_C/2$.

## 2.4. Dynamic characteristics of PIN photodiodes

### 2.4.1. *Intrinsic limitations to the speed of response*

#### 2.4.1.1. *Transit time in the depleted region*

The transit time of photocarriers in the depleted region is the first limit to the response time of the photodiode. Its analysis is carried out here with the same approximations as in section 2.2.1, which is a good representation of the properties of real PIN diodes. Figure 2.11 shows the electric charge profile in the depleted region at a time *t* after a brief impulse signal, of energy $W_0$, which was applied to the photodiode at time 0. The packet of generated electrons is transported at speed $v_s$ and takes time $\tau_t = dv_s$ to cross the depleted region.

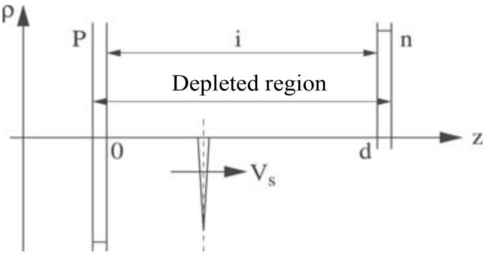


**Figure 2.11.** *Density of mobile charges produced in the depleted region of a PIN photodiode in response to an optical impulse $\delta(z, t)$*

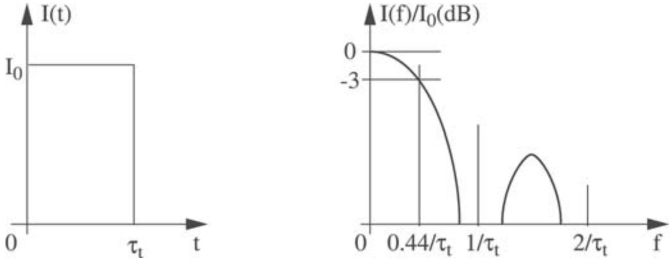


**Figure 2.12.** *Impulse and frequency response of a PIN photodiode limited by the transit time*

The solution of Maxwell's equations for the case $R_L=0$ shows that the current generated by the diode is the rectangular impulse shown in Figure 2.12, of duration $\tau_t$ and of intensity:

$$I_0=(q/h\nu)W_0/\tau_t \qquad [2.28]$$

The frequency response can then be obtained by Fourier transforming the impulse response. This is shown in Figure 2.12 and given by:

$$I(f)=I_0\sin(\pi f\tau_t)/\pi f\tau_t \qquad [2.29]$$

The 3 dB cutoff frequency $\left(I(f)/I_0=1/\sqrt{2}\right)$ is:

$$f_t=0.44/\tau_t \qquad [2.30]$$

We see that this varies as 1/*d*. In the case where $\alpha d\ll 1$ and the generation is uniform in the depleted region, the more accurate model presented in section 2.4.2.2 shows that we obtain the following numerical result:

$$f_t=0.55/\tau_t \qquad [2.31]$$

We see that regardless of the value of $\alpha d$, we can apply this approximate numerical expression, which we will continue to use:

$$f_t \cong 1/2\tau_t \qquad [2.32]$$

Also, equation [2.19] shows that if $\alpha d << 1$ we have:

$$\eta_i \cong \alpha d \qquad [2.33]$$

Thus we see from equations [2.32] and [2.33] that the product $\eta_i f_t$ only depends on the characteristics of the material in the i region:

$$\eta_i f_t \cong \alpha v_s/2 \qquad [2.34]$$

from which we conclude that, for the geometry considered and for a particular wavelength and material, we must accept a compromise between the internal quantum efficiency and the speed of response.

#### 2.4.1.2. *Diffusion time in the neutral regions*

When the optical generation occurs in one of the electrically neutral regions, *n* or *p*, the minority photocarriers must initially be transported by diffusion to the edges of the depleted region, in which they can be transported by the electric field. This process can be characterized by a diffusion time $\tau_d$ which we can show has an order of magnitude:

$$\tau_d \cong d^2/D_{n\,p} \qquad [2.35]$$

for electrons (holes) generated in the p(n) region. In practice the diffusion time is always greater than the transit time in the depleted region and has the effect of reducing the response speed of photodiodes. Therefore, excluding some exceptions (see section 2.6.4), it is desirable to design these devices such that the absorption of photons takes place in the depleted region, both to optimize the speed of response and the quantum efficiency.

#### 2.4.1.3. *Trapping at the Pi heterojunction*

An additional parasitic mechanism can slow the speed of heterojunction PIN photodiodes. The holes generated in the i region, which has a small bandgap, can be trapped at the interface with the P region by discontinuity in the valence band. They can then be emitted by thermo-ionic effects, although with delay, into the P region. This effect is particularly marked for holes, whose effective mass is greater than that of electrons. The proposed remedy [FOR 82] is the use of gradual heterojunction or layers with an intermediate-sized bandgap in order to reduce the potential barrier presented to the holes and to make it easier for them to cross through tunneling or thermal effects. We will assume below that the structures of the devices are designed in such a way that these effects are rendered negligible.

### 2.4.2. Limitations due to the circuit

#### 2.4.2.1. *Response of the basic circuit, optimum thickness of the i region*

In order to produce a signal, a photodiode is inserted into a circuit whose components also contribute to limit the bandwidth of an optical detector. The most simple circuit is shown in Figure 2.13, consisting of the capacitance of the junction $C_J$ in parallel with a load resistance $R_L$, with a negligible impedance of the source at high frequencies. This circuit is driven by the photodiode generated current, whose response is independent of the voltage at its terminals.

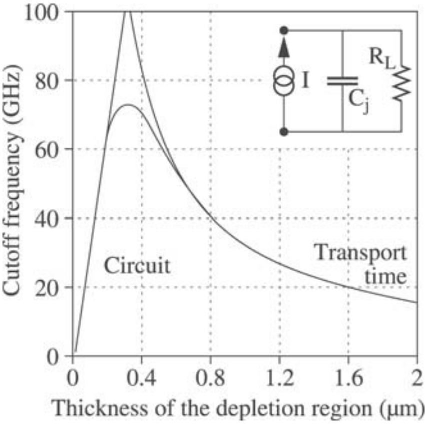


**Figure 2.13.** *3 dB cutoff frequency of a PIN photodiode, as a function of the thickness of the depleted region. The material is assumed to be GaInAs/InP, the width of the active region 10µm and the load resistance 50 Ω*

The impulse and frequency response of such a resistor-capacitor network are well known. The frequency response is given by the expression:

$$I(f) = \frac{I_0}{\sqrt{1 + (2\pi f R_L C_J)^2}} \qquad [2.36]$$

and the 3 dB cutoff frequency is:

$$f_{RC} = 1/(2\pi R_L C_J) \qquad [2.37]$$

We see that $f_{RC}$ varies with $d$. The cutoff frequencies $f_\tau$ and $f_{RC}$ vary inversely with $d$, thus there is an optimal thickness for the depleted region which maximizes the global cutoff frequency $f_{3dB}$. Figure 2.13 shows the cutoff frequencies variation with $d$ for a typical PIN photodiode.

The optimal thickness of $d$ is obtained when $f_\tau = f_{RC}$, giving:

$$\tau_t = \pi R_L C_J \qquad [2.38]$$

The maximum cutoff frequency is then

$$f_{3dB} = \frac{v_s}{2\sqrt{2d}} \qquad [2.39]$$

### 2.4.2.2. *Response of a real PIN photodiode*

It is also interesting to consider the results of a more complete analysis [BOW 87] of a photodiode made of GaInAs/InP, designed for optical telecommunications at a wavelength of 1.3 µm, taking into account:

– the finite value of the optical absorption coefficient: $\alpha = 1.3\ 10^4$ cm$^{-1}$;

– the different saturation velocities for electrons and holes $v_n = 6\ 10^6$ cm/s and $v_p = 4.8\ 10^6$ cm/s;

– the inevitable parasitic capacitances of the circuit: $C_P = 50$ fF;

– a parasitic inductance $L$, caused by the connecting wires, adjustable and kept as a free parameter.

The illumination is assumed to be incident at the N side of the structure, and the diameter of the active region is assumed to equal 30 µm. Figure 2.14 shows the response curves of the photodiode inserted in its circuit, with five examples of the parameters $d$ and $L$ listed in Table 2.1 along with the performances obtained, the cutoff frequency and the internal quantum efficiency. The load resistance was kept constant: $R_L = 50\ \Omega$.

| N° | d (µm) | L (nH) | $f_{3db}$ (GHz) | $\eta_i$ |
|---|---|---|---|---|
| 1 | 10.0 | 0.00 | 2.2 | 1.00 |
| 2 | 0.1 | 0.00 | 3.4 | 0.12 |
| 3 | 1.0 | 0.00 | 17.0 | 0.73 |
| 4 | 1.0 | 0.25 | 24.3 | 0.73 |
| 5 | 1.0 | 2.50 | NA | 0.73 |

**Table 2.1.** *Values of the parameters for the curves in Figure 2.17*

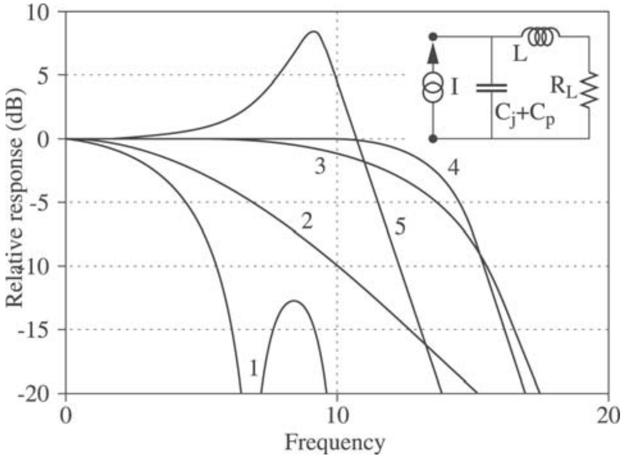


**Figure 2.14.** *Frequency response of a PIN photodiode made of GaInAs/InP at 1.3µm, inserted into the circuit shown. The parameters of each curve are given in Table 2.1*

Curves 1 and 2 show behaviors limited by the transit time and by the circuit respectively. Curve 3 corresponds to a depleted region whose width is close to the optimum, while curve 4 shows that its response can be further improved by parasitic induction. For a higher value of inductance (curve 5), the bandwidth is reduced, but a gain in signal is obtained.

### 2.4.3. *Power-frequency compromise, $Pf^2$ "law"*

For certain optical transmission systems, it is desirable to use a distortion-free increased electrical power in the sinusoidal regime across the resistance $R_L$ loading the photodiode. In particular, this allows us to simplify the electronics of the receiver, and is made possible by the current availability of powerful modulated optical sources or optical preamplifiers. Given the maximum frequency to be transmitted, the problem is now to design a photodiode, defined mainly by the thickness $d$ of the depleted region and by the area $S$ of the active region, as well as select the bias voltage $V_0$ in a circuit, as in Figure 2.7. We assume that the incident optical power is modulated at a level of 100%.

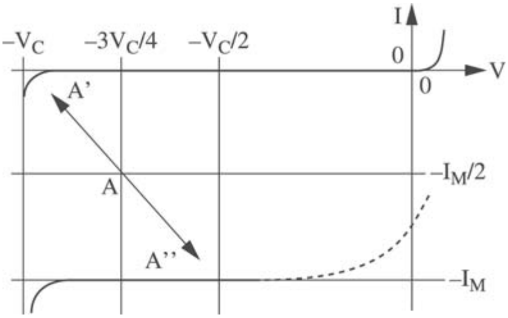


**Figure 2.15.** *Excursion from the operating point of the photodiode in a plot of its I/V characteristics: allowing us to determine the maximum RF electrical power. Points A, A' and A'' are respectively the mean and extreme points*

In order to maximize the RF power dissipated in the load, the photocurrent must vary sinusoidally (see Figure 2.15) between 0 and the maximum value $I_M$ given by equation [2.25], as well as the voltage varying between the breakdown voltage $V_B$ given by equation [2.23] and the voltage $V_B/2$ (equation [2.27]). The maximum effective power is therefore:

$$P_M = \frac{1}{2R_L}\left(\frac{V_C}{4}\right)^2 = \frac{E_B^2 d^2}{32 R_L} \qquad [2.40]$$

If we assume that the photodiode is optimized for its frequency response, and we eliminate *d* from [2.39] and [2.40], it follows that:

$$P_M f_{3dB}^2 R_C \cong \frac{E_C^2 v_s^2}{256} \qquad [2.41]$$

This expression shows that, for a fixed value of the load resistance, the product $P_M f_{3dB}^2$ only depends on the properties of the semiconductor material used for the i region. We should note that the numerical coefficient 1/256 is a maximal value which may in fact depend on the detailed specifications of the nonlinearity of the transmission system considered. Equation [2.41] is analogous to the known result for electronic components such as transistors [JOH 64]. We can note that this limitation a photodiode can generate on the RF power is of a purely electronic origin; in addition, thermal effects can severely limit it, in particular for components designed for low frequencies.

## 2.5. Semiconductor materials used in PIN photodiodes for the visible and near-infrared

### 2.5.1. *Absorption of semiconductors in the range 400-1,800 nm*

The absorption functions of the main semiconductor materials or semiconductor families used to build PIN photodiodes in the spectral range considered in this chapter are shown in Figure 2.16. They can be chosen as a function of the wavelengths to be detected. Other materials have been studied in the past, but have not taken off in practice for numerous reasons; these will be mentioned briefly later on.

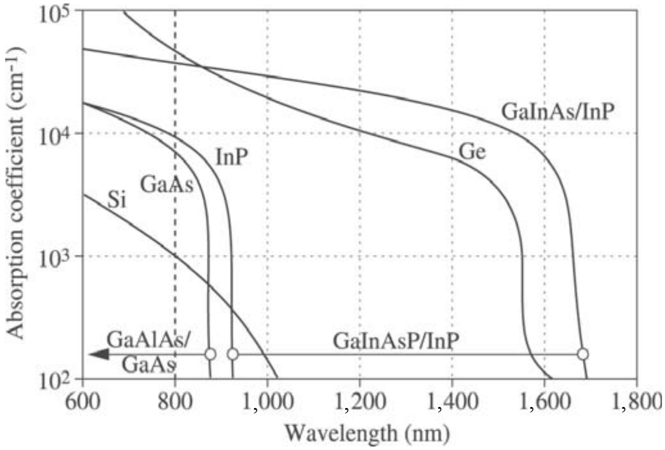


**Figure 2.16.** *Absorption spectra for semiconductor materials used in the visible to near-infrared range*

### 2.5.2. *From 400 to 900 nm: silicon and the GaAlAs/GaAs family*

These two materials have achieved a good technological maturity under the demands of the integrated circuit market, and as such they are the natural choice for this spectral domain. However, the shape of their absorption functions demonstrates their difference in bandgap nature, which is indirect for silicon and direct for gallium arsenide.

#### 2.5.2.1. *Silicon*

Silicon has an indirect bandgap and its absorption coefficient slowly increases below the corresponding wavelength ($\lambda_G$ = 1,100 nm), reaching the value of $10^3$ $cm^{-1}$ for $\lambda$ = 800 nm. The interface with its natural oxide, $SiO_2$, produces very few electronic states in the bandgap, which is the main factor enabling the development of planar and MOS technologies, and ultimately very high density integrated circuits. On the other hand, this semiconductor has not, as yet, contributed to the design of heterojunction optoelectronics with competitive performance.

Figure 2.17 shows the spectral response of a silicon photodiode designed for optical fiber telecommunications. The structure of the device is chosen (with a depleted region thickness around 30 µm) in order to increase quantum efficiency to between 600 and 900 nm where weakly-attenuating fibers and high performance electro-luminescent emitters are available. It is possible to extend the spectral response slightly towards the infrared by increasing the thickness of the intrinsic region, in particular for detection of the $\lambda$ = 1,060 nm radiation from solid-state neodymium-doped YAG lasers. The drop in efficiency at short wavelengths, due to optical absorption in the surface layer, can also be minimized in the near-ultraviolet by several methods.

The response speed is often limited to less than 1 GHz by the transit time, in particular for components whose response has been extended into the infrared. The dark current is mostly caused by current generation in the depleted region, thanks to the effectiveness of the surface silicon passivation, which is of the order $10^{-7}$ A/cm$^2$.

The maturity of silicon technology allows the construction of a variety of photodiode types, potentially with multiple regions of sensitivity, and offered at a low price. As a result, they are used for short-range telecommunications in numerous professional systems (automation, Lidar, etc.) or mass-market systems (optical disk readers, photographic equipment, etc.). CCD and MOS imagers also form part of the family of silicon-based photodetectors.

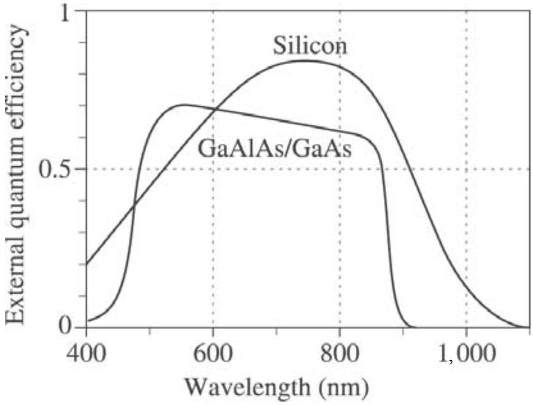


**Figure 2.17.** *Spectral response of photodiodes, made of silicon for optical telecommunications and of experimental GaAlAs/GaAs without antireflection treatment*

### 2.5.2.2. *Gallium arsenide and similar materials*

Gallium arsenide has a direct bandgap; its absorption coefficient is approximately $10^4$ cm$^{-1}$ for $\lambda = 800$ nm and increases very rapidly below the corresponding wavelength $\lambda_G$ ($\lambda_G = 870$ nm). Highly developed for the construction of microwave integrated circuits and diode lasers, it is available as a high-quality substrate for creating defect-free heterojunctions with $Ga_{1-x}Al_xAs$, $Ga_{1-x}In_xAs_{1-y}P_y$ and $Al_xGa_yIn_{1-x-y}As$ alloys.

In fact, there has been interest in heterojunctions from the GaAs/GaAlAs system for photodiode and solar cell fabrication since the 1970s [ALF 70], and components showing good performance have been demonstrated [LAW 79]. A spectral response function is given in Figure 2.17, showing the typical shape for a heterojunction photodiode: rapid cutoff at long wavelengths, absorption in the direct-bandgap i material, internal quantum efficiency close to 100%, and cutoff at short wavelengths due to absorption from the large-bandgap window layer. The density of the dark current is of the order of $10^{-8}$ Acm$^{-2}$ at $V_B/2$ and a product $\eta_i f_t$ of the order of 20 GHz can be achieved for conventionally-structured devices. Alongside heterojunctions, the use of Schottky diodes based on GaAs has also been demonstrated. In addition, the deposition of ITO (indium tin oxide), a conductive and transparent material, has allowed the development of fast photodiodes based on GaAs [PAR 87]. MSM (metal-semiconductor-metal) photodiodes, discussed in Chapter 5, can also be thought of as Schottky photodiodes.

Despite these encouraging and already well-known characteristics, GaAs-based photodiodes have not seen significant development since 2001, except in solar cells. This is probably due to the cheap cost of silicon components whose speed of operation is adequate for real-world systems. This situation may change if the need arises for optical transmissions above 1 Gbit/s within the spectral range covered in this section.

### 2.5.3. *From 900 to 1,800 nm: germanium, GaInAsP/InP…*

This spectral range is mostly used for long distance transmissions and broadband over silica fibers. These systems have driven the development of photodetectors, initially in germanium – which predated the emergence of optical telecommunications – and then in various families of semiconductor compounds, with the GaInAs alloy on InP substrate finally prevailing.

#### 2.5.3.1. *Germanium*

Germanium has an indirect bandgap and its absorption coefficient increases slowly below $\lambda_G$, then more rapidly for $\lambda < 1,500$ nm where its conduction band also has a direct minimum for $\boldsymbol{k} = 0$ (see Figure 2.4). The need for photodetectors for telecommunications was the drive for an important effort in the 1970s in the development of germanium photodetectors [KAN 85]. However, several factors finally brought an end to their application in this domain, along with the increasing use of GaInAs/InP:

– non-existence of materials allowing the construction of heterojunctions;

– limitation of the spectral response around 1,550 nm for fast photodiodes, due to the shape of the absorption spectrum;

– high dark current due to an insufficient control of the surface preparation.

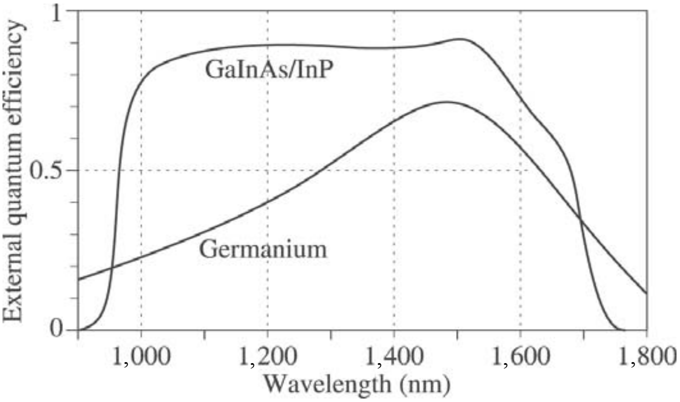


**Figure 2.18.** *Spectral response of state-of-the-art photodiodes in 2000: in germanium for instrumentation and in GaInAs/InP for optical fiber telecommunications*

Figure 2.18 shows the spectral response of a germanium photodiode aimed at scientific instrumentation. The dark current densities are of the order of $10^{-4}$ $Acm^{-2}$ and the cutoff frequencies are limited by the circuit to a few hundred MHz, depending on the active surface.

### 2.5.3.2. *Indium phosphate and related materials*

Solid solution crystals from the $Ga_xIn_{1-x}As_yP_{1-y}$ family formed into a crystal lattice on InP substrates ($y \cong 2.2x$) all have an indirect bandgap, with $Ga_{0.43}In_{0.57}As$ having the smallest bandgap of the family ($\lambda_G$ = 1,650 nm). Its absorption spectrum, as well as that of InP ($\lambda_G$ = 920 nm), is shown in Figure 2.16. The composition of the solid solution can be selected to give a bandgap width anywhere between those two values.

The technology of these materials was initially driven by the development of electroluminescent emitters, electroluminescent diodes and laser diodes, required for long distance transmissions. It has been of benefit to photodetectors which have established themselves as rivals to germanium photodetectors.

The main stages in the "planar" fabrication procedure for PIN heterojunction photodiodes are shown in Figure 2.19. It is analogous to that of the devices described above, but we note in particular that the PIN junction emerges at the surface in the wider-bandgap InP material, which contributes significantly to the reduction in the dark current. The transparency of the substrate allows the possibility of the illumination being in this direction (rear illumination).

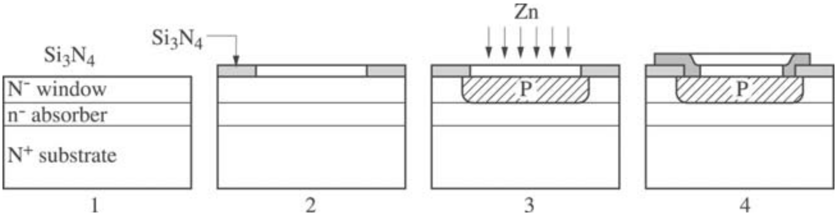


**Figure 2.19.** *Fabrication stages of PIN photodiodes in GaInAs/InP by "planar" technology: 1) epitaxial deposition of the absorber and window layers, 2) deposition of the dielectric and etching of openings, 3) zinc diffusion (p-type impurity), and 4) attachment of metallic contacts*

Figure 2.18 shows the spectral response of a photodiode targeted for optical telecommunications. This diode has had an antireflection treatment applied which allows it to approach an external quantum efficiency of 100% across a large spectral band. The dark current density is of the order of $10^{-5}$ $Acm^{-2}$ and, as we saw in section 2.4.2.2, the cutoff frequency of a photodiode with an optimized structure can surpass 20 GHz while retaining a high quantum efficiency. The spectral response is limited to around 1,700 nm when the absorbing layer is exactly matched to the InP substrate. This can be extended beyond 2,000 nm with the use of unmatched materials, at the expense of the dark current, for certain applications (imaging, spectroscopy, etc.) [LIN 92].

2.5.3.3. *Other materials [PEA 85]*

The integration of the GaAlAsSb family into GaSb substrates was also considered by several groups, towards the end of the 1970s, for the fabrication of photodiodes for optical transmissions. GaSb is a direct semiconductor whose bandgap is close to that of GaInAs ($\lambda_G$ = 1,720 nm). By combining it with GaAlAsSb, it was possible to demonstrate photodiodes with competitive quantum efficiencies, although with a higher dark current (by a factor of ~10) than GaInAs/InP photodiodes. On the other hand, a weaker additional avalanche noise was achieved (see Chapter 3). Nevertheless, the industrial significance of GaInAsP/InP, also used for electroluminescent emitters, seems to have prevented investment in research which might have taken this technology to new levels.

The HgCdTe (MCT, *Mercury Cadmium Telluride*) integrated into CdTe substrates is the dominant material for detection and imaging in the mid-infrared. Attempts were made to re-use the technological achievements in the infrared for the construction of detectors for telecommunications. Although the performances obtained were comparable to those of GaInAs/InP, as in the case of GaInAsSb/GaSb, the potential gains were insufficient to justify its industrial development.

## 2.6. New photodiode structures

### 2.6.1. *Beyond the limits of conventional PIN*

The development in the last decades of new epitaxial techniques for the integration of very thin layers (only a few atoms thick) of semiconductor compounds has enabled the design of optoelectronic devices, including photodetectors with far more complex structures than the simple PIN photodiodes with three useful layers. In this way it has become possible to exceed the limit of the product $\eta_i f_t$, which we examined in section 2.4.1.1 (for the PIN photodiode geometries discussed), whose order of magnitude is 30 GHz for indirect bandgap materials. Several avenues were explored with this aim, all based on making the radiation pass through the narrow i absorption region multiple times (as shown in Figure 2.20) in order to increase the internal quantum efficiency while maintaining a short electronic transit time.

## 2.6.2. *Photodiodes with collinear geometry*

A first approach to increase the absorption of radiation by a layer of thickness $\alpha d << 1$ is to cause the light to pass through the absorbing layer again, after reflection from a mirror at the back (see Figure 2.20), which results in a doubling of the absorbing thickness. Such a mirror can be fabricated either by metalizing the rear face of the structure or, as shown in the figure, by a Bragg semiconductor mirror. These mirrors made from a stack of layers with an alternating sequence of bandgaps (and hence refractive indices) of thickness $\lambda/2n$ allow reflection coefficients approaching 100% to be achieved. They were first used in the fabrication of vertical cavity surface emitting lasers (VCSELs) and more generally in microcavity devices.

One technique for increasing the number of crossings of the absorbing region is to insert the i layer between two mirrors, possibly Bragg mirrors, to form a Fabry Perot cavity (or interferometer), as shown in Figure 2.20.

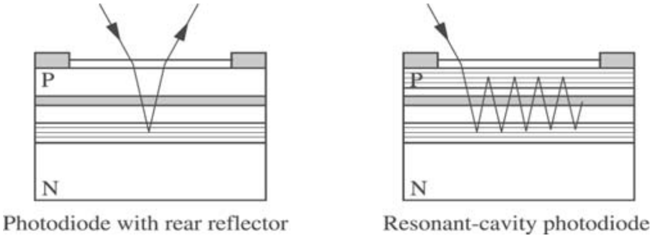


**Figure 2.20.** *Photodiodes of collinear geometry with an improvement in the $\eta_i f_t$ tradeoff*

The analysis of the behavior of such a structure [KIS 91] shows that when the effective distance between the mirrors is an integer number of half-wavelengths of light in the material, and the i layer is placed at an antinode of the electromagnetic field, the absorption can be considerably enhanced, as can the internal quantum efficiency. This enhancement is obtained for a series of resonant wavelengths satisfying the above condition, which gives rise to the name RCE (*Resonant Cavity Enhanced*) photodiodes. However, this resonant behavior has a marked effect on the characteristics of the device:

– the spectral response consists of a series of peaks which become even narrower as the fineness of the cavity is increased, which is in contrast to the normally broad response of conventional photodiodes (see Figures 2.17 and 2.18). RCE photodiodes can be thought of as the monolithic integration of a photodiode and an optical filter whose selectivity can be exploited beneficially, in particular if it can be made tunable;

– for a given wavelength, the resonance condition is satisfied only for certain angles of incidence, such that the directivity is much higher than that of conventional devices.

Analysis of the response speed [UNL 95] shows that a product $\eta_i f_t$ approaching 100 GHz should be obtainable by RCE photodiodes. In addition, other advantages of the use of very thin detector layers are anticipated including:

– the reduction of the contribution of pair generation to the dark current (equation [2.20]);

– the possibility of using partially incommensurate i materials without loss of crystal quality if its thickness is below a critical value.

The experimental results obtained up to now for most of the families of III-V materials have indeed confirmed the potential of RCE photodiodes in terms of quantum efficiency and selectivity, but it turns out to be difficult to show improvement in the $\eta_i f_t$ product.

### 2.6.3. *Waveguide photodiodes*

Another approach consists of using an orthogonal geometry, giving the device the shape of an optical waveguide, which is analogous to that of a diode laser. This is the case in Figure 2.21, where we see an absorbing i layer inserted between several other layers which have a wider bandgap, are transparent and also have a lower refractive index.

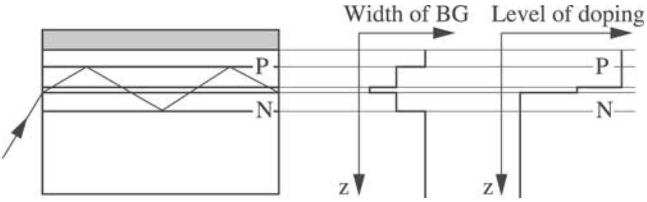


**Figure 2.21.** *Structure, width of the bandgap and doping of the layers, of a waveguide photodiode with orthogonal geometry and with an improved* $\eta_i f_t$ *tradeoff*

In Figure 2.21, a geometric optics representation, the light rays propagate in the structure through a series of total internal reflections, and cross the i layer several times. In reality, since the dimensions of the waveguide are of the order of the wavelength of light in the material, a more rigorous analysis of the electromagnetic field shows that only certain configurations (or *modes*) are supported. The radiation is progressively absorbed with an effective coefficient of absorption depending on the overlap between the guided mode and the absorbing layer; this can be adjusted by the design of the waveguide, but is difficult to reduce below 100 cm$^{-1}$, leading to an internal quantum efficiency of 95% for a waveguide of length 300 µm.

Numerous variations on this principle have been investigated, such as the active part of the photodiode being localized on only part of the waveguide. Figure 2.21 shows a structure with 5 active layers, allowing independent adjustment of the transit time (by varying the thickness of the i depleted region) and the waveguide thickness (which determines the degree of coupling with the incident light and consequently the external quantum efficiency [KAT 92]):

- InP N layer ($\lambda_G = 920$ nm);
- GaInAsP N layer ($\lambda_G = 1{,}300$ nm);
- GaInAs i layer ($\lambda_G = 1{,}650$ nm);
- GaInAsP P layer ($\lambda_G = 1{,}300$ nm);
- InP P layer ($\lambda_G = 920$ nm).

In this way it has been possible to demonstrate a significant gain in the merit factor $\eta_i f_t \cong 30$ for GaInAs. However, it should be noted that the gain in internal quantum efficiency obtained, can in practice be partially undone by the difficulty of coupling the radiation into the mode(s) supported by the waveguide. Although it is in theory possible to couple the mode from a single-mode fiber into a photodiode of single mode with an efficiency of 100%, this can require coupling optics of prohibitive cost.

### 2.6.4. *Traveling-wave photodiodes*

Waveguide photodiodes have the same limitation as conventional structures, in terms of the power/frequency tradeoff discussed in section 2.4.3. This limitation stems from the lumped-element circuit analysis, which is a good approximation as long as the length of the guide remains smaller than the wavelength of the electrical signal ($f = 100$ GHz, $\lambda = 3$ mm). Beyond this, it is necessary to use a distributed element model and to think of a photodiode as a waveguide coupled to a transmission line (see Figure 2.22), terminated at both ends in its characteristic impedance in order to maximize the bandwidth.

Such devices have been proposed and analyzed under the term "traveling-wave photodiodes" [GIB 92]. It follows that, aside from the transit time, the circuit bandwidth is limited by the mismatch between the propagation speeds of the two guided waves, optical and microwave – a limit which is not fundamental in nature, but which is difficult to overcome in practice. The analysis of [GIB 92] predicts values for the product $\eta_i f_t$ of greater than 100 GHz; however, fabrication difficulties appear, until now, to have seriously limited the practical performances relative to the potential performance: the challenge still remains to conceive and then construct structures allowing simultaneous improvement in $\eta_i f_t$ and $Pf^2$ (the figure of merit).

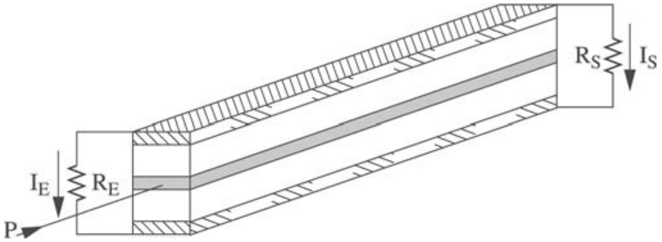


**Figure 2.22.** *Traveling-wave photodiode*

## 2.6.5. *Beyond PIN structures*

Whatever their geometry, the photodiodes considered up to this point have had, by definition, three active regions, with the i region fulfilling both the role of radiation absorption and also the role of the depleted region. It is possible to imagine photodiodes where these functions are separated so that they can be optimized separately, possibly with the additional use of waveguide photodiode or resonant cavity concepts.

### 2.6.5.1. *$P^+p^+N^-N^+$ structures*

The absorbing region, here p+, can be doped without reducing the internal quantum efficiency and the speed, on the condition that it is thin enough to ensure both efficient and rapid transport of photocarriers by diffusion (see sections 2.3.2.2 and 2.4.1.2). An initial example is that of electroluminescent diodes whose optimization leads to a $P^+p^+N^-N^+$ double heterojunction structure, where the emitting layer is designated p+. Under reverse-bias the depleted region grows into the N- region and the photodetection, quantum efficiency and speed characteristics are sufficient for some applications

$(\eta_e \cong 0.6, f_{RC} \cong 300\ MHz)$ [POU 80]. Very fast response speeds have more recently been discovered to be at the expense of low quantum efficiency $(\eta_e \cong 0.06, f_{3dB} \cong 300\ GHz)$ in a conventional geometry, making more use of the high mobility regime of electron transport [ITO 00].

### 2.6.5.2. *$P^+IiN^+$ structures*

The placement of the absorbing i region in a thicker I depleted region, in order to equalize the transit time of electrons generated in the i layer towards the N+ layer and of holes towards the P+ layer, allows us in principle to gain a factor of 2 on the cutoff frequency $f_t$ [UNL 95].

## 2.7. Bibliography

[ALF 70] ALFEROV Z.I., ANDREEV V.M., ZIMOGOROVA N.S., TRET'YAKOV D.N., "Photoelectric properties of $Al_xGa_{1-x}As$ heterojunctions", *Sov. Phys.-Semicond.*, vol. 3, no. 11, p. 1373-1376, May 1970.

[BOW 87] BOWERS J.E., BURRUS C.A., "Ultrawide-band long-wavelength p-i-n photodetectors", *J. Lightwave Technology*, vol. LT-5, no. 10, p. 1339-1350, October 1987.

[FOR 82] FORREST S.R., KIM O.K., SMITH R.G., "Optical response time of $In_{0,53}Ga_{0,47}7As$/InP avalanche photodiodes", *Appl. Phys. Lett.*, vol. 41, no. 1, p. 95-98, July 1982.

[GIB 92] GIBONEY K.S., RODWELL M.J.W., BOWERS J.E., "Travelling-wave photodetectors", *IEEE Photonics Technology Lett.*, vol. 4, no. 12, p. 1363-1365, December 1992.

[ITO 00] ITO H., FURUTA T., KODAMA S., ISHIBASHI T., "InP/InGaAs uni-travelling-carrier photodiode with 310 GHz bandwidth", *Electron. Lett.*, vol. 36, no. 21, p. 1809-1810, 12 October 2000.

[JOH 64] JOHNSON E.O., "Physical limitations on frequency and power parameters of transistors", *IEEE Int. Conf. Rec. pt.5*, p. 27-34, 1965.

[KAN 85] KANEDA T., "Semiconductors and semimetals", *Lightwave Communication Technology,* vol. 22, part D, chap. 3, p. 247-328, Academic Press, 1985.

[KAT 92] KATO K., HATA S., KAWANO K, YOSHIDA J., KOZEN A., "A high-efficiency 50 GHz InGaAs multimode waveguide photodetector", *IEEE J. Quantum Electronics*, vol. 28, no. 12, p. 2728-2735, December 1992.

[KIS 91] Kishino K., Unlu M. S., Chyi J-I., Reed J., Arsenault L., Morkoc H., “Resonant cavity enhanced (RCE) photodetectors”, *IEEE J. Quantum Electronics*, vol. 27, no. 8, p. 2025-2034, August 1991.

[LAW 79] Law H.D., Nakano K., Tomasetta L.R., “State-of-the-art performance of GaAlAs/GaAs avalanche photodiodes”, *Appl. Phys. Lett.*, vol. 35, no. 2, p. 180-182, July 1979.

[LIN 92] Linga K.R., Olsen G.H., Ban V.S., Joshi A.M., Kosonocki W.F., “Dark current analysis and characterization of $In_xGa_{1-x}$ $As/InAs_yP_{1-y}$ graded photodiodes with $x > 0.53$ for response to longer wavelength (> 1.7 µm)”, *J. Lightwave Technology*, vol. 10, no. 8, p. 1050-1055, August 1992.

[MAT 87] Mathieu H., *Physique des semi-conducteurs et des composants électroniques*, Masson, 1987.

[PAR 87] Parker D. G., Say P. G., Hansom A.M., Sibbett W., “110 GHz high-efficiency photodiodes fabricated from indium tin oxide/GaAs”, *Electron. Lett.*, vol. 23, p. 527-528, 1987.

[PEA 85] Pearsall T.P., Pollack M.A., “Semiconductors and semimetal”, *Lightwave Communication Technology*, vol. 22, Part D, Chapter 2, p. 173-247, Academic Press, 1985.

[POU 80] Poulain P., de Cremoux B., “Diffusion limited transient response of heterojunction photodiodes”, *Japanese J. Appl. Phys.*, vol. 19, no. 4, p. L189-L192, April 1980.

[STI 82] Stillman G. E., Cook L.W., Bulman G.E., Tabatabaie N., Chin R., Dapkus P.D., “Long-wavelength (1.3 to 1.6 µm) detectors for fiber-optical communications”, *IEEE Trans. on Electron Devices*, vol. ED-29, no. 9, p. 1355-1371, September 1982.

[SZE 81] Sze S.M., *Physics of Semiconductor Devices*, 2nd edition, John Wiley & Sons, 1981.

[TAK 80] Takanashi Y., Kawashima M., Horikoshi Y., “Required donor concentration of epitaxial layers for efficient InGaAsP avalanche photodiodes”, *Japanese J. Appl. Phys.*, vol. 19, no. 4, p. 693-701, April 1980.

[UNL 95] Unlû S., Onat B.M., Lebleci Y., “Transient simulation of heterojunction photodiodes – Part II: analysis of resonant cavity enhanced photodetectors”, *J. Lightwave Technology*, vol. 13, no. 3, p. 406-415, March 1995.

## Chapter 3

# Avalanche Photodiodes

### 3.1. Introduction

Photodetectors are one of two contrasting families of devices that dominate the field of optoelectronics. Among these, avalanche photodiodes (APDs) are very attractive devices because they benefit from an internal gain due to the multiplication of the charge carriers generated by absorption of incident light. This occurs when these charge carriers cross a region of high electric field ($>10^5$ V/cm), thus acquiring enough energy to ionize the atoms of the crystal lattice, thereby creating new electron-hole pairs which are immediately separated, and can themselves create other electron-hole pairs: this leads step-by-step to an amplification of the primary photocurrent. This mechanism, known by the term *shock ionization* or *impact ionization*, is also at the root of the reversible breakdown of reverse-biased p-n junctions.

In this chapter we will discuss the following subjects:

– theoretical consideration of the avalanche effect (ionization coefficients, multiplication factor, etc.);

– technical conditions necessary for proper operation of APDs;

---

Chapter written by Gérard RIPOCHE and Joseph HARARI.

– APD structure;

– photoelectric properties of APDs;

– state of the art for APDs adapted to different ranges of wavelength, based on Si, Ge, InGaAs, etc.;

– new low-noise APD structures based on superlattices or heterogenous structures (Si/InGaAs).

## 3.2. History

The discovery of solid-state, gas and semiconductor lasers between 1958 and 1962, and the creation at the same time of the first electroluminescent diodes, led to the birth of optoelectronics and to the subsequent boom that it has seen, particularly in the field of optical communications. however to take advantage of the considerable potential of these new light sources, mostly semiconductor lasers and light emitting diodes (with direct modulation at high modulation frequencies, and short response times), there is a need for fast and sensitive detectors operating at the emitted wavelengths. Among the possible candidates, reverse-biased p-n junction photodiodes offer numerous advantages: low bias voltages, small footprint, robustness, ease of use and reliability. These make them very attractive to users compared to photomultipliers which, although benefiting from a high sensitivity thanks to a noise-free internal gain process, suffer from two serious drawbacks for many applications: their size and high operating voltage.

However, the electrical signal generated by the absorption of incident light by a photodiode is, in most cases, too weak to be directly usable and must be amplified. In boosting the signal to a sufficient level, the amplifier stage limits the performance of the receiver in two ways: by reducing the bandwidth of the system and by lowering the global sensitivity due to the additional noise that it introduces.

In order to eliminate these problems and to address the need for amplification of the electric signal generated by the light incident on the photodiode, in 1964 Johnson [JOH 64] highlighted the advantages of primary photocurrent multiplication through the “avalanche effect”; which was reported by McKay and McAfee [KAY 53] in 1953 during their study of reversible breakdown in Si p-n junctions. This internal gain mechanism gives enough improvement to the signal-to-noise ratio to make the use of avalanche photodiodes (APDs) very attractive for many applications, in particular for optical telecommunications, despite the additional noise due to the random nature of the photomultiplication mechanism.

Studies carried out in the 1960s focused on the development of silicon avalanche photodiodes (Si APDs) for 1.06 µm telemetry and for optical communications around 0.85 µm, and also the development of germanium avalanche photodiodes (Ge APDs) whose photosensitivity extends to 1.65 µm. The 1970s saw the optimization of Si APDs for fiber-optic telecommunications at 0.85 µm through the use of “low-noise” designs, as well as the exploratory study of gallium arsenide (GaAs) APDs and those based on GaAs compounds. Since the future of fiber-optic telecommunications is very much based in the near-infrared due to the weak attenuation of optical fibers around 1.3 µm (~2 dB/km) and at 1.55 µm (~0.2 dB/km), several laboratories worked on improving Ge APDs. As the performance of these was limited, even at 1.3 µm, the first investigations were carried out to evaluate the possibilities offered by detectors based on III-V compounds, starting with indium phosphate (InP). The development of working InGaAs/InP APDs required important progress throughout the 1980s, both in the epitaxial growth of heterostructures and in their design and technology, which lead to the creation of devices tailored to the needs of fiber-optic submarine links and in particular to those links that can operate at bitrates of over 10 Gbit/s. From then on, investigations focused on low-noise devices for high bitrates, such as superlattice InGaAlAs/InAlAs APDs or heterogenous Si/InGaAs APDs.

## 3.3. The avalanche effect

The study of reversible p-n junction breakdown by McKay and McAfee in silicon, and by Miller [MIL 55] in germanium, showed that the growth of the dark current, observed when the bias voltage is increased towards the breakdown voltage $V_B$, which is a characteristic of a given device, is caused by a mechanism commonly known as the "avalanche effect", similar to that observed in gas discharges.

The avalanche effect is shown schematically in Figure 3.1. A primary charge carrier, electron or hole, which enters the space charge region (SCR) where there is a fairly strong electric field $E$ ($>10^5$ V/cm), is strongly accelerated and can acquire a mean kinetic energy greater than the characteristic ionization energy $E_i$ of the semiconductor, before colliding with an atom of the crystal lattice, creating an electron-hole pair through ionization and promoting an electron from the valence band to the conduction band. The secondary charge carriers (electron and hole) that are generated, as well as the primary charge carrier, can then independently create new electron-hole pairs, as long as the kinetic energy accumulated by each of the charge carriers between successive collisions is greater than the ionization energy $E_i$. This process is cumulative and can under certain conditions lead to the breakdown of the junction.

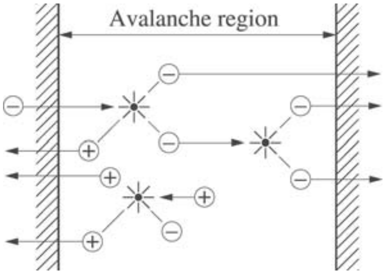


**Figure 3.1.** *Diagram showing the principles of the shock ionization mechanism*

### 3.3.1. *Ionization coefficients*

The ionization energy $E_i$ is, in the simplest case, that of a model with two parabolic energy levels, greater than or equal to $3/2.E_g$, where $E_g$ is the energy corresponding to the width of the bandgap. The ionization energy $E_i$ depends on the type of charge carriers for a given material. If it can be reduced by some means, this can lead to the preferential ionization of one type of charge carrier, as we will see later.

The ionization coefficients of electrons $\alpha_i$ and of holes $\beta_i$ are defined as the average number of ionizing collisions produced per unit length (cm) by a charge carrier moving parallel to the constant electric field $E$, and they quantify the ionizing power of the charge carriers. Depending primarily on the electric field, they grow very rapidly with the field. Their variations, determined experimentally, follow empirical relationships of the form:

$$\alpha_i, \beta_i = a_i \exp\left(\frac{-b_i}{E}\right) \qquad [3.1]$$

where $a_i$ and $b_i$ are parameters that depend on the material and on the type of charge carrier. Figures 3.2a and 3.2b show typical $\alpha_i$, $\beta_i = f(1/E)$ curves, for silicon and germanium on the left and for InP on the right, which are the base materials for APDs developed for optical communications. Silicon is unusual in two ways: the electrons are the most easily ionized carriers and the ratio of the ionization coefficients $k = \beta_i/\alpha_i$ is very high for relatively weak electric fields ($E \sim 3 \times 10^5$ Vcm$^{-1}$). The nature of the carriers triggering the multiplication and the ratio $k$ play an important role in the design of devices, and directly influence the photoelectric performance of APDs.

It is worth noting that the variation seen in the experimental curves published in the literature is a good illustration of the difficulties inherent in determining the coefficients $\alpha_i$ and $\beta_i$ from measurements of photomultiplication or noise.

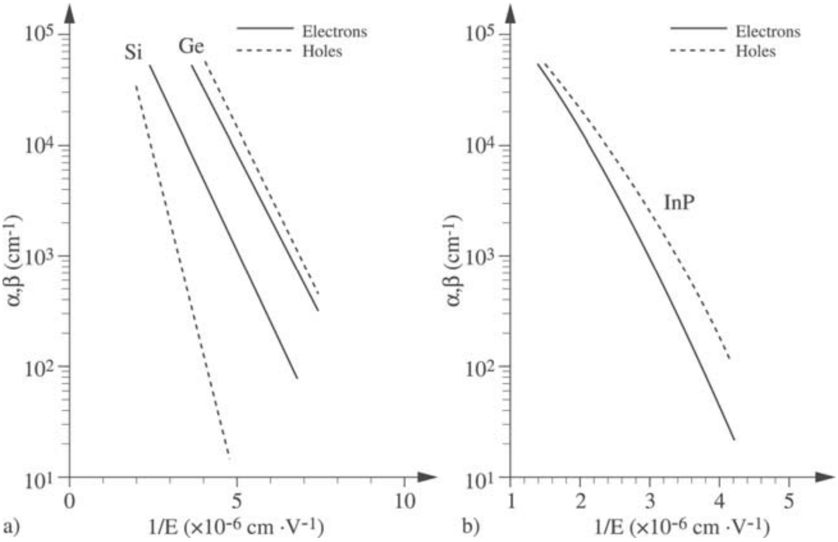


**Figure 3.2.** *Ionization coefficients in silicon, germanium and InP as a function of the inverse of the electric field 1/E*

### 3.3.2. *Multiplication factors*

The impact ionization mechanism leads to a multiplication in the number of charge carriers injected into the SCR, with this taking place in the avalanche region where there is a high electric field $E$. For each type of charge carrier, the multiplication factor $M_n$ or $M_p$ is defined as the ratio of the number of charge carriers extracted to the number of charge carriers injected. The following approximations are normally used in the calculation:

– the ionization coefficients $\alpha_i$ and $\beta_i$ only depend on the electric field $E$, which itself is only a function of the $x$ coordinate for a given junction;

– recombination of carriers is negligible in the depleted region;

– mobile charge carriers do not disturb the electric field distribution;

– the levels of thermal and optical carrier generation in the depleted region are not included.

Calculation of the multiplication factors $M_n$ and $M_p$ is carried out by solving the continuity equations for the electron current $J_n$ and the hole current $J_p$ under a strong electric field. The system of equations describing steady-state behavior is as follows:

$$\frac{dJ_n(x)}{dx} = -\frac{dJ_p(x)}{dx} = \alpha_i(x) \cdot J_n(x) + \beta_i(x) \cdot J_p(x) \qquad [3.2]$$

$$J = J_n(x) + J_p(x) = C^{ste} \qquad [3.3]$$

which can be rewritten as:

$$\frac{dJ_n(x)}{dx} - \left[\alpha_i(x) - \beta_i(x)\right] J_n(x) = \beta_i(x) J \qquad [3.4]$$

Solving this differential equation for the case of multiplication triggered by a pure injection of electrons from the P+ region, by integrating $J_n(x)$ over the avalanche region (between $x = 0$ and $x = w$), leads to the expression:

$$1 - \frac{1}{M_n} = \int_0^w \alpha_i(x) \exp\left[-\int_0^x \left(\alpha_i(x) - \beta_i(x)\right) dx'\right] .dx \qquad [3.5]$$

with $M_n J_{n0}$, where $J_{n0}$ is the current of injected electrons at $x = 0$.

In the case of an injection of holes from the $N^+$ region, the multiplication factor $M_p$ is given by a similar expression:

$$1 - \frac{1}{M_p} = \int_0^w \beta_i(x) \exp\left[-\int_x^w \left(\alpha_i(x) - \beta_i(x)\right) dx'\right] .dx \qquad [3.6]$$

with $M_n\, J_{pw}$, $J_{pw}$ being the current of injected holes at $x = w$.

In reality, carriers of both types are normally injected into the avalanche region, leading to a multiplication factor somewhere between the two which takes into account the "weighting" of each contribution. It also follows that the experimental determination of the ionization coefficients $\alpha_i$ and $\beta_i$ from measurements of multiplication factors is very sensitive, since in addition to a precise knowledge of the electric field distribution $E(x)$, it requires the separation of the effects of both types of charge carrier. In the case where the ionization coefficients are equal ($\alpha_i = \beta_i$), the two expressions take the form:

$$1 - \frac{1}{M} = \int_0^w \alpha_i\left(x\right)dx \qquad [3.7]$$

In the case of a PIN junction with a constant electric field in the depleted region, for a given bias voltage $V$, the above equation becomes:

$$1 - \frac{1}{M} = \alpha_i \cdot w \qquad [3.8]$$

where $\alpha_i\,(E)$ is a function of the bias voltage $V$ as $w$ is constant, hence:

$$E = \frac{V}{w} \qquad [3.9]$$

### 3.3.3. *Breakdown voltage*

As the bias voltage increases, the electric field increases and approaches the critical value for which the dark current grows extremely fast: this is the volume breakdown of the junction, which occurs for a bias voltage $V_B$ – the breakdown voltage. These conditions occur when the multiplications factors $M_n$, $M_n \rightarrow \infty$ which occurs for:

$$1=\int_0^w \alpha_i(x)\exp\left[-\int_0^x(\alpha_i(x)-\beta(x))dx'\right]dx \qquad [3.10]$$

and:

$$1=\int_0^w \beta_i(x)\exp\left[\int_x^w(\alpha_i(x)-\beta(x))dx'\right]dx \qquad [3.11]$$

respectively.

In addition, Sze and Gibbons [SZE 66], studying the behavior of P+N and N+P junctions in the avalanche regime, established a simple semi-empirical relationship between the breakdown voltage $V_B$ and the concentration N in the weakly-doped region of an abrupt junction, or the concentration gradient *a* in a gradual junction. These relations can be applied to different materials. The experimental results are well-matched by the following expressions:

$$V_B=60\left(\frac{E_G}{1.1}\right)^{\frac{3}{2}}\cdot\left(\frac{N}{10^{16}}\right)^{\frac{3}{4}} \qquad [3.12]$$

for an abrupt junction and:

$$V_A=60\left(\frac{E_G}{1.1}\right)^{\frac{6}{5}}\cdot\left(\frac{a}{3.10^{20}}\right)^{\frac{2}{5}} \qquad [3.13]$$

for a gradual junction, where $E_g$ is the width of the bandgap of the material in question, expressed in eV.

Finally, we point out the empirical relationship established by Miller [MIL 55] between the multiplication factor M and the breakdown voltage $V_B$:

$$1-\frac{1}{M}=\left(\frac{V}{V_B}\right)^n \qquad [3.14]$$

where $n$ is an exponent which depends on the material, the type of junction and the nature of the charge carriers triggering the multiplication. This expression satisfactorily describes the variation of the multiplication factor $M$ with the bias voltage $V$, and it is often applied to classical $P^+N$ and $N^+P$ junctions.

## 3.4. Properties of avalanche photodiodes

The behavior of avalanche photodiodes differs from that of standard PIN photodiodes only in terms of the properties that are involved in the avalanche multiplication mechanism: the gain, the noise linked to the photocurrent, and the frequency response, these being at operating voltages in the pre-avalanche regime of the junction, in contrast to the low voltages at which PIN photodiodes are normally used.

### 3.4.1. *Current-voltage characteristics and photomultiplication*

For a PIN junction, the photocurrent $I_{ph}$ generated by the light incident on the active region barely varies with the operating voltage. In contrast, for APDs, above a certain bias voltage the photocurrent $I_{ph}$ grows steadily, reaching its maximum value when the bias voltage is equal to the breakdown voltage $V_B$ of the junction. The gain $M$ is given by the equation:

$$M = \frac{I_{ph}(V)}{I_{ph}(V_0)} = \frac{I(V) - I_d(V)}{I(V_0) - I_d(V_0)} \qquad [3.15]$$

where $I_{ph}$, $I$ and $I_d$ are respectively the photocurrent, the total current flowing across the junction, and the dark current; $V$ and $V_0$ are the operating voltage and the reference voltage for which $M = 1$. In the following, we will refer to $I_{ph0}$ and $I_{d0}$ as the primary photocurrent and the primary dark current (i.e. the non-multiplied currents). These are denoted as $I_{ph}(V_0)$ and $I_d(V_0)$ in equation [3.15]. This equation allows us to plot the $M(V)$ characteristics of an APD measured under well-defined experimental conditions. These may have a decreasing or increasingly strong influence on the measured values. In particular, although an APD may behave fairly linearly for relatively low values of gain ($M < 10$), for which $M$ depends little on the incident light power, this contrasts with significant saturation effects that are seen at high gains. The maximum gain obtained ($M_{\max}$) is very sensitive to the conditions of the measurement. A reduction in $M_{\max}$ can be partially explained by a reduction in the avalanche electric field caused by an opposing electric field created by the mobile charge carriers, which are present in significant numbers in the depleted region. Also, close to breakdown, the applied bias voltage is reduced because of the voltage drop across the equivalent resistance $R$ of the diode.

Taking these considerations into account, equation [3.14], which applies for weak values of the current $I$, can for high currents be written:

$$1-\frac{1}{M}=\left[\frac{V-\left(I_{d0}+I_{ph0}\right)\cdot M\cdot R}{V_B}\right]^n \qquad [3.16]$$

for $V = V_B$, $M = M_{\max}$ and this can be simplified to:

$$M_{\max}\approx\sqrt{\frac{V_B}{n\left(I_{d0}+I_{ph0}\right)R}} \qquad [3.17]$$

This expression shows the value of devices with a weak dark current ($I_{d0} << I_{ph0}$) in order to promote photomultiplication. In these conditions, the maximum gain is inversely proportional to the square root of the primary photocurrent $I_{ph0}$, and hence the power of the incident light $P_0$:

$$M_{\max}\propto\frac{1}{\sqrt{I_{ph0}}} \text{ or } M_{\max}\propto\frac{1}{\sqrt{P_0}} \qquad [3.18]$$

Taking into account equation [3.12], which expresses the breakdown voltage $V_B$ as a function of the concentration $N$, we find:

$$M_{\max} \propto N^{\frac{3}{8}} \qquad [3.19]$$

Weak concentrations $N$ in the depleted region, corresponding to relatively high breakdown voltages $V_B$, would be of benefit in obtaining high gains $M$. However, for convenience of use, the diode breakdown voltage is limited, where possible, to the range 50–150 V.

Finally, for devices with a strong dark current ($I_{d0} >> I_{ph0}$), the limitation on the gain is imposed by the dark current and in these conditions, which can occur with Ge APDs, we have:

$$M_{\max} \approx \sqrt{\frac{V_B}{n \cdot I_{d0} \cdot R}} \qquad [3.20]$$

### 3.4.2. *Noise in avalanche photodiodes*

Noise in detectors is discussed in Chapter 7 and we will limit ourselves here to revisiting the main results for noise associated with the operation of APDs in the avalanche regime. In the absence of multiplication, the APD behaves like a traditional photodiode whose noise is essentially shot noise caused by the dark current $I_{d0}$ and the primary photocurrent $I_{ph0}$ across the junction, and whose spectral density is:

$$\left\langle i^2 \right\rangle = 2 \cdot q \cdot \left(I_{d0} + I_{ph0}\right) \qquad [3.21]$$

When the APD operates in the avalanche system, the currents across the junction undergo a multiplication $M$, and hence the spectral density of the noise is multiplied by $M^2$. The random nature of the impact ionization process results in a fluctuation in the gains associated with the different carriers, contributing to the appearance of extra noise linked to the multiplication. The spectral density of the shot noise of the multiplied current can be written:

$$\left\langle i^2 \right\rangle = 2 \cdot q \cdot \left(I_{d0} + I_{ph0}\right) \cdot M^2 \cdot F\left(M\right) \tag{3.22}$$

where *F(M)*, the excess noise factor, represents the contribution of multiplication to the noise. The value of *F(M)* depends on the characteristics of the p-n junction and on the nature of the primary carriers. *F(M)* expressions for the two types of primary injection, electrons and holes, have been derived by McIntyre [INT 66] under the assumption of an avalanche region with a constant electric field. These are expressed as a function of *M* and the ratio of the ionization coefficients k = $\beta i / \alpha i$, which is considered constant. Instances of a pure injection of electrons or of holes are of particular interest. The excess noise factor for a primary injection of electrons can be expressed by the following:

$$F_n\left(M\right) = M_n \left[1 - \left(1 - k\right)\left(\frac{M_n - 1}{M_n}\right)^2\right] \tag{3.23}$$

and for a primary injection of holes:

$$F_p\left(M\right) = M_p \left[1 - \left(1 - \frac{1}{k}\right)\left(\frac{M_p - 1}{M_p}\right)^2\right] \tag{3.24}$$

Where $M_{n,p} >> 1$, these equations can be expressed in the following simplified forms:

$$F_n\left(M\right) = kM_n + 2\left(1 - k\right) \tag{3.25}$$

$$F_p\left(M\right) = \frac{1}{k} M_p + 2\left(1 - \frac{1}{k}\right) \tag{3.26}$$

and for very high values of multiplication:

$$F_n(M) \rightarrow kM_n \text{ and } F_p(M) \rightarrow \frac{1}{k} M_p \qquad [3.27]$$

In the specific case of ionization by only one type of carrier, electrons or holes, the excess noise factor takes a minimum value of *F(M)*=2, independent of *M*. Attempts can be made to approach this value through modification of the ionization energy or by careful design of the device. The typical variation of the noise factor as a function of *M* for different values of *k* is shown in Figure 3.3.

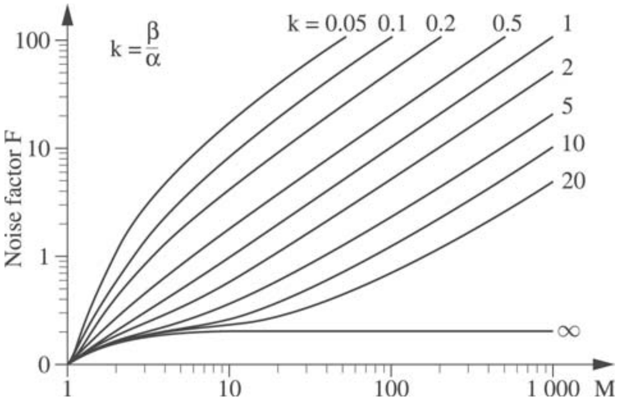


**Figure 3.3.** *Excess noise factor as a function of different values of k for a pure injection of holes*

In order to obtain low excess noise values, two conditions must be met: the multiplication must be triggered by the more ionizing carrier, and a material must be used whose ionization coefficients are very different. Among the normal semiconductors (Si, Ge, GaAs, InP, etc.), only silicon satisfies these two requirements. Furthermore, consideration of the excess noise factor allows us to approach high values of the ratio $k = \beta_i / \alpha_i$, with this parameter playing an important role in the optimization of APDs. Finally, in the literature, the excess noise factor *F(M)* is often written in the form:

$$F(M) = M^x \qquad [3.28]$$

which comes directly from the spectral density of the noise, expressed as:

$$\left\langle i^2 \right\rangle = 2 \cdot q \cdot I_{ph0} \cdot M^{2+x} \qquad [3.29]$$

The exponent $x$, which depends on the value of the $k$ ratio, allows us to compare the different devices. This can often be found in device specifications alongside the value of $F(M)$ for a given $M$. For $k = 1$, the excess noise factor is $F(M) = M$ for the two types of injection, and the exponent $x$ therefore takes the value $x = 1$. For low values of $k$, the multiplication is triggered by electrons, whereas for high values of $k$, the injection of holes is favored.

### 3.4.3. *Signal-to-noise ratio in avalanche photodiodes*

In wideband applications, the limiting noise of traditional photodiodes comes from the thermal noise associated with the load resistance $R_L$ or, more commonly, from the noise of the first stage of the amplification circuit required to make use of the electrical signal produced by the photodiode. For some application types, the internal gain of APDs allows us to improve the signal-to-noise ratio (SNR) of the detector. This detector, consisting of the APD and amplifier, includes two sources of noise outside the diode, independent of the multiplication (the load resistance and the amplifying stage), given by the equation:

$$\left\langle \overline{i_{ext}^2} \right\rangle = \frac{4 k_B T_e}{R_L} \cdot \Delta f \qquad [3.30]$$

where $T_e$ is the equivalent temperature of the noise, and also includes sources of shot noise connected with the multiplication process:

$$\left\langle \overline{i_M^2} \right\rangle = 2q\left(I_{d0} + I_{ph0}\right) \cdot M^2 \cdot F\left(M\right) \cdot \Delta f \qquad [3.31]$$

where $\Delta f$ is the bandwidth of the system.

The mean squared value of the photocurrent is given by:

$$\overline{i_{ph}^2} = \frac{1}{2} m^2 \cdot I_{ph0}^2 \cdot M^2 \qquad [3.32]$$

where *m* is the modulation index of the incident light signal. The SNR for a system with bandwidth *Δf* therefore follows the relationship:

$$\frac{S}{B} = \frac{\frac{1}{2}\left(m^2 \cdot I_{ph0}^2 \cdot M^2\right)}{\left[2q\left(I_{d0} + I_{ph0}\right) \cdot M^2 \cdot F\left(M\right) + \frac{4k_B T_e}{M^2 \cdot R_L}\right] \cdot \Delta f} \qquad [3.33]$$

which can be rearranged into the following form:

$$\frac{S}{B} = \frac{\frac{1}{2}\left(m^2 \cdot I_{ph0}^2\right)}{\left[2q\left(I_{d0} + I_{ph0}\right) \cdot F\left(M\right) + \frac{4k_B T_e}{M^2 \cdot R_L}\right] \cdot \Delta f} \qquad [3.34]$$

demonstrating the advantage of APDs over PIN photodiodes, since for low values of *F(M)* the thermal noise contribution is divided by $M^2$.

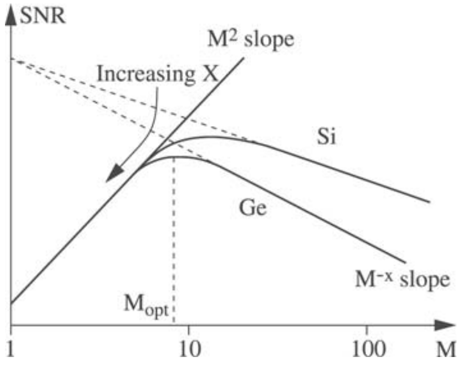


**Figure 3.4.** *Variation of the SNR as a function of the gain M for two semiconductors, silicon and germanium*

As the gain increases, the SNR grows and reaches a maximum for an optimal gain value $M_{opt}$, which depends closely on the point at which the multiplied shot noise and the thermal noise are equal, such that, in the case where $I_d0 << I_{ph}0$:

$$M_{opt} \approx \left( \frac{2k_B T_e}{q I_{ph0} R_L} \right)^{\frac{1}{2+x}} \qquad [3.35]$$

For higher multiplication values, the SNR drops, with the contribution of shot noise becoming the dominant factor. The variation of the SNR with the gain is shown in Figure 3.4. The optimal gain value $M_{opt}$ depends on the incident radiation power $P_0$ through the equation:

$$I_{ph0} = \frac{q \cdot \eta \cdot \lambda}{h \cdot c} P_0 \qquad [3.36]$$

As a result, $M_{opt}$ will increase as $P_0$ decreases. Thus, APDs allow us to improve the sensitivity of photodetectors as long as the shot noise remains lower than the thermal noise entering the amplifier stage.

### 3.4.4. *Speed, response time and frequency response of avalanche photodiodes*

The ability of avalanche photodiodes to detect and amplify optical impulses with short rise and fall times, or light signals modulated at high frequency, is a topic of research for numerous applications. The response time of an APD operating in the avalanche regime is limited by:

– the transit time $t_{tr}$ of carriers in the SCR;

– the diffusion time $t_{Dif}$, the time taken by the electron-hole pairs created outside the depleted region to reach the p-n junction;

– the RC time constant of the APD/load system;

– the avalanche build-up time $t_{av}$, the time needed to reach a stable multiplication value for a given bias voltage.

The transit time $t_{tr}$, the diffusion time $t_{\mathrm{Dif}}$ and the RC time constant all depend on the width $w$ of the bandgap. The transit time $t_{tr}$ can be expressed by the equation:

$$t_{tr} = \frac{w}{v_{sat}} \qquad [3.37]$$

where $v_{sat}$ is the saturation velocity for the slowest charge carriers.

In order to increase the speed of APDs for high-frequency applications or broadband applications, we will seek to reduce the thickness $w$. This reduction will in turn produce an increase in the capacitance $C_j$ of the junction and a reduction in the response coefficient, because the light will be absorbed in a thinner width of material. This results in a compromise between the quantum efficiency and the speed which must be treated on a case by case basis, in a similar way to the tradeoff in PIN photodiodes (see section 2.4).

The diffusion time $t_{Dif}$ of electron-hole pairs towards the junction is dominated by the lifetime of the minority charge carriers in the lightly-doped region of the junction. Its value (between a few µs and a few hundred µs depending on the material) makes some devices unusable for many applications. This limitation appears for certain Si and Ge APDs. In practice, as far as possible, the thickness $w$ of the depleted region is adjusted so that all the light is absorbed there. This result is approximately achieved for:

$$w = \frac{2}{\alpha_\lambda} \qquad [3.38]$$

where $\alpha_\lambda$ is the absorption coefficient for light of a given wavelength $\lambda$ in a semiconductor.

The RC time constant is only significant at high frequencies. In this case R~50 Ω, the value of the load resistance. The capacitance of the junction $C_j$ is related to the area $A$ of the junction and the width $w$ of the depleted region by:

$$C_j = \frac{\varepsilon \cdot A}{w} \qquad [3.39]$$

where $\varepsilon$ is the permittivity of the semiconductor. As a result, we need to minimize the area of the active region and that of the guard ring, and select a value of $w$ compatible with an acceptable transit time $t_{tr}$. This optimization is equivalent to that for PIN photodiodes (see section 2.4).

As the multiplication process is not instantaneous, the response time of APDs can be greater than that of PIN photodiodes of comparable structure, and is very sensitive to the ionization conditions. Hence, in the case of ionization by only one type of carrier, which is similar to that of silicon, the impulse response time is controlled by the transit time across the depleted region of the non-ionizing secondary carriers created by the primary carriers. Under these conditions, the bandwidth of the APD, independent of the multiplication $M$, is not very different from that of a PIN photodiode. When the two types of carriers contribute to the multiplication, the steady-state value of $M$ will be reached after a time which will be longer if $M$ is higher and if the ionization coefficients of the carriers $\alpha_i$ and $\beta_i$ are close to each other. This is because the high number of carrier-lattice interactions means the carriers "bounce" back and forth, prolonging the time that secondary carriers are present in the avalanche region. The avalanche build-up time $t_{av}$ is therefore expressed by the equation:

$$t_{av} = M\tau \qquad [3.40]$$

where $\tau$ is the effective transit time of a carrier in the avalanche region between two ionizing collisions. This depends very strongly on the value of $k = \beta_i / \alpha_i$ and can become very small if $\alpha_i$ and $\beta_i$ are very different. Emmons [EMM 67] analyzed the frequency response of PIN photodiodes operating in the avalanche system and established the following expression:

$$\tau = N_b \left( \frac{\beta_i}{\alpha_i} \right) \cdot \tau_1 \qquad [3.41]$$

where $N_b$ is a number which varies slowly between 1/3 and 2 as the ratio $k = \beta_i / \alpha_i$ varies from $10^{-3}$ to 1, and $\tau_1$ is the actual transit time in the avalanche region. A low value of $k$ minimizes the effective transit time $\tau$ and is also favorable for the response time and for the excess noise factor. Emmons [EMM 67] showed that, for $M > \alpha_i/\beta_i$, the gain-bandwidth product is constant. The gain $M$ varies with frequency according to the following law:

$$M(\omega) = \frac{M_0}{\left[1 + M_0^2 \cdot \omega^2 \cdot \tau^2\right]^{\frac{1}{2}}} \qquad [3.42]$$

where $M(\omega)$ and $M_0$ are the values of the gain as a function of the frequency $f = \omega/(2\pi)$ and for a constant signal. The gain-bandwidth product is given by:

$$M(\omega) \cdot \Delta f = \left(N_b \cdot k \cdot \tau_1\right)^{\ 1} \qquad [3.43]$$

and depends on the material through $k$ and the characteristics of the avalanche region through $\tau_1 = l_a/v_{sat}$, where $l_a$ is the length of the avalanche region. It follows that the product will be much larger if the ratio $k$ of the coefficients of ionization is weak and if the avalanche region is narrow. A high value of the gain-bandwidth product is desirable as it allows a higher value of gain for a given frequency.

## 3.5. Technological considerations

APDs work best when they are biased with a voltage close to the volume breakdown voltage in the active region. As a result there is a very intense electric field (~4-5 x $10^5$ V/cm) present in the avalanche region, which must be as homogenous as possible throughout the active region. Any inhomogenity present in the SCR that can cause a local increase in the electric field will lead to a premature breakdown which will degrade the APD performance. These inhomogenities can be caused by the structure itself, as in a traditional planar junction obtained by localized diffusion in a circular geometry, or linked to imperfections in the crystal structure. Regardless of their origin, they always need to be eliminated.

In the first case, the use of junctions with a guard ring or the use of a *mesa* structure allows the suppression of peripheral breakdown, as long as a geometry is used which is free from angular parts liable to cause point effects. The suppression of crystal defects depends closely on the progress of technology, particularly in the fields of crystallogenesis and techniques of epitaxial growth and doping.

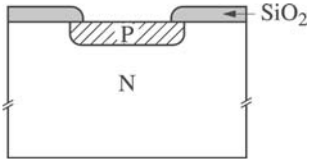


**Figure 3.5.** *Cross-section of a planar photodiode*

### 3.5.1. *Guard ring junctions*

For a junction constructed using planar technology (Figure 3.5), the electrical field inside the depleted region is higher in the curved region around the edge of the diffused area than in the active region, as can be seen by solving Poisson's equation [HAR 91a]. Breakdown due to edge effects therefore occurs for a voltage sufficiently below the breakdown voltage of the active region so that light falling on the central region is not amplified or barely amplified. On the other hand, if the illuminated region extends beyond the diffused area, the increase in the injected global photocurrent due to the high electric field present in a very narrow annular region around the edge can give the illusion of a photomultiplication effect. Regardless, such a photodiode cannot operate as an APD. In order to suppress peripheral breakdown, two kinds of structures are commonly used, which have annular guard rings modifying the electric field distribution around the outer perimeter.

In the first structure (see Figure 3.6a), the guard ring is formed by an annular diffused junction which is relatively weakly doped, and quite deep in order to increase the radius of curvature and thus reduce the peripheral field; the active region is obtained by a shallower diffusion at a higher concentration, partially overlapping the guard ring. The breakdown voltage of the gradual junction of the guard ring is much larger than that of the active region, because for a given bias voltage, the extent of the SCR is larger in the former than in the latter.

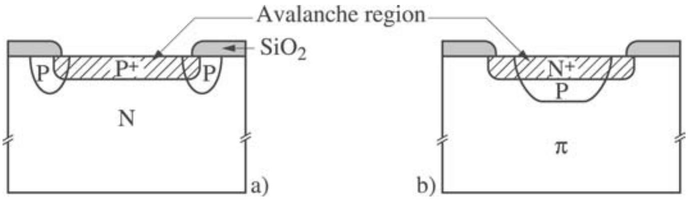


**Figure 3.6.** *Cross-section of (a) an APD with a traditional guard ring and (b) a PπPN+ APD*

In the second structure (see Figure 3.6b), the guard ring and active region junctions are created by diffusion or ion implantation within a weakly-doped substrate $\pi$. The n$^{+}$–$\pi$ guard ring junction which surrounds the n$^{+}$p junction of the active region is much deeper than the latter, but the weak concentration in the $\pi$ substrate precludes breakdown. This is the P$\pi$PN$^{+}$ structure. When the n$^{+}$–$\pi$ junction is not very deep, or even superficial, the addition of a guard ring of the first type is necessary to avoid peripheral breakdown. This is referred to as a P$^{+}\pi$P$\pi$N$^{+}$ structure.

### 3.5.2. *"Mesa" structures*

In mesa structures, the footprint of the active region of the APD is defined by chemical etching or by dry etching followed by surface reconstruction using a suitable chemical bath. This structure, shown in Figure 3.7, offers the sometimes deceptive advantage of simplicity. It allows us to reduce the capacitance of the device, while eliminating the guard ring junction. However, the extension of the avalanche region at high electric fields outside the edges of the etched region can lead to migration of susceptible ion species, causing a growth in the dark current which is often accompanied by instability of the diode. In order to avoid these drawbacks, different passivation techniques have been used with moderate success, which makes such a structure difficult to commercialize on account of its lack of reliability.

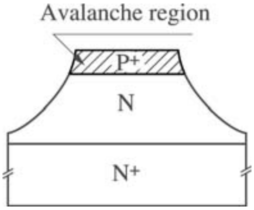


**Figure 3.7.** *Cross-section of a mesa APD*

### 3.5.3. *Crystal defects and microplasmas*

Crystal defects such as dislocation, stacking faults, metallic precipitation, etc., present in a PN junction introduce localized or microplasma breakdowns: isolated regions of a few µm$^2$ where high current densities cross, and which appear at well-defined bias voltages which may or may not be close to the volume breakdown voltage. Unless these microplasmas appear at a voltage very close to the breakdown voltage, their presence decreases APD performance by limiting the photomultiplication factor to that of the breakdown voltage of the first microplasma. At higher voltages, the noise linked to electrical discharges triggered by the random injection of charge carriers in the microplasma is so significant that it precludes any interest in the use of such diodes. In the case of Si APDs, inspection under a microscope allows the microplasmas to be visualized, appearing in the form of bright red spots. In the absence of microplasmas, a diffuse, pale light uniformly illuminates the whole active surface when the breakdown voltage is reached. Observation of the current induced in a scanning electron microscope also allows the detection of inhomogenities in the electric field. In addition, the presence of microplasmas leads to vertical jumps of several µA in the inverse characteristics $I_d(V)$, interspersed with resistive sections in the relationship, thus enabling the electrical sorting of APDs through systematic measurement of the $I_d(V)$ characteristics via probes. Microplasmas can be eliminated using substrates with a low level of dislocations and non-aggressive fabrication techniques tailored to the specific materials used. The construction of high-performance, reliable devices with a high fabrication yield depends closely on the quality of the base material and on the standard of technology used in each stage of the fabrication.

## 3.6. Silicon avalanche photodiodes

Silicon, an indirect-bandgap semiconductor, with a bandgap energy $E_g$ = 1.1 eV at ambient temperature, is sensitive to light in the wavelength range of 0.4–1.1 µm. The absorption length, considerably less than 1 µm at $\lambda$ = 0.4 µm, varies between 10 and 10 µm in the wavelength range 0.8–0.9 µm, the first transmission window for optical fibers (attenuation ~ 2 dB/km), and is more than 200 µm at $\lambda$ = 1.06 µm, which is the emission wavelength of the Nd lasers that are very popular in military applications. Also, the ionization coefficients of electrons and holes are very different, with electrons having the higher ionization energy ($\alpha_i >> \beta_i$). These characteristics play an important role in the design of devices targeted at different applications, with the aim of achieving a weak excess noise factor and an acceptable compromise between sensitivity and response time.

### 3.6.1. *Si $N^+P$ APDs*

Si APDs with an abrupt-junction $N^+P$ structure are fabricated on a P substrate with a concentration around $10^{16}$ cm$^{-3}$. As the width of the depleted region is between 2 and 3 µm, these APDs are well-suited to the detection of light signals with wavelengths below 0.7 µm (ruby lasers emit at 0.63 µm). Commercialized in 1966, they were initially developed for military applications such as laser telemetry at 1.06 µm, taking advantage of an atmospheric transmission window allowing targeting at 20–25 km with high precision (of the order of a meter).

For the first APDs we will study, the guard ring and active region junctions are made by a double diffusion of phosphor impurities, masked using a layer of silica obtained by thermal oxidation; the fabrication conditions for the different stages of the construction procedure have been developed to ensure high performance in terms of the dark current and photomultiplication, along with a high fabrication yield. The homogenity and reproducibility of phosphor ion implantation used as a predeposition technology have greatly helped the achievement of microplasma-free APDs. Additionally, the choice of substrates with fairly high dislocation levels (~3,000 $cm^{-2}$), but which are stable under various thermal treatments at high temperatures (T ≈ 1,000–1,200°C), has allowed the attainment of fabrication yields considerably higher than 50%, whereas they were at best around 20% with the “zero dislocation” substrates advocated in the literature.

These APDs, with a breakdown voltage $V_B \sim 60 - 80$ V, offer $M(V)$ characteristics which vary extremely rapidly as breakdown is approached, with the exponent $n$ in the Miller equation taking a value close to 4. At short wavelengths absorbed in the $N^+$ layer, the multiplication is initiated by holes. For $\lambda > 0.7$ µm, the multiplication is mostly initiated by electrons and takes higher values. At the operating point, $M \sim 50$ to 100, with the gradient $dM/dV$ very large. To avoid issues with variations in $M$, the ratio $V/V_B$ must be kept constant. Since the breakdown voltage $V_B$ follows variations in temperature, some system of regulation is required. The involvement of a Zener diode with a breakdown voltage close to that of the APD has allowed the multiplication to be regulated within the temperature range from –30°C to +80°C. Other regulation systems make use of the noise linked with multiplication.

The impulse response of Si $N^+P$ APDs (see Figure 3.8) is dominated by the diffusion time $t_{Diff}$ of the charge carriers generated outside the depleted region. It consists of a brief rise to the signal (<0.5 ns) followed by a relatively long tail, with the duration of the impulse generally being insufficient to reach the steady-state value. We then see a similar behavior for the fall of the signal. The diffusion time limits the repetition frequency, and this can be improved by limiting the thickness of the absorbing region with the use of epitaxial P/P+ structures or by localized thinning of the substrate.

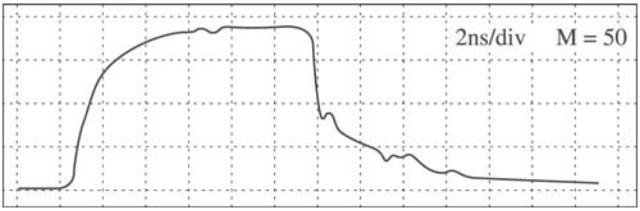


**Figure 3.8.** *Typical impulse response of a Si N+P photodiode*

### 3.6.2. *Si $N^+P\pi P^+$ APDs*

To compensate for defects in the N+P structure, in 1966 Ruegg [RUE 66] showed the potential of an $N^+P\pi P^+$ structure, which combines the advantages of an N+P structure (internal gain) with those of a PIN structure (high quantum efficiency and speed). The result of this is that the depleted region consists of two adjacent regions: one relatively narrow, with an intense electric field, where the multiplication of injected charge carriers takes place, and another relatively extensive region, with a weak electric field, where light is absorbed and the photocarriers collected. This separation of the roles of multiplication and absorption gives the structure a high gain-bandwidth product. Absorption takes place in the very weakly-doped $\pi$ region ($<10^{14}$ cm$^{-3}$), whose thickness is adjusted as a function of the operating wavelength (0.85 µm or 1.06 µm) and the desired performance. The doping and thickness of the P layer are absolutely critical.

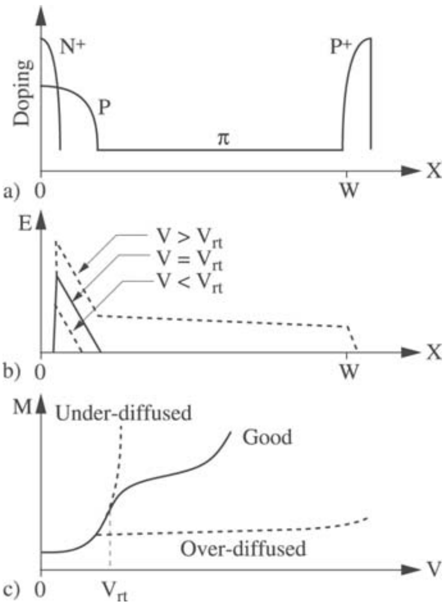


**Figure 3.9.** *Operating principles of a Si P+πPN+ APD: a) doping profile, b) typical electric field, c) typical M(V) characteristics.* $V_{rt}$ *is the reach-through voltage*

The operation of $N^+P\pi P^+$ APD structures is shown in Figure 3.9, which displays the doping profile, the electric field distribution for three significant bias voltages and a typical multiplication curve. At weak bias voltages, the depleted region is restricted to the P layer. As the voltage increases, the edge of the depleted region moves and enters the $\pi$ layer for a bias voltage $V_{rt}$ known as the *reach-through* voltage, for which the multiplication is around 10 to 20, just before the breakdown of the $N^+P$ junction. Then, with the depleted region extending across the whole $\pi$ layer, the increase in the bias voltage will only cause a weak increase $\Delta E$ in the electric field. The multiplication will vary relatively slowly, especially if the $\pi$ layer is wide, and the *M(V)* characteristic involves a more or less pronounced "plateau". Under normal operating conditions, the field in the absorption region is sufficient to ensure carrier transport at their saturation velocity. However, in order to benefit from a high gain, the operating point must be beyond the "plateau".

Consistent achievement of the desired *M(V)* multiplication characteristics assumes that we can accurately control (± 10%) the concentration profiles of the $N^+$ and P layers. Specifically, if the P layer is too wide, the breakdown of the $N^+P$ junction occurs before the depleted region extends in to the $\pi$ layer, but if it is too narrow, the extension into the $\pi$ layer occurs too soon and the multiplication will be weak (Figure 3.9c).

For applications at 0.85 µm, the thickness of the π layer can be adjusted between 15 and 50 µm, as a function of the chosen level of tradeoff between speed and response. In order to avoid the critical stage of localized thinning, these APDs are best fabricated on an epitaxial $\pi/P^+$ plate, with illumination of the $N^+P$ junction only slightly degrading the excess noise factor.

### 3.6.3. *Si $N^+\pi P\pi P^+$ APDs*

An $N^+\pi P\pi P+$ structure with buried P layer is a refined version of the $N^+P\pi P+$ structure. Proposed and studied by Lecrosnier [LEC 75], it offers a significant reduction in the excess noise factor *F(M)* thanks to its narrow multiplication region with a constant electric field limited by the buried P layer: the field is weaker when this is far below the surface. The technique of implanting high-energy $B^+$ ions (1 to 2 MeV), chosen for constructing this P layer, is very well-suited because by modifying the implantation energy and dose parameters it is possible to change the breakdown voltage and excess noise behavior of these APDs. For a given Si $\pi/P^+$ structure, the multiplication electric field decreases as the implantation energy is increases, since the P layer lies further from the surface, at distances of 2.75 and 3.1 µm respectively for implantation energies of 1.5 MeV and 1.8 MeV.

The typical behavior of $N^+P\pi P^+$ APDs is only reached over a small range (± 15%) on either side of the optimal implantation dose. If the dose is too large, the behavior is like that of a $N^+P$ diode, or of a PIN diode if it is too weak. Figure 3.10 shows a schematic doping profile, the associated electric field distribution, and the *M(V)* characteristics for different doses at a given implantation energy.

The fabrication procedure for APDs involves two diffusion phases in order to construct stopper ring and guard ring junctions, followed by two phases of ionic implantation and a phase of electrical activation annealing: firstly the low-energy implantation of phosphor ions to build the active $N^+\pi$ junction close to the surface ($x_J \sim 0.2$ µm), then the high energy implantation of Boron ions (1.5 or 1.8 MeV, dosage $1.2 \times 10^{12}$ ions/cm$^2$) for the buried P layer.

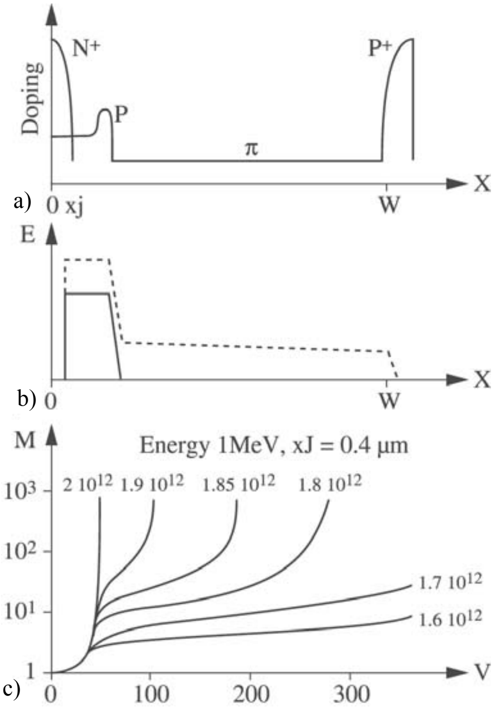


**Figure 3.10.** *Operation principles for a Si N+πPπP+ APD: a) doping profile, b) typical electric field, c) typical M(V) characteristics for different implantation doses of Boron (in cm$^{-2}$)*

Control of the conditions for the two implantations ensures good reproducibility of the procedure, which is particularly easy to do for epitaxial $\pi/P^+$ wafers, and also a good fabrication yield. A model of this structure, based on complete photoelectric characterization (*I(V)*, *C(V)*, *M(V)*, noise) for various APDs was developed by Maille [MAI 80].

APDs obtained in this way offer excellent photoelectric performance, normally with 250 µm diameter diodes:

– weak dark current ($I_0 < 1$ nA) close to breakdown;

– useful gain $M_u > 150$, resulting in a response coefficient greater than 75 A/W at $\pi = 0.85$ µm and a rise time $\tau_r \sim 0.5$ ns for a $\pi$ layer of thickness 30 µm;

– an excess noise factor $F(M) = M^x$ typically with $F(M = 100) = 3 - 3.5$, corresponding to $x = 0.25$. This value, close to the lower limit of $x = 0.2$ is a result of the low ratio $k \sim 0.01$ of the ionization coefficients $\beta_i$ and $\alpha_i$, due to the fact that the multiplication electric field is such that $E_m < 3 \times 10^5$ $\text{Vcm}^{-1}$;

– a uniformity in the photoresponse of ± 5%.

Diodes with an active zone diameter of 800 µm and even 1,500 µm have been achieved thanks to the excellent spatial homogenity offered by the implantation technique. To compensate for the lack of suitable high-energy implantation equipment, a number of solutions have been researched. One of them consists of implanting doubly-ionized $B^{++}$ Boron ions at relatively high energies (~300 keV), or alternatively implanting Boron ions at 800 keV in a preferred direction in a $\pi$/P+ epitaxial plate with <110> orientation, to take advantage of a "channeling" effect, resulting in the P layer lying 2.2 µm below the surface, instead of 1 µm as would be the case under normal conditions. The excess noise factors at $M = 100$ are $F = 4$–5 with "channeling", instead of $F = 6$–7, with $x = 0.30$–0.35 and $x = 0.40$ respectively [KAN 78]. However, although "channeling" implantation gives promising results, these are not easily reproducible.

Another process exists where the buried P layer is fabricated in two stages: first localized low-energy implantation of Boron ions on the $\pi/P^+$ plate, followed by an epitaxial stage to obtain the upper $\pi$ layer for multiplication. The growth of this layer is delicate. APDs made in this way have an excess noise factor $F = 5$ at $M = 100$, with $x = 0.35$ [YAM 76], which could be improved upon by reducing the doping and increasing the thickness of the upper $\pi$ layer. These results confirm that the excess noise factor *F(M)* is weaker when the P layer is buried more deeply, and hence the electric field is only marginally increased.

### 3.6.4. *SiPt-Si N Schottky APDs*

SiPt-Si N Schottky APDs, obtained through the formation of a very thin, transparent and non-absorbing layer (20 to 40 nm) of platinum silicate on top of N-type silicon, with resistivity $\rho \sim 0.5$–1 Ωcm, are well-suited to the detection of ultraviolet (UV) radiation around $\lambda = 0.5$ µm, for which the absorption occurs very close to the surface (absorption length of the order of 0.1 µm). Their performance depends closely on the quality of the SiPt-Si N interface, and hence on the standard of the fabrication technology.

After diffusion of the guard ring, the silica protecting the active central region is chemically removed. Then the residual native oxides and the first few atomic layers of silicon (~10 nm) are removed by ion bombardment during the chemical deposition phase. Then, a 10 to 20 nm platinum film is deposited by radio-frequency sputtering. The platinum silicate is then formed by *in situ* annealing at 600°C, with the thickness of the silicate obtained being twice that of the platinum. The metal present in the passivation layer is eliminated using aqua regia. The upper contact is formed with a ring of aluminum deposited by evaporation. The presence of an anti-reflection coating tuned to $\lambda = 0.5$ µm is needed to obtain a good response coefficient, as the reflection coefficient of silicon is around 45% in the near ultraviolet. Under these conditions, a response coefficient above 0.25 A/W at $M = 1$ can be obtained for $\lambda = 0.4$ µm.

APDs fabricated in this way possess excellent photoelectric characteristics [ALA 74]. A barrier height $\phi_{BN} = 0.85$ V and a good ideality factor $n = 1.02$ measured from the *I-V* characteristics confirms the good quality of the SiPt-Si N interface, despite a dark current greater than that of traditional $N^+P$ or $P^+N$ junctions in their pre-breakdown regime. At short wavelengths, multiplication is triggered by electrons, and values of multiplication $M \sim 300$ have been measured at $\lambda = 0.435$ µm with an excess noise factor $M^x$ such that $x \sim 0.7$. The limit on the bandwidth is imposed by the time constant RC and is several GHz. In contrast, at $\lambda = 0.8$ µm, the light is mostly absorbed outside the depleted region and multiplication is thus triggered by holes leading to performance degradation: a weaker *M*, higher excess noise and a reduced bandwidth.

## 3.7. Avalanche photodiodes based on gallium arsenide

Si APDs are almost perfect photodetectors for the 0.80–0.85 µm transmission window in fiber-optics, simultaneously possessing a weak dark current, a high gain and low noise, all thanks to the maturity of the technology. Its only drawback stems from the response coefficient/speed tradeoff imposed by the weak absorption coefficient, which is in fact not a great problem because the applications are at low or moderate bitrates (<500 Mbit/s).

Nevertheless, investigations have been carried out into gallium arsenide (GaAs), a direct-bandgap semiconductor with a bandgap $E_G = 1.40$ eV and photosensitive up to 0.88 µm, with respect to possible use in integrated circuits and on ternary structures based on GaAs, materials with a high absorption coefficient. GaAs APDs, either of the Schottky type or diffused with an implanted guard ring ($H^+$ protons or $Mg^+$ ions) offer performances below that of Si APDs within the range 0.80–0.85 µm, except in terms of speed. On the other hand, Stilman *et al.* [STI 74] showed the presence of a very pronounced Franz-Keldysh effect for APDs made of lightly-doped material ($n \sim 5 \times 10^{14}$ cm$^{-3}$), when biased in the multiplication regime.

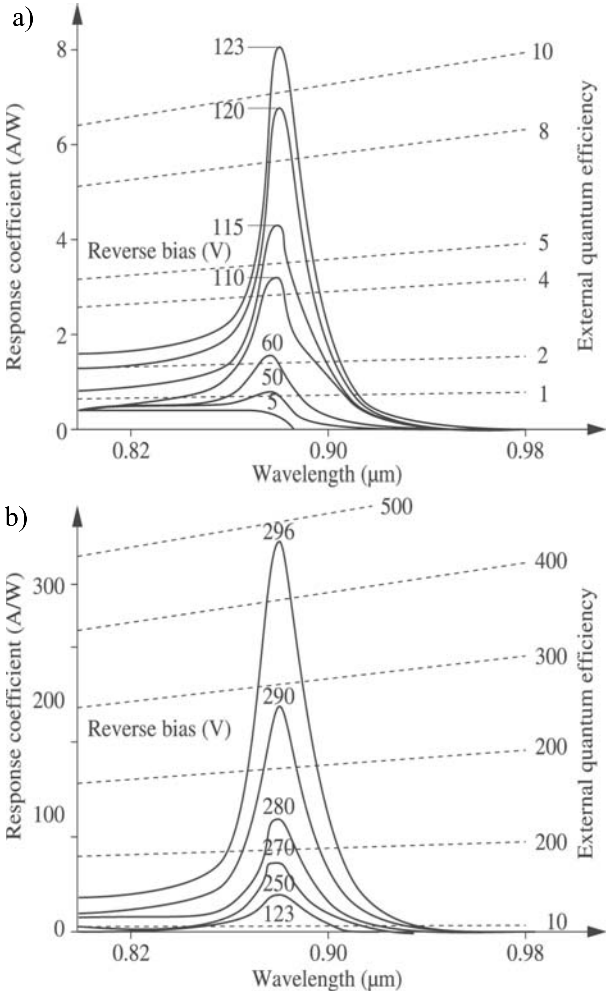


**Figure 3.11.** *Spectral response of electroabsorption GaAs Schottky APDs: a) weak bias voltages, b) strong bias voltages*

More recently, GaAs/Al$_x$Ga$_{1-x}$As APDs with low noise, using a new concept (see section 3.10), have opened a new research avenue. They offer a multi-quantum well (MQW) or "staircase" multiplication region with a bandgap width varying continuously between GaAs and $Al_xGa_{1-x}As$. The need to cross discontinuities favors the ionization of one type of charge carrier. Thus, for the $x = 0.45$ composition of Al, the discontinuities are respectively $\Delta E_C = 0.48$ eV and $\Delta E_V = 0.08$ eV, giving a value of $k = \beta_I/\alpha_I \sim 0.15$ and an excess noise factor $F = 3$ at $M = 10$ [KAG 89]. The presence of an adjacent 1.5 µm absorbing layer of GaAs enables high quantum efficiency devices that are fast and low-noise.

## 3.8. Germanium avalanche photodiodes

Germanium, an indirect-bandgap semiconductor with a small bandgap $E_g = 0.67$ eV at ambient temperature and photosensitive in the spectral range 0.4 µm to 1.65 µm, became established in the middle of the 1970s as the photodetector for the first generation of receivers for fiber-optic links at 1.3 and 1.5 µm, a region where silicon does not work.

Starting in 1966, Melchior and Lynch [MEL 66] showed that Ge APDs could be used as fast and sensitive detectors, and various devices were brought to market. However, competition with Si APDs limited their use to very specific areas of application. The problem was that Ge APDs suffered a number of drawbacks linked to the intrinsic properties of germanium, including:

– a high dark current, dominated by the charge carrier diffusion current, which is proportional to $n_i^2$, $n_i$ being the intrinsic concentration for germanium;

– an insufficiently high ratio $k$ of the ionization coefficients of electrons and holes: $k \approx 2$, with the holes being the more easily ionized, in contrast with silicon;

– the absence of stable native oxides, which makes the process of surface passivation very difficult;

– a difficulty growing high quality epitaxial structures.

On the other hand, due to the high absorption coefficient $\alpha_\lambda$ up to 1.50 µm, depleted region widths $w$ between 2 and 3 µm are sufficient to absorb the light signal, thus allowing a good response coefficient and a high speed. In addition, the dark current does not include a tunneling current contribution, despite the weak value of $E_g$.

### 3.8.1. *Ge APDs with $N^+P$, $N^+NP$ and $P^+N$ structures for 1.3 µm communication*

The three structures $N^+P$, $N^+NP$ and $P^+N$ were developed in succession for use in optical links at 1.3 µm. They have a planar construction with a guard ring to ensure good reliability. Figure 3.12 shows a cross-section view. The presence of a stopper ring decreases the surface current.

$N^+P$ structures were the first to be used, since the diffusion of $n$-type dopants (As, Sb) is easier to achieve. The conditions for the surface preparation and the deposition of dielectric layers ($SiO_2$, $Si_3N_4$), used to limit the diffusion, for passivation, and also for the antireflection coating, are very critical. For an active region of diameter ∅ = 100 µm, the total dark current measured at 0.9 $V_B$ is around 150 nA and the dark current $I_{d0}$ around 50 nA at $M = 1$. The excess noise factor for $M = 10$ is such that $F(10) \approx 10 - 11$, corresponding to an effective ratio between the ionization coefficients of $k_{eff} \approx$ 1.3 µm. This high value of $F(M)$, due to the injection of electrons instead of holes as primary charge carriers, increases still further with the wavelength.

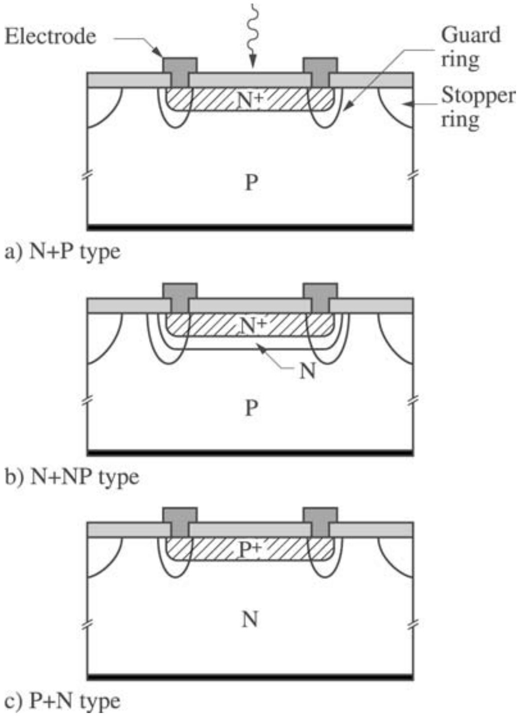


**Figure 3.12.** *Cross-section of Ge APDs: $N^+P$, $N^+NP$ and $P^+N$ structures*

The $N^+NP$ structure was proposed to improve the noise properties of Ge APDs, with multiplication taking place close to the NP junction lying about 2.5 µm from the surface. The N layer is obtained by implantation of arsenic ions followed by annealing, leading to a gradual junction. Since the light is mostly absorbed in the N layer, the holes play the role of primary charge carriers triggering multiplication. In addition, the electric field $E$ at the junction is weaker ($E \approx 2 \times 10^5$ V/cm). This results in a sensitive increase in the excess noise factors: $F(10) = 7$ instead of 10 as previously. On the other hand, the total dark current at $0.9V_B$ is around 1 µm and the response time is limited by the diffusion time in the N layer. The main drawback of the $N^+NP$ layer is the lack of reproducibility caused by the fabrication conditions for the N layer.

The $P^+N$ structure represents the final success of Ga APDs for 1.3 µm applications. It possesses the respective advantages of the $N^+P$ and $N^+NP$ structures in terms of speed and excess noise factor, with the light being absorbed in the N-type depleted region and the multiplication being triggered by holes. The active region and guard ring junctions, as well as the stopper ring, are made by ionic implantation followed by annealing at relatively low temperature. This "fully implanted" technique offers good reproducibility of the performance. For diodes with an active region diameter of 100 µm and 30 µm, the excess noise factor $F(10)$ is 8 and 7 respectively, and the cutoff frequency is at least 1 GHz at $M = 10$.

### **3.8.2.** ***Ge APDs with $P^+NN^-$ structures for 1.55 µm communication***

Above L=1.52 µm the N+P, N+NP and P+N photodiode structures become "blind" if the temperature drops below -20°C, because the absorption coefficient of germanium falls dramatically. The absorption length at ambient temperatures is around 25 µm at $\lambda = 1.55$ µm, a window of weak absorption for optical fibers (0.2 dB/km). Structures developed for 1.3 µm communication are not suitable because of their low depleted region width $w \approx 2$ to 3 µm. Only a structure analogous to $N^+P\pi P^+$ Si APDs, where absorption and multiplication take place in separate regions, could attain the required performance of the response coefficient and speed.

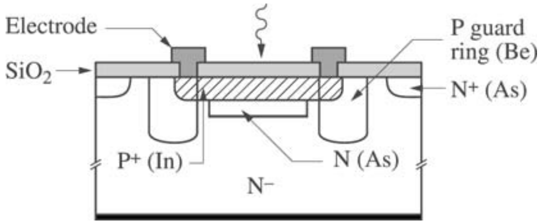


**Figure 3.13.** *Cross-section of a $P^+NN^-$ structure Ge APD for 1.55 µm communication*

$P^+NN^-$ structured Ge APDs on a weak-concentration N substrate ($\approx 5 \times 10^{14}$ cm$^{-3}$) have been built to meet these requirements. The active region of these planar-structure devices with guard rings (Figure 3.13), of diameter Ø = 80 µm or Ø = 30 µm for coupling to multimode or single-mode fibers, is best obtained by a double implantation of boron or indium for the surface $P^+$ junction and of arsenic for the N-type multiplication region, followed by activation and diffusion annealing. The guard ring junction is made either by a long period of zinc diffusion or by implantation of beryllium ions followed by annealing. The best performances obtained for 80 µm diameter APDs at $\lambda = 1.55$ µm are as follows:

– total dark current at 0.9 $V_B$: 1.3 µm;

– dark current $I_{d0}$ at M = 1: 70 nA;

– quantum efficiency $\eta = 80\%$ at $M = 1$;

– excess noise factor $F = 6.1$ at $M = 10$;

– cutoff frequency above 500 MHz at $M = 10$;

– breakdown voltage $V_B \approx 70$–85 V.

The limitation on the response time is imposed by the transit time in the depleted region. The impulse response of Ge $P^+NN^-$ APDs at 1.55 µm does not contain a “lag” due to the collection of charge carriers created outside the depleted region. This is also the case for Ge $P^+N$ photodiodes, whose cutoff frequency is only 10 MHz. Additionally, Ge APDs often operate with an optimal gain between 5 and 10, with the response time depending little on the multiplication factor.

Nevertheless, with the available performance gains for Ge APDs limited by the fundamental properties of germanium, investigations carried out by various laboratories have yielded APDs based on III-V compounds with very promising performances as a second generation of structures for optical communications at 1.3 and 1.55 µm.

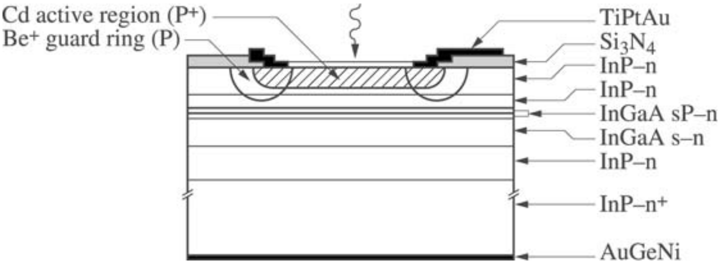


**Figure 3.14.** *Cross-section of an InGaAs/InP planar APD with front illumination, using a SAGM structure*

## 3.9. Avalanche photodiodes based on indium phosphate (InP)

$In_{0.53}Ga_{0.47}As$/InP heterojunction APDs rapidly came to the forefront in the 1980s as second generation detectors for optical communications at 1.3 and 1.55 µm. This was because the ternary compound $In_{0.53}Ga_{0.47}As$, with a direct bandgap of $E_G = 0.75$ eV, offers strong absorption up to 1.65 µm. Formed into a lattice using InP, it is well-suited to epitaxial growth of high crystal quality layers. However, the appearance of a high tunneling current for an electric field $E > 1.5 \times 10^5$ V/cm precludes all avalanche multiplication, despite a more favorable ratio of ionization coefficients than that of germanium ($\alpha_I \sim 2.\beta_I$).

On the other hand, $P^+N$ APDs based on InP, a material with a large bandgap ($E_G = 1.35$ eV) and transparent above 0.95 m, offer a low dark current along with an ionization coefficient for holes that is higher than that of electrons, such that $\beta_I \sim 2\alpha_I$.

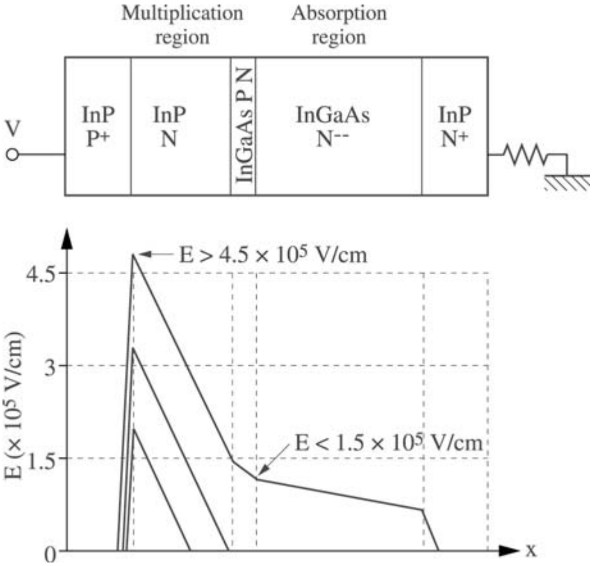


**Figure 3.15.** *Operation principles of an InGaAs/InP APD with a SAGM structure. The electric field is given: 1) for a weak bias voltage, 2) for the punch-through voltage, and 3) in the case of multiplication*

The use of a structure analogous to the $P^+\pi PN^+$ structure developed for silicon, consisting of a separate absorption region ($In_{0.53}Ga_{0.47}As$) and multiplication region (InP), led to the SAGM (*Separated Absorption Graded Multiplication*) structure. Figure 3.14 shows a cross-section through a planar InGaAs/InP APD with front-face illumination which has a SAGM structure. The "graduated" quaternary InGaAsP layers, with bandgap widths intermediate between those of InGaAs and InP, fulfill a double function: to ensure a "gentle" transition between InGaAs and InP, to facilitate the transport of holes by reducing the valence band discontinuity at the InGaAs/InP interface, and to avoid the decomposition of the ternary layer during growth of the InP layers above it. Their presence is vital for the growth of a vertical heterostructure by liquid phase epitaxy. This structure, suited for illumination through front or back faces, offers the advantage of a pure injection of holes, which are the most ionizing carriers in InP, thus allowing a reduction in the excess noise factor. It is suited to illumination through the front or rear faces. Proper operation of these APDs, whose principle is summarized in Figure 3.15, requires strict conditions to be met in terms of doping and thickness of the epitaxial layers, especially of the multiplication layer, since during avalanche the electric field $E_{QT}$ at the quaternary/ternary absorption interface must be below $1.5 \times 10^5$ V/cm in order to retain a low dark current.

Construction of the guard ring, required to reach the necessary reliability levels for detectors targeted at submarine communications, is even more delicate than for Si and Ge APDs. Specifically, the guard ring junction, which is slightly deeper than the active region, must remain within the InP, and the depleted region under the guard ring must reach the ternary layer before that of the active region's junction. This requires that the guard ring junction be gradual with a weak level of doping. For a long time, the implantation of beryllium $Be^+$ ions, followed by redistribution annealing [SHI 83], was the only method of fabricating effective guard rings. More recently, better understanding of epitaxial growth techniques has allowed good results to be obtained by combining localized implantation and epitaxial regrowth [WEB 88]. A realistic predictive model for the behavior of planar InGaAs/InP APDs was put forward by Harari [HAR 91b] after a detailed analysis of their operating conditions.

### 3.9.1. *InGaAs/InP APDs for optical communications at 2.5 Gbit/s*

An APD heterostructure based on InGaAs/InP on a <100> oriented InP-$N^+$ substrate consists of a stack of 6 layers: an InP-$N^+$ buffer layer ($n = 5 \times 10^{17}$ cm$^{-3}$, 0.5 μm), an InGaAs absorption layer ($n = 5 \times 10^{15}$ cm$^{-3}$, 2.5 μm), 2 InGaAsP layers (not intentionally doped, 0.2 μm), the InP-N multiplication layer ($n = 1.5 \times 10^{16}$ cm$^{-3}$), and the InP-$N^-$ covering layer ($n = 8 \times 10^{15}$ cm$^{-3}$, 2 μm).

The guard ring junction is obtained by implantation of beryllium ions ($Be^+$, 150 keV, $5 \times 10^{13}$ cm$^{-2}$) followed by annealing at 700°C in a sealed ampoule. The active junction is formed by diffusion of cadmium at 620°C in a sealed ampoule, for a duration which depends on the required characteristics of the multiplication layer. A nitride layer of $Si_3N_4$ ensures surface passivation, and another acts as an antireflection coating.

Typical performances of such 70 µm diameter APDs are:

– a response coefficient (at $M = 1$) of 0.85 A/W and 1 A/W at 1.3 µm and 1.55 µm respectively;

– a dark current at 0.9 $V_B < 30$ nA;

– a maximum gain $M_{max} = 30$;

– an excess noise factor $F = 5$ at $M = 10$;

– a bandwidth of 2.5 GHz.

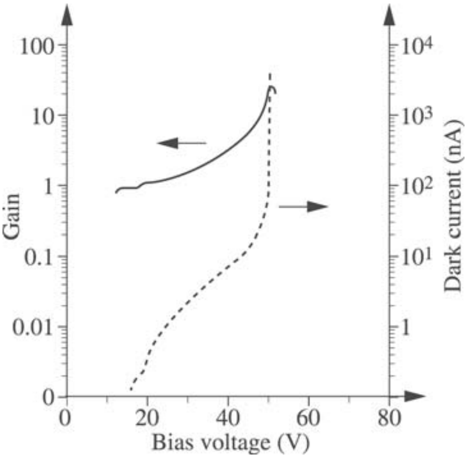


**Figure 3.16.** *M(V) and $I_d(V)$ characteristics of a fast InGaAs/InP APD*

Experiments on accelerated aging under extreme conditions (175°C, 100 µA) have demonstrated the robustness of these APDs, whose median lifetime under these conditions is greater than 4,000 hours. In addition, thanks to their high degree of reliability, they are one of the key devices for fiber-optic communications at 2.5Gbit/s. To begin with, the variation in concentration and thickness of the layers deposited through liquid phase epitaxy had a beneficial effect, with the devices present on a treated wafer offering a wide range of behaviors. This helped develop and understanding of their operating conditions and to acquire a good understanding of the fabrication requirements for such APDs. In this way, fabrication yields greater than 50% were obtained.

More sophisticated growth techniques, such as molecular jet epitaxy or vapor phase epitaxy, allowing the deposition of layers that are very uniform in thickness and concentration have led to even higher fabrication yields (>80%), while tightening the positioning requirements for the junction layout with respect to the InP/InGaAs interface.

### **3.9.2.** ***Fast InGaAs/InP APDs***

In order to meet the need for fast detectors for high-bitrate communications at 5 or 10 Gbit/s, the normal bandwidth of APDs (typically 2.5 GHz) must be increased. In order to do this, we must reduce the "accumulation" of holes due to the large valence band discontinuity at the InP/InGaAs interface, which is the main limitation of the bandwidth of these APDs. It is also necessary to reduce the transit time across the InGaAs absorption layer while keeping its thickness between 1.0 and 1.5 µm in order not to degrade the response coefficient. Furthermore, the increase in doping of the InP multiplication layer, along with a reduction in its thickness, contributes to a reduction in the avalanche trigger time. Finally, an active region of 30 µm diameter, comparable with single-mode fiber illumination, minimizes the capacitance of the junction. APDs with one, two or three thin quaternary layers, or with a quaternary layer whose composition varies continuously between InGaAs and InP, have been fabricated with bandwidths between 7 and 9 GHz. With only one thin 1.18 µm quaternary layer (the wavelength corresponding to the bandgap energy of the GaInAsP compound) ensuring equipartition of the discontinuity of the valence band (0.19 eV instead of 0.38 eV), a theoretical study [RIP 83] shows performances equivalent to two- or three-layer structures which are more complicated to fabricate. After fabrication, the measured performances are as follows:

– dark current < 60 nA at $M = 10$, with the best values being around 10 nA;

– dark current $I_{d0}$ lying between 0.8 and 4 nA;

– maximum gain $M_{\max} > 30$ for the majority of diodes;

– response coefficient at $M = 1$ of 0.9 A/W at 1.3 µm and 0.95 A/W at 1.55 µm;

– bandwidth > 6 GHz at $M = 8$;

– excess noise factor $F = 6$ at $M = 10$, with a fabrication yield greater than 75%.

However, these InGaAs/InP APDs targeted at 10 Gbit/s communications are reaching the limits of performance. System designers often prefer to use InGaAs/InP APDs at 2.5 Gbit/s in a 4 × 2.5 Gbit/s configuration, when an infrastructure for 2.5 Gbit/s optical communications is already in place.

## 3.10. III-V low-noise avalanche photodiodes

The performance of APDs in terms of noise and bandwidth depends strongly on the ratio of the ionization coefficients of the charge carriers. For a given material, they will be higher when the ratio between the ionization coefficients of the least and more easily ionized charge carrier is weaker. This is the case for silicon, where at weak electric fields ($E < 3 \times 10^5$ Vcm$^{-1}$), $k \sim 0.01$, which is close to the behavior of ionization only by electrons. In contrast, in InP where the holes are the more easily ionized, this ratio is around $k \sim 0.5$, a value already corresponding to a significant noise factor.

In order to reduce the excess noise factor and increase the gain-bandwidth product for III-V APDs based on InP or GaAs, it is necessary to strongly increase the ionizability of one type of charge carrier, using a modified structure or a material which has this property under specific conditions. Chin *et al.* [CHI 80] suggested reducing the ionization energy of holes or electrons by taking advantage of the numerous discontinuities in the valence or conduction band which can be found in multi-quantum-well (MQW) structures. Alternatively, Hildebrandt *et al.* [HIL 81] showed that ternary GaAlSb structures offer a significant increase in the ionizability of holes close to a specific composition ratio. This is also the case for the II-VI ternary compound HgCdTe.

### 3.10.1. *III-V super-lattice or MQW APDs*

The multiplication region of a MQW APD consists of a stack of layers with alternately wide bandgaps ("barriers") and narrow bandgaps ("wells") which produce discontinuities at the interface, between conduction bands $\Delta E_c$ and between valence bands $\Delta E_v$, whose values depend on the nature of the materials used. The "barrier/well" pairs favor the transport of electrons when $\Delta E_c > \Delta E_v$ or of holes when $\Delta E_c < \Delta E_v$. For example, in the case of the InAlAs/InGaAs pair, an electron which undergoes a transition in a barrier region crosses the interface and "falls" into the adjacent well, gaining an energy $\Delta E_c$, or alternatively benefits from an artificial reduction in its ionization energy, thus favoring ionization in the wells. Conversely, movement from a well to a barrier reduces the ionizing power of the electron within the barrier. The crossing of numerous barrier-well interfaces in succession translates into an increase in the multiplication triggered by the electrons, equivalent to an increase in $\alpha_I$. On the other hand, the ionizing power of the holes traveling in the opposite direction is less favored, and overall this results in a reduction of the effective ratio of ionization coefficients $k = \beta_I/\alpha_I$, leading to a reduction in the excess noise factor. When $\Delta E_v > \Delta E_c$, such as in InP/InGaAs, it is the ionizing power of holes which is favored, leading to a reduction in the ratio $\alpha_I/\beta_I$ for this structure. Pairs of materials with very asymmetric bandgap discontinuities are favored; thus, the combination of $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As$ ($\Delta E_c$ = 0.55 eV, $\Delta E_v$ = 0.2 eV) is more favorable than $InP/In_{0.53}Ga_{0.47}As$ ($\Delta E_c$ = 0.22 eV, $\Delta E_v$ = 0.38 eV). The number of barrier-well layers and the thickness of those layers both have a significant influence on performance.

Up to now, the best performances have been obtained [WAT 97] for a planar APD with an implanted Ti guard ring with a MQW multiplication layer of InAlGaAs/InAlAs (0.27 µm, 13 repeats) and an absorption region of InGaAsP (1.3 µm, $p = 8 \times 10^{15}$ cm$^{-3}$) separated by a thin doped layer of InP-P$^+$ (0.04 µm, $p = 4 \times 10^{17}$ cm$^{-3}$) in order to reduce the electric field between the two regions. These APDs are characterized by a weak dark current of 0.36 µA at $M = 10$, a bandwidth of 15 GHz, a gain-bandwidth product of 110 GHz, a response coefficient of 0.8 A/W at 1.55 µm, an excess noise factor $F(M) \sim 3$ at $M = 10$ and a high reliability. They are well fitted to the specifications of high-sensitivity 10 Gbit/s receivers.

### 3.10.2. *Spin-orbit resonance APDs*

#### 3.10.2.1. *GaAlSb/GaSb APDs*

The strong increase in the ionizing power of holes which is seen in $Ga_{1-x}Al_xSb/GaSb$ APDs with $x \sim 0.05$, sensitive at 1.3 and 1.55 µm, is linked, according to the interpretation of Hildebrand *et al.* [HIL 81], to the ionization mechanism for holes which causes the valence band to be split in two via spin-orbit coupling. The result is that, around this composition, the structure of the energy bands specific to this material has the unusual property of having a bandgap energy $E_g$ close to the difference in energy $\Delta$ separating the maximum values for the valence band and the additional split band.

If $\Delta E_g \rightarrow 1$, through the resonance effect, the ratio of the ionization coefficients $\beta/\alpha \rightarrow \infty$ and the ionization energy of holes is equal to $E_g$. Subsequent construction of $Ga_{1-x}Al_xSb/GaSb$ APDs with $x \sim 0.05$ has made it possible to obtain devices with a weak excess noise factor $F(M) = 3.8$ at $M = 10$, corresponding to a value of $k = \beta_I/\alpha_I \sim 5$, but much lower than the value of 20 to 30 that has been anticipated, and a gain-bandwidth product of 90 GHz [KAT 90]. These performances are higher for these parameters than those obtained with comparable

InGaAs/InP APDs; however, the inability to reduce and stabilize their dark current or to fabricate a planar APD has led to these investigations being abandoned.

### 3.10.2.2. *HgCdTe avalanche photodiodes*

Research into ternary, mercury-rich alloys $Hg_{1-x}Cd_xTe$, which had already been explored with the aim of fabricating detectors for military applications in the infrared windows at 3.5 µm and 8–14 µm, was extended at the start of the 1980s for the development of alternatives to Ge APD detectors for fiber-optic telecommunications at 1.3 µm and 1.55 µm. The alloys with cadmium-rich compositions ($x \sim 0.7 - 0.6$) corresponding to these two transmission windows offer the same electronic properties of spin-orbit coupling, which offers the opportunity to develop APDs with performances similar to $Ga_{1-x}Al_xSb$.

A method has been developed for growing $Hg_{1-x}Cd_xTe$ wafers with a given stochiometric composition, allowing the construction of p-type substrates with doping levels around $2 \times 10^{16}$ cm$^{-3}$, 40 mm in diameter, with a high homogenity in terms of doping, and a low level of dislocations. The devices are constructed in a planar structure using traditional techniques of insulator deposition by cathode bombardment, and photolithography. The n-type guard ring is obtained by mercury diffusion. The active region junction is made by ionic implantation ($Al^{++}$, 65 keV, $10^{14}$ ions cm$^{-2}$) followed by redistribution annealing. The result is a PIN junction. Compositions with $x$ close to 0.6 are the most suitable, because they cover both transmission windows and also give the highest values of the ratio $\beta_I/\alpha_I$ [ORS 87], with the best being the composition with $x = 0.62$, $E_g = 0.72$ eV and $\beta_I/\alpha_I = 30$. Encouraging initial results obtained in terms of dark current, response coefficient and excess noise factor have not been confirmed at the predevelopment stage.

## 3.11. Prospects

A large proportion of the work in progress is aimed at the development of devices for high-bitrate communications (20 to 40 Gbit/s) that are both fast and low-noise. Particularly significant is the work on Si/InGaAs APDs, "waveguide" MQW APDs and low noise APDs with a very thin multiplication region.

### 3.11.1. *Si/InGaAs APDs*

These APDs, combining the very low excess noise factor of Si APDs with the strong absorption of InGaAs, should simultaneously allow high quantum efficiency and a high bandwidth, a high gain-bandwidth product and a low level of noise. Preliminary results obtained for Si/InGaAs structures fabricated using the technique of *wafer fusion* are very promising: bandwidths of 13 GHz for gains of 135, corresponding to a gain-bandwidth product of 315 GHz have been measured. However, the dark current is raised very close to breakdown and is unstable under a bias voltage [HAW 97].

Important work remains to be done to develop enough understanding of wafer construction and the stages of APD fabrication in order to overcome the limitations linked to the large discrepancy in the lattice parameter between Si and InGaAs on the one hand and, on the other hand, also linked to the difficulties of passivation.

### 3.11.2. *"Waveguide" MQW APDs*

In order to reduce the maximum transit time of charge carriers without reducing the quantum efficiency, the light signal is injected laterally, perpendicular to the electric field, in a vertical "waveguide" structure whose InGaAs core is very thin (~0.5 µm).

Since the MQW InAlAs/InGaAlAs or InAlAs/InGaAs multiplication region is also very thin, the transit time of charge carriers is very short and bandwidths of 25 GHz have been obtained. The gain-bandwidth product is more than 5. For a diode of length 20 µm, the response coefficient at $M = 1$ is 0.9 A/W at 1.55 µm for a fiber-diode coupling factor of 80% [COH 97].

### 3.11.3. *Low-noise APDs with a very thin multiplication region*

The study of APDs with a very thin multiplication region (0.2 to 0.5 µm) has demonstrated the possibility of making devices with a low excess noise factor out of materials such as InP, GaAs, InAlAs whose ionization coefficients are similar, even in the case of a primary injection by the least ionizing charge carriers. Under such conditions, McIntyre's analysis is not relevant because the "dead space", the distance required for a charge carrier to acquire enough energy to ionize, is no longer negligible compared to the thickness of the multiplication region. Consequently, the ionization process is less random, and as a result there is less noise.

Excess noise factors comparable to those measured for multi quantum well APDs ($F \sim 4$ at $M = 10$) have been published for traditional [DRI 99] and waveguide [KIN 00] $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As$ APDs with a high bandwidth (~20 to 30 GHz), whose multiplication region in the high-electric-field monolithic InAlAs material is separated from the low-field InGaAs absorption region by a very thin but heavily doped InP layer. The primary injection of electrons makes the presence of quaternary transition layers unnecessary. This makes such structures very attractive for high-bitrate applications as they are simpler to fabricate.

Additionally, a record gain-bandwidth value of 290 GHz was obtained for InAlAs/InGaAs APDs with a vertical structure incorporating a resonant cavity centered at $\lambda = 1.55$ µm, a high response coefficient (above 0.8 A/W) and a weak dark current for a thickness of 0.2 µm for the InAlAs multiplication region [LEN 99]. However, these APDs require an excellent understanding of epitaxial growth in order to achieve stacking of the “mirror” layers and the active region, and are much more delicate to fabricate than the traditional structure in [DRI 99].

## 3.12. Conclusion

Initially responding to the demands of military applications, the development of avalanche photodiodes has occurred in parallel to that of optical fiber telecommunications, with one not being possible without the other. The result is that for each generation of fibers corresponding to a spectral transmission window of silica, 0.85 µm, 1.3 µm and 1.55 µm respectively, there is a specific APD associated, be it $P^+\pi PN^+$ Si APDs, Ge APDs, and then SAGM InGaAs/InP APDs, tailored to the relevant emitter. The perfection of high-performance APDs has been one of the important factors in the achievement of competitive optical communications. An increase in the sensitivity of the photoreceiver, consisting of the APD combined with an associated amplifier, allows either a reduction in the power of the signal emitted in the first place or a reduction in the number of repeaters along the link.

In addition, the high level of reliability of SAGM InGaAs/InP APDs has contributed to the expansion of submarine optical fiber links over the last decade, which has seen the installation of many intercontinental links at higher and higher bitrates (5 to 10 Gbit/s) with shorter and shorter latencies. Traditional SAGM InGaAs/InP APDs have met, and continue to meet, the needs of designers up to 5 Gbit/s and even 10 Gbit/s. However, the advent of optical links at bitrates of 20 or 40 Gbit/s and beyond leaves no question that there is no current APD structure which is really applicable, and further research is required.

## 3.13. Bibliography

[ALA 74] Alabedra R., Maille C., Lecoy G., Ripoche G., "Photodiode Schottky SiPt-Si N à avalanche", *ESSDERC*, 1974.

[CHI 80] Chin R., Holonyak N., Stillman G., Tang J., Hess K., "Impact ionisation in multilayered heterojunction structures", *Electronics Letters*, vol. 16, no. 12, p. 467, 1980.

[COH 97] Cohen-Jonathan C., Giraudet L., Bonzo A., Praseuth J.P., "Waveguide AlInAs/GaAlInAs avalanche photodiode with a gain-bandwidth product over 160 GHz", *Electronics Letters*, vol. 33, no. 17, p. 1492, 1997.

[DRI 99] Dries J., Thomson K., Forrest S., "$In_{0,53}Ga_{0,47}As/In_{0,52}Al_{0,48}As$ separate absorption, charge, and multiplication layer long wavelength avalanche photodiode", *Electronics Letters*, vol. 35, no. 4, p. 334, 1999.

[EMM 67] Emmons R., "Avalanche photodiode frequency response", *Journal of Applied Physics*, vol. 38, no. 9, p. 3705, 1967.

[HAR 91a] Harari J., Etudes théoriques et expérimentales de photodiodes à avalanche planaires GaInAs-InP, PhD Thesis, Lille Flandre Artois University of Science and Technology, 1991.

[HAR 91b] Harari J., Decoster D., Vilcot J-P., Kramer B., Oguey C., Salsac P., Ripoche G., "Numerical simulation of avalanche photodiodes with guard ring", *IEEE Proceedings*, vol. 138, no. 3, p. 211, 1991.

[HAW 97] Hawkins A., Wu W., Abraham P., Streubel K., Bowers J., "High gain-bandwidth product silicon heterointerface photodetector", *Applied Physics Letters*, vol. 70, no. 3, p. 303, 1997.

[HIL 81] Hildebrand O., Kuebart W., Benz K., Pilkuhn M., "$Ga_{1-x}Al_xSb$ avalanche photodiodes: resonant impact ionization with very high ratio of ionization coefficients", *IEEE Journal of Quantum Electronics*, vol. QE-17, no. 2, p. 284, 1981.

[INT 66] McIntyre R., "Multiplication noise in uniform avalanche diodes", *IEEE Trans. Electron, Devices*, vol. ED-13, no. 1, p. 987, 1966.

[JOH 64] Johnson K., "High speed photodiode signal enhancement at avalanche breakdown voltage", *Digest of Technical Papers, International Solid State Circuits Conference*, vol. 7, p. 64, 1964.

[KAG 89] Kagawa T., Iwamura H., Mikami O., "Dependence of the GaAs/AlGaAs superlattice ionization rate of Al content", *Applied Physics Letters*, vol. 54, no. 1, p. 33, 1989.

[KAN 78] KANEDA T., KAGAWA S., YAMAOKA T., NISHI H., INADA T., "Low noise avalanche photodiodes by channeling of 800 keV boron into <110> silicon", *Journal of Applied Physics*, vol. 49, no. 12, p. 6199, 1978.

[KAT 90] KATSUWAKA H., MIKAWA T., MIURA S., YASUOKA N., TANAHASHI T., WADA O., "An $Al_xGa_{1-x}Sb$ avalanche photodiode with a gain bandwidth product of 90 GHz", *IEEE Photonics Technology Letters*, vol. 2, no. 1, p. 54, 1990.

[KAY 53] MCKAY K., MCAFEE K., "Electron multiplication in silicon and germanium", *Physical Review*, vol. 91, no. 5, p. 1079, 1953.

[KIN 00] KINSEY G., HANSING C., HOLMES A., STREETMAN B., CAMPBELL J., DENTAI A., "Waveguide $In_{0,53}Ga_{0,47}As/In_{0,48}Al_{0,48}As$ avalanche photodiode", *IEEE Photonics Technology Letters*, vol. 12, no. 4, p. 416, 2000.

[LEC 75] LECROSNIER D., PELOUS G., AMOUROUX C., BRILMAN M., RIPOCHE G., "Optimization of avalanche silicon photodiodes: a new structure", *Technical Digest of International Electron Devices Meeting*, Washington D.C., p. 595, 1975.

[LEN 99] LENOX C., NIE H., YUAN P., KINSEY G., HOMLES A., STREETMAN B., CAMPBELL J., "Resonant cavity InGaAs- InAlAs avalanche photodiodes with gain-bandwidth product of 290 GHz", *IEEE Photonics Technology Letters*, vol. 11, no. 9, p. 1162, 1999.

[MAI 80] MAILLE C., Modélisation à l'obscurité et sous éclairement des photodétecteurs à avalanche $N^+\pi P\pi P$ au silicium, PhD Thesis, Languedoc University of Science and Technology, Montpellier, July 1980.

[MEL 66] MELCHIOR H., LYNCH W., "Signal and noise response of high speed germanium avalanche photodiodes", *IEEE Transactions on Electron Devices*, vol. ED-13, no. 12, p. 829, 1966.

[MIL 55] MILLER S., "Avalanche breakdown in germanium", *Physical Review,* vol. 99, no. 4, p. 1234, 1955.

[ORS 87] ORSAL B., ALABEDRA R., VALENZA M., LECOY G., MESLAGE J., BOISROBERT C., "Les photodiodes à avalanche $Hg_{0,4}Cd_{0,6}Te$ à $\lambda = 1{,}55$ µm. Bruit près de la résonance liée au couplage spin-orbite", *Revue de Physique Appliquée*, vol. 22, p. 227, 1987.

[RIP 83] RIPOCHE G., PEYRE J.-L., LAMBERT M., MOTTET S., "High speed ($\geq$ 6 GHz) InGaAs/InP avalanche photodiodes grown by gas source molecular beam epitaxy with a thin quaternary grading layer for high bit rate ($\geq$ 5 Gbit/s) systems", *Journal de Physique*, III, vol. 3, p. 1761, September 1983.

[RUE 66] RUEGG H., "A fast high gain silicon photodiode", *Digest of Technical Papers*, *International Solid State Circuits Conference*, vol. 9, p. 56, 1966.

[SHI 83] SHIRAI T., MIKAWA T., KANEDA T., MIYAUCHI A., "InGaAs avalanche photodiodes for 1 µm wavelength region", *Electronics Letters*, vol. 19, no. 14, p. 534, 1983.

[STI 74] STILLMAN G.,WOLFE C., ROSSI D., DONNELLY J., "Electroabsorption avalanche photodiode", *Applied Physics Letters*, vol. 25, no. 11, p. 671, 1974.

[SZE 66] SZE S., GIBBONS G., "Effect of junction curvature on breakdown voltage in semiconductors", *Solid State Electronics*, vol. 9, p. 831, 1966.

[WAT 97] WATANABE I., NAKATA T., TSUJI M., MAKITA K., TAGUCHI K., "High reliability and low dark current 10 Gb/s planar superlattice avalanche photodiodes", *IEEE Photonics Technology Letters*, vol. 9, no. 12, p. 1619, 1997.

[WEB 88] WEBB P., MCINTYRE R., SCHEIBLING J., HOLUNGA M., "Planar InGaAs/InP avalanche photodiode fabrication using vapor phase epitaxy and silicon implantation techniques", *Technical Digest of Optical Fiber Communications Conference*, OFC 88, 1988.

[YAM 76] YAMAOKA T., MATSUMOTO H., KANEDA T., "Silicon avalanche photodiodes", *Fujitsu Scientific & Technical Journal*, vol. 12, p. 87, September 1976.

# Chapter 4

# Phototransistors

## 4.1. Introduction

The effects of light on transistors have been studied since transistors were first created, with the main motivation for this research being to produce a combination of photodetection and signal amplification in single device. The first work on phototransistors dates from 1951 when Shockley *et al.* [SHO 51] proposed the use of a bipolar n-p-n or p-n-p structure as a phototransistor operating with a base current generated by optical means. The first demonstration of this type of photodetector was achieved two years later when Shive [SHI 53] described a Ge-based n-p-n phototransistor, operating in the spectral region of 1.2 µm wavelength, with an optical gain of the order of 100. In the 1960s, it was mostly bipolar Si phototransistors, operating in the near-infrared, which were developed. However, since the optical absorption coefficient of silicon is weak ($<10^3$ cm$^{-1}$), the foreseeable applications were not those requiring high speeds.

This interest in phototransistors has recently been revived with the development of optical fiber transmission systems and the evolution of the III-V materials technology. This is because, on the one hand,

---

Chapter written by Carmen GONZALEZ and Antoine MARTY.

the best attenuation and absorption properties of optical fibers are found in the near-infrared, with the three minimal absorption windows at 0.85, 1.3 and 1.55 µm corresponding to the emission and absorption ranges of semiconductor materials based on GaAs and InP. On the other hand, III-V materials have absorption coefficients much higher than that of silicon ($>10^4$ cm$^{-1}$). These two characteristics have opened up the possibility of applying phototransistors based on III-V materials to the field of optical telecommunications. Heterojunction bipolar transistors and field effect transistors, made with these materials, began to be studied as phototransistors in the 1970s [ALF 73, BAA 77].

In section 4.2, we present a summary of the different types of phototransistors based on the materials used for their fabrication and as a result their sensitivity to optical wavelengths. We will then classify them according to their structure. In section 4.3 we describe the mechanisms of operation of a bipolar phototransistor and its main properties. Section 4.4 is dedicated to some examples of circuits based around bipolar phototransistors. Finally, in section 4.5, we review the main fields of application of this device.

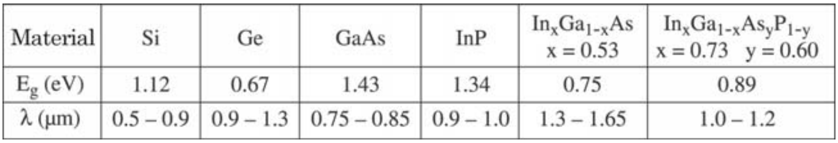
## 4.2. *Phototransistors*

In terms of photodetectors, phototransistors can be classed according to their fabrication material and, just like their transistor analogs, they can also be sorted according to structure into two categories: unipolar field effect devices and bipolar devices. This section reflects this dual classification.

### 4.2.1. *Phototransistors according to their fabrication materials*

The first phototransistors were based on a bipolar homojunction silicon transistor. In this indirect bandgap material, the energy of the bandgap is 1.12 eV. As a result, silicon is photosensitive to wavelengths in the near-infrared, 0.6–0.8 µm. Later, starting in the 1970s, technological progress achieved with III-V compounds and their ternary and quaternary alloys enabled the development of heterojunction devices: field effect transistors and bipolar transistors. These materials have a direct band structure, and some of their ternary compounds have bandgap energies which match the spectral windows of lowest attenuation in optical fibers: 0.85, 1.3 and 1.55 µm. This is why two variants based on III-V materials have been developed: the GaAs variant (for $\lambda = 0.85$ µm) and more recently the InP variant (for $\lambda = 1.3$ and 1.55 µm). Heterojunction phototransistors based on AlGaAs/GaAs are photosensitive to wavelengths between 0.8 and 0.9 µm, with the GaAs bandgap being 1.43 eV. This value of bandgap is compatible with the spectral window of optical fibers around 0.85 µm, for which there is an attenuation of around 2 dB/km. The optical absorption coefficient of GaAs for this wavelength is relatively high: of the order of $10^4$ cm$^{-1}$. The first optical systems, operating initially at 0.85 µm, were later targeted at 1.3 and 1.55 µm where the attenuation in optical fibers is 0.5 and 0.2 dB/km respectively.

Thus, it is the ternary and quaternary alloys based on InP which are best suited for photodetection at these wavelengths. In particular, the ternary alloy $In_xGa_{1-x}As$ with $x = 0.53$, which matches the lattice spacing of InP, has a bandgap energy of 0.75 eV. This value of energy is compatible with photodetection at 1.3 and 1.55 µm. The optical absorption coefficient of InGaAs is $1.16 \times 10^4$ cm$^{-1}$ for $\lambda = 1.3$ µm and $6.8 \times 10^3$ cm$^{-1}$ for $\lambda = 1.55$ µm. The bandgap energy ($E_g$) and the applicable spectral range for the main materials used in phototransistor fabrication are shown in Table 4.1.



| Material | Si | Ge | GaAs | InP | $In_xGa_{1-x}As$<br>$x = 0.53$ | $In_xGa_{1-x}As_yP_{1-y}$<br>$x = 0.73$ $y = 0.60$ |
|---|---|---|---|---|---|---|
| $E_g$ (eV) | 1.12 | 0.67 | 1.43 | 1.34 | 0.75 | 0.89 |
| λ (µm) | 0.5 – 0.9 | 0.9 – 1.3 | 0.75 – 0.85 | 0.9 – 1.0 | 1.3 – 1.65 | 1.0 – 1.2 |

**Table 4.1.** *Bandgap energy and applicable spectral range for various materials used in the construction of phototransistors*

### 4.2.2. *Phototransistors classified by structure*

Homojunction or heterojunction bipolar transistors (photo-HBTs) and unipolar field effect transistors (photo-FETs) are three-terminal devices. When illuminated, the optical input acts as an additional terminal across which the device can be controlled optically. These devices can be integrated into MMICs (*Monolithic Microwave Integrated Circuits*) to achieve optically-controlled amplification or switching, or even more sophisticated microwave functions such as optical locking of oscillators and the mixing of optical and electrical signals [SEE 90].

#### 4.2.2.1. *Unipolar field effect transistors*

Unipolar field effect transistors (FETs), based on III-V materials, only create a single type of charge carrier. Out of the family of field effect transistors, most research has been focused on heterojunction phototransistors which are Schottky gate transistors made with GaAs (MESFETs, *Metal-Semiconductor Field Effect Transistors*) and high electron mobility transistors (HEMT). This popularity is mostly due to the possibilities of using the phototransistor effect in MMICs (Monolithic Microwave Integrated Circuits).

Figure 4.1 shows the cross-section of an illuminated MESFET device (photo-MESFET). Two distinct phenomena occur in the photo-MESFET during illumination by light: a photoconductor effect which is the result of the increase in conductivity due to the creation of photocarriers, and a photovoltaic effect which occurs close to the gate/channel and channel/substrate junctions. Several investigations [GAU 85, MAD 92] have shown not only that the photovoltaic effect dominates the photoresponse of the device, but also that this effect is the cause of the mediocre dynamic performance of photo-MOSFETs. Physically, we can explain this degradation by the poor coupling between the incident light and the gate junction (poor overlap between the optical absorption region and the region with an electric field). A large proportion of the photocarriers end up trapped in low field regions (e.g. the barrier), thus distorting the static characteristics of the device.

Compared to a photo-MESFET without illumination, an important reduction in performance has been observed for the transition frequency $f_T$ and the maximum frequency of oscillation $f_{MAX}$ when illuminated [SIM 86]. The best dynamic performance obtained under illumination is for a bandwidth less than 100 MHz [BAR 97]. Dynamic optical performances are very poor compared to those obtained electrically.

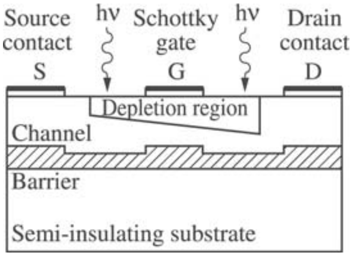


**Figure 4.1.** *Structure of a GaAs MESFET under illumination. The channel is an n-type semiconducting region sandwiched between a semi-insulating substrate and the space charge region (depletion region) of the reverse-biased Schottky junction*

MESFET-based phototransistors have shown themselves to be a poor approach for developing photodetectors in the microwave and millimeter range. Other devices in the FET family, notably HEMT, suffer from the same limitation [ROM 96]

### 4.2.2.2. *Bipolar phototransistors*

An alternative to the photo-MOSFET is the silicon-based homojunction bipolar phototransistor or the heterojunction bipolar phototransistor (based on GaAs or InP). In the bipolar phototransistor, there is an overlap between the region of light absorption and the high-electric-field depletion region. This overlap is as good as in PIN photodiodes. As will be shown in section 4.3, the frequency response to a modulated optical signal is directly related to its purely electrical properties. To make the best use of the performance of a transistor, different illumination approaches have been considered, such as vertical illumination through the front or rear face, lateral illumination, or even the use of a waveguide.

#### 4.2.2.2.1. Traditional surface illumination

The first phototransistors were surface-illuminated, which is the traditional situation, particularly for silicon phototransistors. This vertical illumination allows easy coupling between the optical fiber and the phototransistor, making the integration of the device easier. Several solutions have been proposed: in the case of an opening in the metallic contact of the emitter, the incident light flux crosses the emitter without being absorbed, and the electron-hole pairs are created in the active (or intrinsic) base-collector region of the photo-HBT, that is, the region which is just below the emitter. If a base contact is removed, as shown in Figure 4.2a, the optical flux directly illuminates the base and the collector in the “extrinsic” region of the phototransistor.

#### 4.2.2.2.2. Rear-face illumination

The aim of rear-face illumination of a transistor is the creation of photocarriers in the active region of the transistor (see Figure 4.2b) without modifying the emitter contact. The traditional HBT structure is used with a substrate which does not absorb the incident light. The absorption of light takes place in the collector and the base. Furthermore, the response coefficient is improved because the metallic contact of the emitter acts as a mirror.

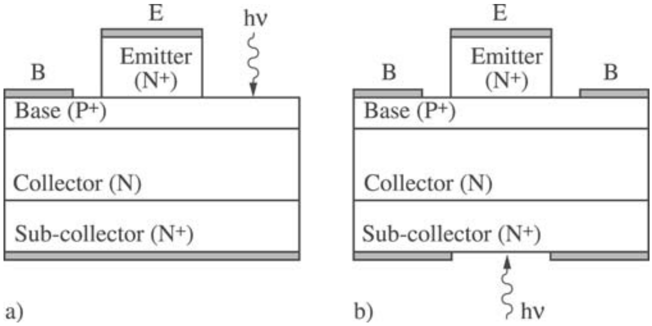


**Figure 4.2.** *Cross-section of a HBT phototransistor with (a) front-face illumination and (b) rear-face illumination*

#### 4.2.2.2.3. Lateral illumination

Lateral illumination (parallel to the layers) offers another approach for the injection of light into a transistor. It is inspired by the technological processes developed for photodiodes. This type of illumination requires a face very vertical to the base-collector island, in order to obtain the best possible injection efficiency. In laterally-illuminated phototransistors, the photons and charge carries no longer propagate in the same direction, as is the case of vertically illuminated photo-HBTs. The aim of lateral illumination is to simultaneously improve the conversion of optical power into electrical power (quantum efficiency) and the speed performance (high operating frequencies) of the device.

#### 4.2.2.2.4. Lateral illumination with an integrated waveguide

However, lateral illumination does pose significant difficulties in terms of achieving a good optical coupling between the fiber and the device. This is why one trick involves the fabrication of a waveguide integrated into the structure of the device, as shown in Figure 4.3. With this approach, the matching of the mode leaving the optical fiber to the propagation mode inside the device is made easier [FRE 96]. In the best cases, an improvement to the injection efficiency is observed which rises from 50% to 90%, with an improved frequency response.

## 4.3. The bipolar phototransistor: description and principles of operation

The bipolar phototransistor is a transistor designed such that the "signal" current which feeds the base terminal is mostly provided by photoelectric effects. This "signal" is then amplified by the transistor effect of the device. First we revisit the principle of operation of a bipolar n-p-n phototransistor, which consists of two distinct p-n junctions, as shown in Figure 4.4. Then, we shall discuss the parameters characterizing the phototransistor: the response coefficient, the static and dynamic gains, the response time, the conversion gain and the noise. A cross-section of an n-p-n type bipolar phototransistor, using mesa technology, is shown in Figure 4.4, along with its bias circuit.

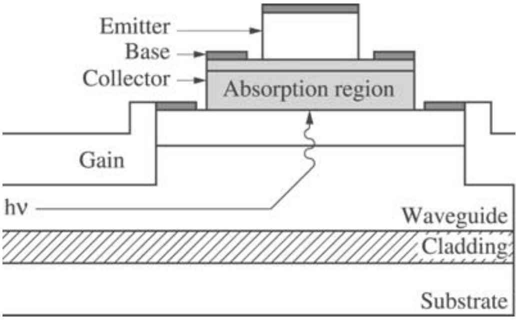


**Figure 4.3.** *Cross-section of an integrated waveguide phototransistor under illumination*

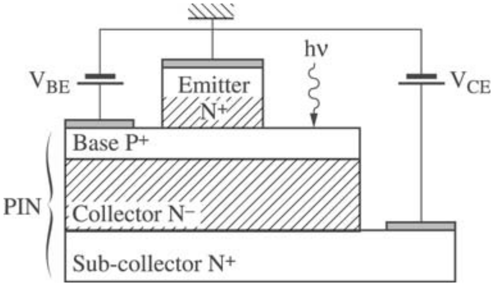


**Figure 4.4.** *Schematic diagram of an n-p-n phototransistor, along with its bias circuit in common-emitter configuration. The shaded regions are free from mobile charges. The base, collector and sub-collector parts can be thought of as a PIN photodiode*

### 4.3.1. *The phototransistor effect*

In Figure 4.4, the n-p-n phototransistor is biased in the common-emitter configuration, with voltages $V_{BE} > 0$ and $V_{CE} > 0$. It differs from traditional bipolar structures, having a relatively large side area of the base-collector junction; this is called the optical window and is what becomes illuminated. This part of the device is effectively a photodiode, connected between the collector and base contacts of the active transistor. When illuminated, electron-hole pairs are created by the photoelectric effect, in the base and in the space charge region (SCR) associated with the base-collector junction (we assume for the sake of simplicity that no absorption takes place in the sub-collector and that the collector is entirely free of mobile charges). A photocurrent $I_{\Phi}$, known as the "primary" photocurrent, is established between the base and collector regions; it is transported by minority electrons from the base which diffuse towards the collector, and by the carriers generated in the SCR, which are separated and moved by the electric field. The flow of this photocurrent $I_{\Phi}$ generates a voltage across the base-emitter and base-collector junctions such that the transistor finds itself in its normal operating system ($V_{BE} > 0$ and $V_{CE} > 0$).

The holes attracted by the base will therefore find themselves blocked by the emitter-base junction. This excess of holes will cause a reduction in the emitter-base potential barrier, which results in an injection of electrons from the emitter into the base, from where the majority will diffuse until they are at the level of the collector. Thus, this is the traditional behavior of a bipolar transistor. The amplification of the photocurrent is a purely electrical phenomenon due to the transistor effect.

#### 4.3.1.1. *The main currents in the phototransistor*

In the normal operating system, the phototransistor can be characterized in terms of the following currents (see Figure 4.5):

– Firstly, the emitter injects electrons into the base region. These, minority carriers in this region, diffuse perpendicular to the junction layout and, if the base is thin enough that recombination can be ignored, they reach the depletion region of the base-collector junction, where the high electric field present in this region clears them out towards the collector region. The flow of these charge carriers gives the contribution $I_{ne}$, which is shown in Figure 4.5.

– Conversely, a current $I_{pe}$ of holes, majority carriers in the base, is injected from the base towards the emitter.

– Generation-recombination phenomena mostly occur at the level of the emitter-base junction, $I_{reb}$, and in the base, $I_{rb}$. $I_{reb}$ comes from the recombination of electrons in the SCR of the emitter-base junction. $I_{rb}$ is caused by the recombination of electrons with holes, majority carriers, in the base.

– Illumination produces the primary photocurrent $I_{\Phi}$. Mostly created in the SCR of the base-collector junction, $I_{\Phi}$ originates from a current of electrons which migrate directly toward the collector contact, and from a current of holes which accumulate at the level of the base.

Thus, balancing the different contributions to the current crossing the two junctions, listed above, allows us to calculate the total currents at the emitter, collector and base; in this way, we can use the continuity of the electron and hole currents across the transition regions to write the following equations. For the base:

$$I_E = I_{ne} + I_{pe} + I_{reb} \qquad [4.1]$$

the emitter:

$$I_B = I_{pe} + I_{rb} + I_{reb} - I_{\Phi} \qquad [4.2]$$

the collector:

$$I_C = I_{ne} - I_{rb} + I_{\Phi} \qquad [4.3]$$

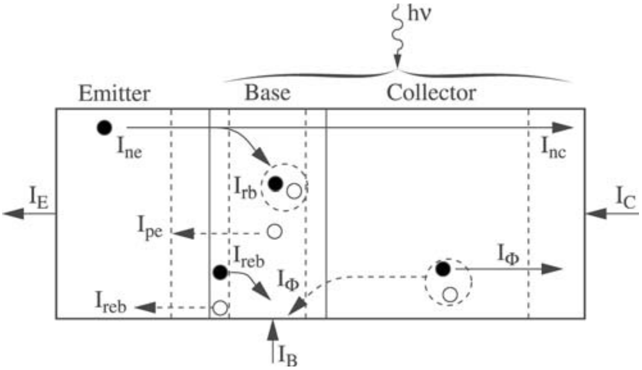


**Figure 4.5.** *Distribution of the different currents of electrical and optical origin in the phototransistor. The SCRs around the emitter-base and base-collector junctions are shown by dotted lines*

Equations [4.1], [4.2] and [4.3] will be used in the following sections to define the optical gain and the electrical gain of the phototransistor.

### 4.3.1.2. *Injection efficiency from the emitter in a homojunction and a heterojunction*

The standard parameters for a bipolar transistor, the injection efficiency from the emitter $\gamma$, and the transport factor in the base $\alpha_T$, can be extended to the phototransistor with the help of equations [4.1], [4.2] and [4.3] where the photocurrent term $I_\Phi$ was introduced. The injection efficiency $\gamma$ relates the ratio of the electron current $I_{ne}$ injected into the base to the total emitter current $I_E$, by the following equation:

$$\gamma = \frac{I_{ne}}{I_{ne} + I_{pe} + I_{reb}} \tag{4.4}$$

Recombination currents reduce the injection efficiency and they must be minimized. If in equation [4.4] we ignore the effect of recombination in the emitter-base SCR, we obtain:

$$\gamma = \frac{1}{1 + \frac{I_{pe}}{I_{ne}}} \qquad [4.5]$$

This ratio must be as close as possible to 1 ($I_{pe} << I_{ne}$) in order to achieve the maximum injection efficiency. Starting from the transport equations established for a bipolar phototransistor, it has been shown that the injection efficiency closely follows the following expression [CAM 85, MOR 72]:

$$\gamma = \frac{A.\gamma_0}{1 + A \cdot \gamma_0} \qquad [4.6]$$

where $A$ is a constant, a function of the thickness of the base and the diffusion length of electrons in the base, and $\gamma_0$ is the Kroemer factor [KRO 57a and b] established for an emitter-base heterojunction. The Kroemer factor is a function of the physical parameters of the materials making up the heterojunction, and can be expressed in the following way:

$$\gamma_0 = \frac{D_{nb} \cdot L_{pe} \cdot n_e}{D_{pe} \cdot L_{nb} \cdot p_b} \cdot \left( \frac{m^*_{nb} \cdot m^*_{pb}}{m^*_{ne} \cdot m^*_{pe}} \right)^{3/2} \cdot \exp\left( \frac{\Delta E_g}{kT} \right) \qquad [4.7]$$

with:

– $\Delta E_g$ the difference in bandgap between the emitter and the base;

– $D_{nb}$ the diffusion coefficient of electrons in the base;

– $D_{pe}$ the diffusion coefficient for holes in the emitter;

– $L_{pe}$ the diffusion length for holes in the emitter;

– $L_{nb}$ the diffusion length for electrons in the base;

– $n_e$ the electron density in the emitter;

– $p_b$ the density of holes in the base;

– $m^*_{nb}$ and $m^*_{pb}$ the effective masses of the electrons and holes in the base;

– $m^*_{ne}$ and $m^*_{pe}$ the effective masses of the electrons and holes in the emitter.

For a homojunction bipolar transistor, there is no variation in the bandgap between the emitter and the base, $\Delta E_g$ is zero, and the exponential factor in equation [4.7] is equal to 1. It is clear from [4.6] and [4.7] that to obtain a high injection efficiency, $\gamma \approx 1$, it is crucial that the emitter should be much more heavily doped than the base. Conversely, for a heterojunction, $\gamma$ depends mostly on $\Delta E_g$, and the term $\exp(\Delta E_g/kT)$ becomes dominant compared to the $n_e/p_b$ ratio; thus, to obtain a $\gamma$ close to 1, it is no longer necessary to under-dope the base relative to the emitter and/or overdope the emitter relative to the base. In heterojunctions based on the materials GaAs/AlGaAs and InGaAs/InP, the size of the emitter bandgap (AlGaAs and InP) is more important than that of the base (GaAs and InGaAs), thus offering a reduction in the injection of majority carriers from the base into the emitter. For the heterojunction phototransistor, the base can be doped at high levels without compromising the efficiency of the junction, leading to a reduction in the resistance of the base. The doping of the emitter can remain within relatively low limits, thus reducing the capacitance of the emitter. These two effects combine to give an improvement in the current gain and an increase in the high frequency performance of the heterojunction bipolar phototransistor.

### 4.3.1.3. *Transport factor in the base*

The transport factor in the base $B$ is defined by the ratio between the electron current gathered by the collector and the electron current injected from the emitted into the base:

$$B = \frac{I_{nc}}{I_{ne}} = \frac{I_{ne} - I_{rb}}{I_{ne}} \qquad [4.8]$$

Due to the recombination current, $B$ is always less than 1. Beginning with equation [4.8], $B$ can be expressed as a function of the transit time in the base $t_B$ and the lifetime of electrons in the base $\tau_n$ using the following equation [CAM 85, CAS 89, POU 94]:

$$B = 1 - \frac{t_B}{\tau_n} \qquad [4.9]$$

According to [4.9], $B$ is closer to 1 when the transit time $t_B$ is small compared to the electron lifetime $\tau_n$. The transit time $t_B$ is smaller when the base thickness is small. As a result, it is necessary for the base thickness to be smaller than the diffusion length of the electrons. For a phototransistor, as for a transistor, the base should therefore be as thin as possible.

### 4.3.2. *The response coefficient of a phototransistor*

The response coefficient of the photodiode base-collector part of the phototransistor is defined as the ration of the primary photocurrent $I_\Phi$ to the received optical power $P_{opt}$:

$$S_0 = \frac{I_\Phi}{P_{opt}} \left(A/W\right) \qquad [4.10]$$

The quantum efficiency, which is the ratio of the number of electrons collected to the number of incident photons, is also used to characterize the optical-electrical conversion of the base-collector photodiode. This efficiency, often called the external quantum efficiency, is expressed in the following way:

$$\eta = \frac{I_\Phi / q}{P_{opt} / h\nu} = \frac{h\nu}{q}.S_0 \qquad [4.11]$$

and the response coefficient of the phototransistor is characterized by the ratio of the component, due to optical excitation $(I_C)_{opt}$ of the current leaving the device $I_C$ to that same incident optical power:

$$S = \frac{\left(I_C\right)_{opt}}{P_{opt}} = \frac{\left(I_C\right)_{opt}}{I_\Phi} \cdot \frac{I_\Phi}{P_{opt}} \left(A/W\right) \qquad [4.12]$$

As we will see in the following section, the ratio $(I_C)_{opt}/I_\Phi$ defines the optical gain of the phototransistor.

Taking into account equations [4.11] and [4.12] we can express $(I_C)_{opt}$ in the following way:

$$\left(I_C\right)_{opt} = \frac{q}{h\nu} \cdot \eta \cdot G_{opt} \cdot P_{opt} = G_{opt} \cdot S_0 \cdot P_{opt} \qquad [4.13]$$

This last equation clearly shows that the phototransistor has an effective response coefficient $G_{opt}$ times greater than that associated with the base-collector photodiode.

## 4.3.3. *Static electrical and optical gains of the phototransistor*

### 4.3.3.1. *Static electrical gains $\beta_0$ and $\alpha_0$*

In the static system, the current gain of the phototransistor is obtained without illumination and, as a consequence, it is defined in the same way as for a bipolar transistor. In the common-emitter configuration, the static electrical gain $\beta_0$ is given by:

$$\beta_0 = \frac{I_C}{I_B} = \frac{I_{ne} - I_{rb}}{I_{pe} + I_{rb}} \qquad [4.14]$$

This gain can also be expressed as a function of the injection efficiency $\gamma$ and of the transport factor in the base $B$ taking into account equations [4.5], [4.8] and [4.14]:

$$\beta_0 = \frac{B \cdot \gamma}{1 - B \cdot \gamma} \qquad [4.15]$$

Because of this dependence on $\gamma$, the current gain in the common-emitter configuration, $\beta_0$, can reach very high values for the heterojunction bipolar transistor. We also define the current gain in common-base configuration $\alpha_0$, which is the ratio of the collector current to the emitter current. Its relationship with the gain $\beta_0$ is:

$$\alpha_0 = \frac{I_C}{I_E} = \frac{\beta_0}{1 + \beta_0} \qquad [4.16]$$

#### 4.3.3.2. *Static optical gain $G_{opt}$*

The static optical gain of a phototransistor is obtained under continuous illumination. It is defined as the ratio between the component linked to the collector current $(I_C)_{opt}$ and the primary photocurrent $I_\Phi$, obtained at the level of the base-collector photodiode:

$$G_{opt} = \frac{\left(I_C\right)_{opt}}{I_\Phi} = \frac{h\nu}{q\eta} \cdot \frac{\left(I_C\right)_{opt}}{P_{opt}} \qquad [4.17]$$

$G_{opt}$ is the equivalent of the electrical gain $\beta_0$ under illumination. It links the gain obtained through amplification, due to the transistor effect, to the primary photocurrent $I_\Phi$. Several authors have shown that, for a bipolar transistor, $G_{opt}$ is proportional to the product of the external quantum efficiency and the electrical current gain [CAM 85, CHA 85]:

$$G_{opt} \cong \eta \cdot \left(1 + \beta_0\right) \qquad [4.18]$$

Equation [4.18] shows that to achieve a given value of $G_{opt}$, it is necessary to establish a compromise between the quantum efficiency $\eta$ and the electrical gain $\beta_0$, depending on the intended application. The result is that the larger the value of $\eta$, the smaller the value of $\beta_0$, and vice versa. In addition, $G_{opt}$ is always smaller than $\beta_0$.

### 4.3.4. *Dynamic characteristics of phototransistors*

In the dynamic system, just as for a normal transistor, the phototransistor effect can be put to good use in the quasi-linear "small-signal" mode, in the non-linear mode to achieve multiplication and mixing behavior and in the "large signal" switching system. In these three cases, the illumination varies with time and we are interested either in (amplitude) modulated light or in an impulse. These three dynamic modes of operation of a phototransistor will be analyzed below.

### 4.3.4.1. *Small-signal operation*

In the context of small-signal applications, the excitation by light may correspond directly to the signal to be amplified, or alternatively it may include a continuous component (modulated signals) which can be put to the particular use of ensuring a pre-bias in addition to that obtained by electrical access to the base. Phototransistors whose base is electrically accessible have the advantage of allowing, by selecting the bias to the base, the choice of an operating point which ensures optimal linearity. The global response time of a phototransistor is no different to the "electrical" response time of the transistor element triggered by the current $I_{\Phi}$ induced in the base-collector photodiode. Two time constants are therefore associated with the dynamic operating system of a phototransistor: the first is due to the intrinsic transport of photocurrent carriers, while the second is due to the electrical response of the phototransistor. The first time constant may be more or less important depending on the geometry of the device; however, to begin with, we can presume that it is the electrical response time which is most significant. The electrical response time of a phototransistor is linked to its transition frequency $f_{T0}$, which is the frequency at which the optical gain is equal to 1. In an equivalent manner to a transistor, this sets the limit for the use of a phototransistor as a photocurrent amplifier. This transition frequency can be calculated in a similar way to the calculation of the equivalent small-signal method for the $\pi$ hybrid configuration.

#### 4.3.4.1.1. Transition frequency calculated using the π hybrid model

The behavior of a phototransistor in the small-signal system is shown in the equivalent circuit diagram of Figure 4.6b. This is based on an electrical model of a bipolar transistor known as the Giacoletto circuit, which is established for the common-emitter arrangement [LET 78]. The incident illumination is modeled by a photocurrent source $i_{\Phi}$ placed between the base B and the collector C. The output signal is taken across the terminals of a load resistance $R_L$ connected to the collector.

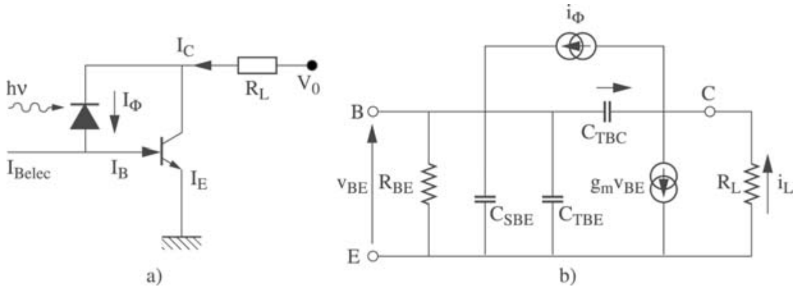


**Figure 4.6.** *(a) Electrical model and (b) its equivalent electrical circuit in the small-signal system of a phototransistor, for a common-emitter configuration*

The elements of the above equivalent circuit can be described as follows:

– $R_{BE}$ is the dynamic input resistance of the phototransistor;

– $R_{BE} = \beta_0 U_T / I_C$;

– $C_{TBE}$ and $C_{TBC}$ are the transition capacitances of the emitter-base and base-collector junctions;

– $C_{SBE}$ is the diffusion capacitance, which represents the effects of accumulation of minority charge carriers in the base, effects linked to the transit time $\tau_B$ in this region, $C_{SBE} = \tau_B . I_C / U_T$;

– $g_m$ defines the transconductance, $g_m = I_C / U_T$.

Under these conditions the input photocurrent $i_\Phi$, called the primary, can be expressed in the following form:

$$i_\Phi = v_{BE} \cdot \left[ \frac{1}{R_{BE}} + j\omega \cdot \left( C_{TBE} + C_{SBE} \right) \right] + \left[ v_{BE} + R_L \cdot i_C \right] \cdot j\omega \cdot C_{TBC} \qquad [4.19]$$

in which the second term (corresponding to the current component $i$ indicated in Figure 4.6) represents the internal reaction mechanism induced by the transition capacitance of the base-collector junction $C_{TBC}$ and by the load resistance $R_L$, commonly referred to as the Miller effect. If we now ignore the current contributions $i$ and $i_\Phi$

compared with the amplified current $g_m v_{BE}$, the output collector current is found to be:

$$i_C \cong g_m \cdot v_{BE} \quad [4.20]$$

Given this, the ratio $i_C/i_\Phi$, which defines the dynamic current gain of the phototransistor $g_{opt}$, known as the optical gain in the small-signal system, can be written in the form:

$$\frac{i_C}{i_\Phi}(j\omega) = g_{opt}(j\omega) = \frac{\beta_0}{1 + j \cdot \omega \cdot \beta_0 \cdot \left(\dfrac{C_{TBE} + C_{TBC}}{g_m} + \tau_B + R_C \cdot C_{TBC}\right)} \quad [4.21]$$

At high frequencies, the imaginary part of the denominator of equation [4.21] becomes dominant and we can write:

$$g_{opt}(j\omega) \cong \frac{1}{j\omega\left(\dfrac{C_{TBE} + C_{TBC}}{g_m} + \tau_B + R_L \cdot C_{TBC}\right)} \quad [4.22]$$

If we assume that the product $R_C.C_{TBC}$ is weak compared to the two first terms, $|g_{opt}|$ is obtained when:

$$\frac{1}{\omega} = \frac{1}{\omega_{TO}} = \frac{C_{TBE} + C_{TBC}}{g_m} + \tau_B \quad [4.23]$$

Equation [4.23] defines the angular frequency of the transition $\omega_{T0}$, with the transition frequency $f_{T0}$ being $\omega_{T0}/2\pi$. Physically, the transition frequency, $f_{T0}$, corresponds to the total transit time, $\tau_T$, of photocarriers from the emitter to the collector. It is interesting to analyze this time constant $\tau_T$, associated with $f_{T0}$. It is defined as:

$$\tau_T = \frac{1}{\omega_{TO}} = \frac{1}{2 \cdot \pi \cdot f_{TO}} \qquad [4.24]$$

If we substitute equation [4.23] into [4.24] we find:

$$\tau_T = \frac{C_{TBE}}{g_m} + \frac{C_{TBC}}{g_m} + \tau_B \qquad [4.25]$$

The first two terms of equation [4.24] correspond to the charging times of the base-emitter and base-collector junctions respectively. Equation [4.25] shows that $\tau_T$ depends on the collector current $I_C$ (via $g_m$) up to a certain threshold and, if we take into account equation [4.13], $\tau_T$ is inversely proportional to the incident optical power $P_{opt}$. For high values of $P_{opt}$, $\tau_T$ approaches the limiting value $\tau_B$. On the other hand, at low values of $P_{opt}$, the terms in $C_{TBE}$ and $C_{TBC}$ become dominant and the transition frequency $f_{T0}$ is proportional to $P_{opt}$.

This behavior of $f_{T0}$ as a function of the incident optical power has been experimentally observed, as shown in Figure 4.7. This dependence makes the use of relatively high optical powers (>10 µW) necessary. Alternatively, we can use an electrical current via the metallic base contact to pre-bias the structure, thus reducing the "threshold" optical power needed to reach the maximum value of $f_{T0}$.

With the help of equation [4.22] we can also define the cutoff angular frequency $\omega_\beta$ as being the angular frequency at which $|g_{opt}|$ falls by 3 dB. This is equal to:

$$\frac{1}{\omega_B} = \beta_0 \left( \frac{C_{TBE} + C_{TBC}}{g_m} + \tau_B + R_L \cdot C_{TBC} \right) \qquad [4.26]$$

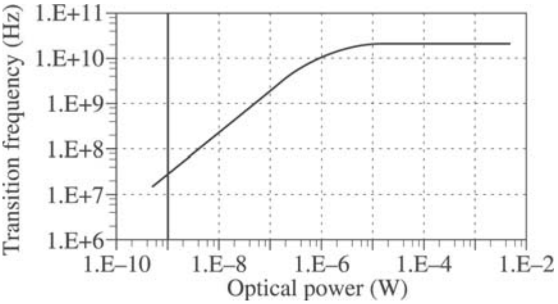


**Figure 4.7.** *Variation in transition frequency $f_{TO}$ with incident optical power $P_{opt}$. At weak $P_{opt}$, $f_{TO}$ varies rapidly up to a certain limiting value of $P_{opt}$, after which it remains constant*

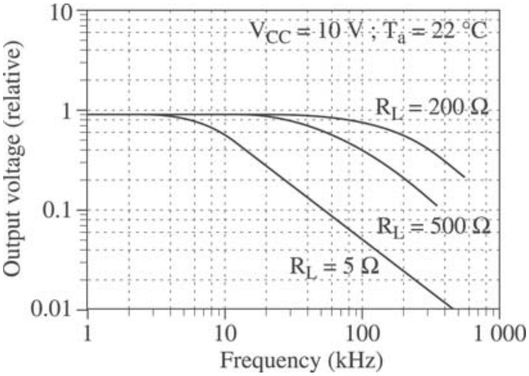


**Figure 4.8.** *Frequency response of a phototransistor for different load resistances*

It can be clearly seen from equation [4.26] that the cutoff angular frequency $\omega_\beta$ depends on the operating point (as does $\omega_{TO}$) and on the value of the load resistance $R_C$, as shown in Figure 4.8.

### 4.3.4.2. *Nonlinear operation*

A bipolar phototransistor behaves similarly to a bipolar transistor, and as a result possesses nonlinear characteristics which allow it to act as an optical-electrical mixer. It can mix a modulated optical signal, carrying information, with an electrical signal from a local oscillator.

This mixing behavior has been the subject of recent research [BET 98, GON 98, SUE 96] and has been developed for telecommunication applications. Two mixing configurations are possible:

– transposition of a low frequency input signal into a higher frequency signal. This mode of operation will be referred to below as up-conversion;

– transposition of a high frequency input signal into a lower frequency signal. This mode will be referred to as down-conversion;

#### 4.3.4.2.1. Mixing principles

Any device able to transpose an input signal from a frequency $f_E$ into another higher or lower frequency is known as a mixer. This change in frequency originates in the nonlinear properties of the mixer and can be explained in the following manner: consider a circuit element whose $I/V$ current-voltage characteristics are nonlinear (a nonlinear resistance, a Schottky diode, etc.). If it is subjected to a voltage $V_{OL} = v_{OL}\sin(\omega_{OL}t)$, its $I/V$ nonlinearity can be described by expanding the current flowing through the element in terms of a discrete series that is a function of the voltage $V$:

$$I = I(V) = I_0 + a_1 V + a_2 V^2 + a_3 V^3 + \ldots \qquad [4.27]$$

where $I_0$ is a continuous bias current and $a_1$, $a_2$ and $a_3$ are real constant coefficients. [4.27] shows that we see an infinite number of powers of $V$ appearing at the output of the circuit, these will enrich the spectrum of the input signal $V$. If we add a second signal to the input, $V_{FI} = v_{FI}\sin(\omega_{FT}t)$, the spectrum of the output signal becomes even more complicated due to the presence of the products of $V_{OL}$ and $V_{IF}$ and their respective harmonics.

If we now replace $V$ with $V_{OL}$ + $V_{FI}$ in [4.27], the current (V) becomes:

$$
\begin{aligned}
I &= I_0 + a_1\left(V_{OL} + V_{FI}\right) + a_2\left(V_{OL} + V_{FI}\right)^2 + \ldots \\
&= I_0 + a_1\left(v_{OL}\sin\omega_{OL}t + v_{FI}\sin\omega_{FI}t\right) + a_2\left(v_{OL}\sin\omega_{OL}t + v_{FI}\sin\omega_{FI}t\right)^2 + \ldots \\
&= I_0 + a_1\left(v_{OL}\sin\omega_{OL}t + v_{FI}\sin\omega_{FI}t\right) \qquad [4.28] \\
&+ a_2\left\{\frac{1}{2}v_{OL}^2\left(1-\cos 2\omega_{OL}t\right) + v_{OL}v_{FI}\left[\cos\left(\omega_{OL}-\omega_{FI}\right)t - \cos\left(\omega_{OL}+\omega_{FI}\right)t\right]\right. \\
&\left. + \frac{1}{2}v_{FI}^2\left(1-\cos 2\omega_{FI}t\right)\right\} + \ldots
\end{aligned}
$$

The interesting frequencies are ($\omega_{OL}$ + $\omega_{FI}$) and ($\omega_{OL}$ – $\omega_{FI}$) and they can be extracted by a filter tuned to the target frequency. In a bipolar mixing transistor, the main nonlinearity contributing to the mixing behavior is transconductance. For a bipolar phototransistor, the nonlinearity in the current gain is the most important parameter.

#### 4.3.4.2.2. Performance criteria

As a first approximation, we can say that the performance criteria established for a mixing transistor can be applied to a mixing phototransistor. We will discuss two of these criteria: conversion gain and insulation.

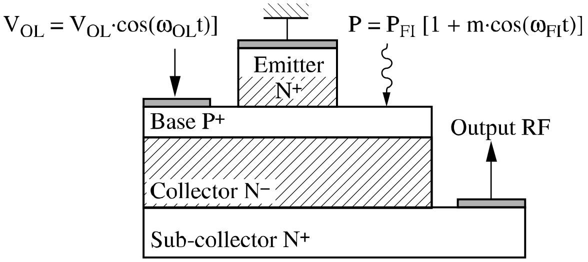


**Figure 4.9.** *Optical and electric inputs of the phototransistor (in place of the mixing phototransistor)*

*Conversion gain*

In a mixing phototransistor, the optical input is physically separated from the electrical input (see Figure 4.9). The optical signal "enters" through the optical window and the electrical signal is applied to the metallic contact of the base. In the common-emitter configuration, the mixed signal is obtained at the output of the collector. The traditional definition of the conversion gain $G_{con}$, applied to a phototransistor, can be expressed as the ratio between the output electrical power at the "mixed" frequency and the effective available input optical power at the modulation frequency of the light. Taking the case of up-conversion, the light is modulated at an intermediate frequency $F_{IF}$, the frequency of the local oscillator $F_{OL}$ is much higher than $F_{IF}$, and the mixed frequencies $F_{RF}$ are:

$$F_{RF} = F_{OL} \pm F_{FI}$$

The conversion gain can be written as:

$$G_{con} = \frac{P_{elec}\left(F_{RF}\right)}{P_{opt}\left(F_{FI}\right)} \qquad [4.29]$$

In equation [4.29], the denominator refers to an optical power. In order to determine the electrical power generated by the incident optical power, we must take into account the primary photocurrent $I_{\Phi}$ generated by the illumination at the level of the base-collector junction, and of the input impedance associated with this junction, $Z_{BC}$. The electrical power can thus be expressed by the equation:

$$P_{elec} = \frac{1}{2}\cdot \mathrm{Real}\left(Z_{BC}\right)\cdot\left|I_{\Phi}\right|^2 \qquad [4.30]$$

and the electrical conversion gain can be expressed as:

$$G_{con} = \frac{P_{RF}}{\left(P_{elec}\right)_{FI}} = \frac{2P_{RF}}{\mathrm{Reel}\left(Z_{BC}\right)\cdot\left|I_{\Phi}\right|^2} \qquad [4.31]$$

Another definition of the conversion gain is possible. It is defined as being the ratio between the power of the mixed signal $F_{RF}$ and the “primary” power of the incident signal $F_{IF}$, in other words, the power of $F_{IF}$ obtained when the base-emitter junction of the phototransistor is short-circuited (base-emitter voltage $V_{BE} = 0$ V).

$$G_{con} = \frac{P_{RF}}{\left(P_{FI}\right)_{V_{BE}\ 0}} \qquad [4.32]$$

When describing experimental results this second definition of $G_{con}$ is the most widely used.

*Insulation*

All the frequencies generated by the mixing are present at each of the terminals of the mixer, and this has a significantly detrimental affect on its conversion gain performance. As Figure 4.9 shows, the phototransistor has three points of access, two electrical, the base contact (*OL*) and the collector contact (*RF*), and one optical, the optical window (*IF*). The insulation is measured between two of the mixer’s points access, *IF-OL* and *IF-RF*, by the ratio between the power at the frequency $F_{OL}$ (or $F_{RF}$) present at the access point *OL* (or *RF*) and the power at the frequency $F_{OL}$ (or $\mathrm{F_{RF}}$) present at the IF access point. Let us consider the *OL* and *IF* access points at which the $F_{OL}$ and $F_{IF}$ signals are injected. If we hypothesize that no electrical signal originating from the access point of the base can be converted into an optical signal leaving through the optical window, the *IF-OL* insulation is infinite. The same reasoning can be applied to the *IF-RF* access point. As a result, we obtain a maximum insulation between these two pairs of terminals without needing to use external circuitry. However, for insulations in the opposite direction, *OL-IF* and *RF-IF*, there can be a loss of $F_{IF}$ signal power in each of these two electrical access points.

### 4.3.4.3. *Operation in the strong-signal system*

As was mentioned previously, the other application of a phototransistor concerns the detection of radiation in the strong-signal system. In this case, the presence or not of the illumination induces a switching type of behavior: the device initially blocks a signal, or is in the weak conduction system if it is pre-biased. It then enters the normal, or even saturated, system when light is applied; it will then return to the initial mode of operation when the input signal ceases. These system changes give rise to transitory phases, as shown in Figure 4.10, which defines a response time for each: the lag time $t_\tau$ is inherent to the time it takes the transistor to start conducting (in the case of operation without pre-bias) and is linked to the charging of the transition capacitances $C_{TBE}$ and $C_{TBC}$. By considering the mean values of these parameters, this time can be calculated using the expression:

$$t_r = \frac{0.7 \cdot \left(C_{TBE} + C_{TBC}\right)}{I_\Phi} \qquad [4.33]$$

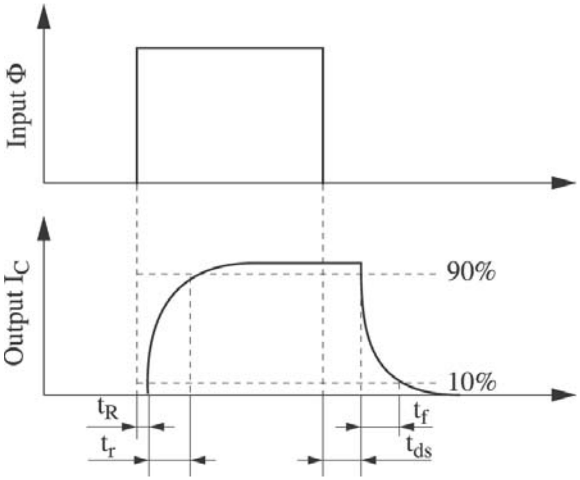


**Figure 4.10.** *Representation of the different switching response times of phototransistors*

The rise time $t_r$ in the normal active system (up to saturation, if applicable) is more difficult to evaluate, bearing in mind the variation in numerous electrical parameters that apply in this transitory system; nevertheless, the dynamic behavior of the phototransistor can be approached in a simple manner, using the small-signal model. If we assume that the current gain and its cutoff frequency are independent of the level of injection; the result is that the first-order model, which relates $I_C$ to $I_\Phi$ in the context of this assumption, suggests that the collector current response to a step change to the primary photocurrent in the base follows an exponential law with time constant $1/\omega_\beta$. Thus, the rise time which separates the start of the growth in the collector current (conventionally, $I_C$ equal to 10% of its final value) from the end of this first transient phase ($I_C$ equal to 90% of its steady-state value), being either $\beta_0.I_\Phi$ or $I_{Csat}$, can be calculated as:

$$t_m = \frac{2.9}{\omega_\beta} \qquad [4.34]$$

when saturation is not reached and:

$$t_m = \frac{0.8}{\omega_T} \cdot \frac{I_{C_{sat}}}{I_\Phi} \qquad [4.35]$$

in the alternative case.

For opposite switching, the desaturation time corresponds to the removal of the accumulated surplus charge in the form of the minority carriers (electrons) in the vicinity of the base, and is therefore defined by the lifetime of these charge carriers.

Over the course of the decaying phase of the collector current, the transistor is once more in the normal active system, up to the point it stops conducting. The fall time $t_f$ can again be expressed by equations [4.34] and [4.35].

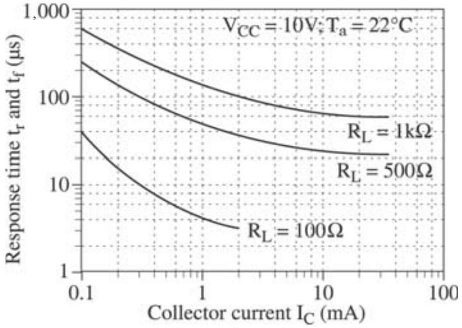


**Figure 4.11.** *Response time of a phototransistor as a function of the collector current and for different values of the load resistance*

Here once more, in the switching system, the rise time $t_r$ and the fall time $t_f$ increase with the load resistance $R_L$. We also observe that, while it leads to an increase in response time, the presence of saturation leads to a reduction in these rise and fall times. Finally, the addition of a pre-bias also contributes to an improvement in $t_r$ and $t_f$, as can be seen in Figure 4.11. Overall, depending on the type of phototransistor and the circuit used, along with its load resistance and operating point, these response times vary from a few tenths to tens of microseconds.

## 4.3.5. *Noise in phototransistors*

The phototransistor is the first element in a photoreceiver system. This means that there is a certain threshold power below which the photoreceiver cannot detect a signal: a power which is determined by the power of the noise in the phototransistor. The minimum detectable power in the photoreceiver system is limited by the noise in the phototransistor and its load circuit. The noise in the phototransistor is hence an important criterion for judging the performance of the whole photodetection apparatus. The most commonly-used approach for characterizing the noise of a photodetector is the noise equivalent current generator, applied at the input [CAM 82, SMI 80, WAN 86]. This is the approach that we will develop over the rest of this section.

Four main noise sources can be associated with a HPT:

– the shot noise due to the base current $I_B$ entering the phototransistor, this current is the sum of the photocurrent $I_\Phi$ and the electrical bias current $I_{Belec}$;

– the shot noise due to the collector current $I_C$, at the output of the HPT;

– the thermal noise due to the load resistance $R_C$;

– the thermal noise generated in the base region and, in contrast to the other quasi-neutral regions, the base, although heavily doped, presents an effective base resistance $\mathrm{R_B}$ which can reach values as high as hundreds of Ohms.

The spectral densities of the noise power, $\overline{i^2}/\Delta f$ in $A^2/Hz$, associated with each source of noise are, for the base, the following:

$$\frac{\overset{-2}{i}_{base}}{\Delta f} = \underline{2}\cdot\left(2q\cdot\left(I_\Phi + I_{Belec}\right)\right) + \frac{4kT}{R_{BE}} \qquad [4.36]$$

The factor of $\underline{2}$ in equation [4.36] is due to the correlation which exists between the base-emitter and base-collector junctions [MON 71]. Physically, this arises from the fact that each fluctuation in the base current $\Delta I_B$ simultaneously produces another equivalent fluctuation in the emitter current to compensate for $\Delta I_B$. These two currents, in the opposite direction, cancel through recombination, but their shot noises are independent and add up. For the collector:

$$\left.\frac{\overset{-2}{i}_{C}}{\Delta f}\right|_{output} + \left.\frac{\overset{-2}{i}_{th}}{\Delta f}\right|_{output} = 2qI_C + \frac{4kT}{R_L} \qquad [4.37]$$

These two spectral densities associated with the collector at the output of the transistor are linked to the optical input of the phototransistor in the following manner [THU 99]:

$$\left.\frac{\overset{-2}{i}_{C}}{\Delta f}\right|_{input} + \left.\frac{\overset{-2}{i}_{th}}{\Delta f}\right|_{input} = \frac{1}{\left|g_{opt}\right|^2}\cdot\left(2qI_C + \frac{4kT}{R_L}\right) \qquad [4.38]$$

where $g_{opt}$ is the dynamic optical gain of the phototransistor, defined in equation [4.21].

Some degree of correlation can exist between the sources of shot noise in the base and collector. Nevertheless, for frequencies above the noise cutoff frequency 1/*f*, the correlation between these two sources can be assumed to be negligible [ESC 95]. As a result, the total spectral density of noise power of the phototransistor, in terms of its optical input, can be expressed as the sum of all the spectral densities, equations [4.36] and [4.38]:

$$\frac{\overset{-2}{i_{Tot}}}{\Delta f} = 4 \cdot q \cdot \left(I_{\Phi} + I_{Belec}\right) + \frac{4kT}{R_B} + \frac{1}{\left|g_{opt}\right|^2} \cdot \left(2 \cdot q \cdot I_C + \frac{4kT}{R_L}\right)\left(\frac{A^2}{Hz}\right) \qquad [4.39]$$

According to equation [4.39], the total spectral noise density of the phototransistor is dominated at low frequencies by the sources of noise linked to the photocurrent $I_{\Phi}$, to the bias current of the phase $I_{Belec}$ and to the base resistance $R_{BE}$; at high frequencies, the noise source due to the collector current becomes dominant.

## 4.4. Photodetector circuits based on phototransistors

### 4.4.1. *Amplification circuits*

Phototransistors can be used either in the linear detection system (continuously modulated signals or pulsed signals) or in the switching system, in circuits with (see Figure 4.12a) or without (see Figure 4.12b) pre-bias, with of course the possibility of adding an additional level of amplification (see Figure 4.12c). It has been shown that in every application – detection of weak or strong signals – the speed is limited by the value of the load resistance of the phototransistor. In order to improve sensitivity, it is common nowadays to use either an active load with low input impedance (a common-base transistor) (see Figure 4.12d) or an operational amplifier providing the current-voltage conversion (see Figure 4.12e). Finally, we should note that the phototransistor can be integrated with a second transistor to form a photo-Darlington pair (see Figure 4.12f). This structure clearly gives a greater response coefficient at the cost of deterioration in the operation speed.

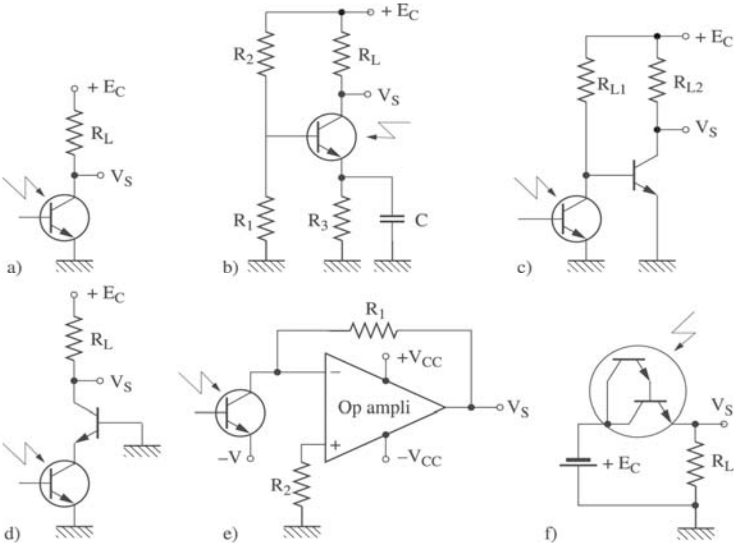


**Figure 4.12.** *Main circuit diagrams for phototransistors: a) structure with two electrodes E and C, b) pre-bias using access to the base, c) supplementary amplification with a second stage, d) improvement in speed using a low impedance load (common-base transistor) or e) by current-voltage conversion, f) integration on silicon of a second stage: the photo-Darlington pair. Note: structures c) and d) are used in fast circuits*

Structures c) and d) have been used to make monolithic amplification circuits integrating the phototransistor and several HBTs, for wideband [CHA 92, KAM 00, WAN 86a] and narrowband [GON 00a, KAM 95] applications.

### 4.4.2. *Nonlinear circuits*

Phototransistors can also be used in optoelectric mixing and auto-oscillation configurations. The simplest circuit is shown in Figure

4.13a and consists of a phototransistor and two matching cells. The first cell is placed between the source *OL* and the input to the base of the phototransistor. The second cell is placed between the output of the collector and the load resistance $R_L$. The optical signal is modulated at the intermediate frequency IF, and the mixed signals at frequencies *OL* ± *IF* are recovered at the output of the collector. The second circuit, described by Sawada *et al.* [SAW 99] and depicted in Figure 4.13b, consists of a phototransistor and a circuit associated with its base which allows it to auto-oscillate at the *OL* frequency of 9.567 GHz. This frequency is stabilized using a dielectric resonator *DR*. The optical signal is modulated at the frequency of 0.2 GHz, and the mixed signals recovered with up-conversion are at frequencies of 9.367 and 9.767 GHz. The third circuit, described by Lasri *et al.* [LAS 00], is shown in Figure 4.13c. It consists of two phototransistors with their emitters joined together. One phototransistor is used as an auto-oscillator at 30 GHz locked optically to an optical signal $\lambda_1$ modulated at the same frequency *OL*. This signal *OL* is coupled with the second phototransistor, which acts as an optoelectrical mixer between the 30 GHz signal and a second optical signal, $\lambda_2$, modulated at 300 MHz. The two optical signals $\lambda_1$ and $\lambda_2$ could be at wavelengths corresponding to a WDM (wavelength division multiplexing) scheme. These circuits demonstrate the abilities of phototransistors to combine the functions of photodetection, amplification, mixing and auto-oscillation.

## 4.5. Applications

### 4.5.1. *Galvanic isolation*

Silicon phototransistors have been used for a long time in optical isolators. These devices may or may not have a base contact. Figure 4.14 shows an optical isolator based on a phototransistor. For this application, all that is required is to connect a structure upstream which takes care of the conversion of an input electrical signal into a light beam. This is in order to achieve a transfer of information which avoids any form of electrical connection. In this way, galvanic isolation between the controller and the receiver is ensured, and the achievable voltage difference, which in the static system is determined by the distances which separate the two components or the connections to the different electrodes, can reach several kilovolts.

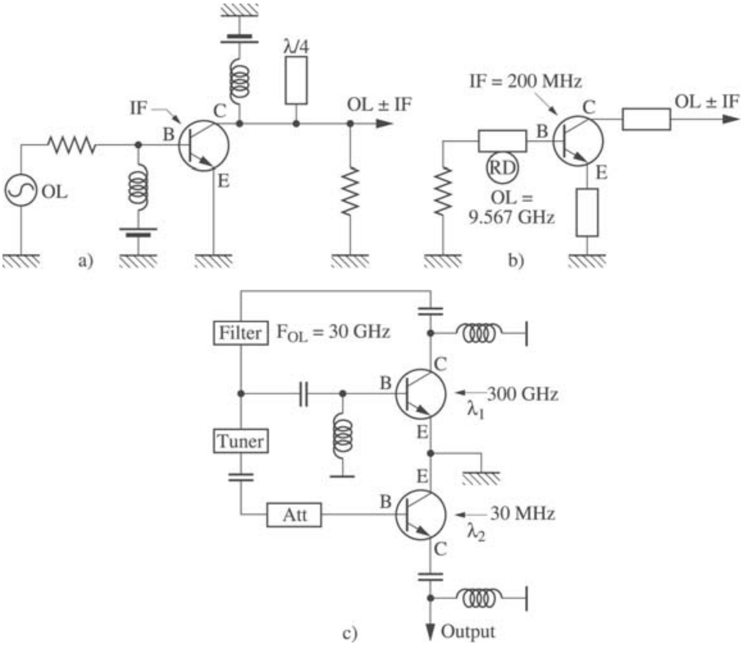


**Figure 4.13.** *Various circuit diagrams for phototransistors acting as (a) a simple optoelectrical mixer, (b) auto-oscillation and mixing with only one phototransistor and (c) auto-oscillation and mixing with two phototransistors*

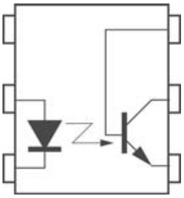


**Figure 4.14.** *Schematic representation of an optical isolator based on a phototransistor*

The device providing the electron-photon conversion at the input is normally a light-emitting diode (LED) based on GaAs, whose emission line of around 850 nm is well suited to optical detection by the phototransistor or the photo-Darlington pair. The current conversion gain can reach 500%. An example of its variation as a function of the input current is shown in Figure 4.15.

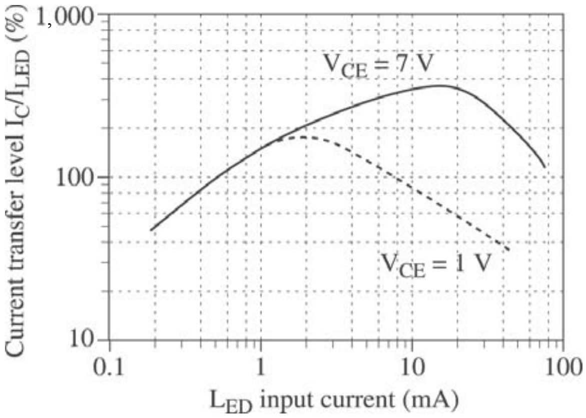


**Figure 4.15.** *Current transfer level (current from the collector of the phototransistor/current in the input LED) as a function of the input current to the optocoupler*

Another application of silicon phototransistors involves power electronics. These types of devices must often meet strict requirements of galvanic isolation, particularly for control applications and energy conversion from the industrial electricity grid. Control through optical means allows this specific problem to be solved in an efficient and elegant manner: with the thyristor and the triac – which are the two basic bipolar components in this application domain – constructed around a photodiode, whose response replaces the current from the control electrode, the gate. A schematic cross-section for a photothyristor and the equivalent circuit for a phototriac are shown in Figure 4.16.

Photothyristors present very interesting possibilities for direct current power transmission at high voltages (1,000-1,500 A and 5-8 kV) [ARN 92].

The phototriac is constructed by combining two photothyristors in parallel, in opposing directions. The near-obligation, for monolithic integration of this device onto a silicon substrate, of using only one surface of the substrate minimizes the current capacities and reduces the voltage stability. Because of this, these devices are applied in the domain of static relays which involve currents and voltages not exceeding 1 A and 1 kV.

### 4.5.2. *Phototransistors for optical telecommunications*

This application, which requires both a good response coefficient and speed, mostly involves heterojunction bipolar transistors based on InP and InGaAs.

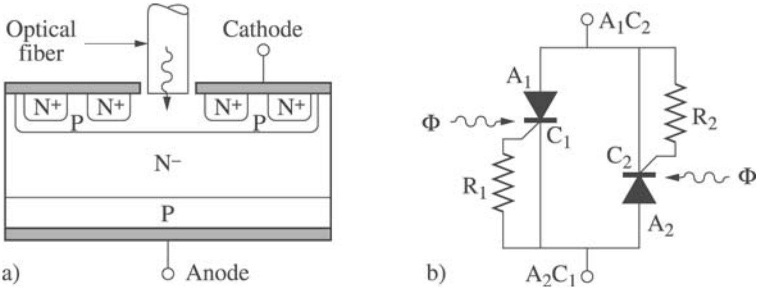


**Figure 4.16.** *(a) Schematic cross-section of a photothyristor and (b) the equivalent electric circuit for a phototriac*

In the basic process of an optical transmission system, shown in Figure 4.17, the phototransistor is the device which converts the optical signal into an electrical signal. It is situated at the input of the receiver, acting as an optical detector. In the transmitter, the light is modulated (directly or indirectly via an optical modulator) with the signal carrying the information. This signal can be of an analog or digital. After transport through the optical fiber, the optical signal is demodulated by the photodetector in order to recover the electrical signal. This signal is subject to noise and distortion, and so the receiving circuit may need to amplify and reconstruct the signal in order to extract the original information.

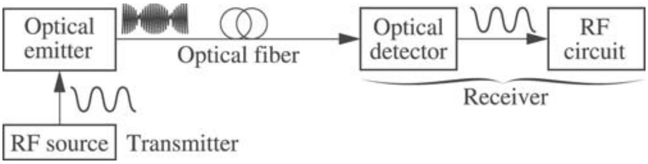


**Figure 4.17.** *Schematic diagram of an optical transmission system with an analog transported signal*

An InP/InGaAs phototransistor is an alternative to traditional photodetection, consisting of a PIN photodiode associated with a preamplifier. As was previously stated, this device is of interest because of its combination of photodetection and amplification behavior and, since its structure is similar to that of a heterojunction bipolar transistor, it can attain a frequency performance comparable to that of a HBT based on InP [BLA 00]. In addition, when acting as an optical-electrical mixer, it can convert the demodulated signal to other frequency ranges depending on the required output.

#### 4.5.2.1. *InP/InGaAs phototransistors as pre-amplifying photodetectors*

The transport of analog and digital data takes place at frequencies and bitrates that are constantly increasing. In digital transmission, for example, currently 10 and even 50 Gbit/s can be achieved. As a result, phototransistors must become faster and faster while maintaining a good level of sensitivity. The structure of a phototransistor based on InP/InGaAs is shown in Figure 4.18. This device is sensitive to optical wavelengths, $\lambda$ = 1.3 and 1.55 µm. The base and collector layers are made of InGaAs and the emitter layer is InP. The optical window sits directly above the base. This has been optimized to achieve an optical gain of 30 dB and a transition frequency of 88 GHz [GON 00]. These high performances are the result of many refinements, particularly to reduce the dimensions of the optical window, the base-emitter and base-collector junctions (in order to reduce their capacitances), as well as the use of a base layer with a graded composition, in order to increase the gain and speed.

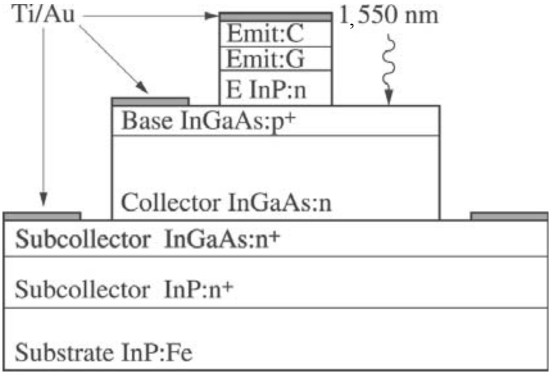


**Figure 4.18.** *Vertical structure of a phototransistor based on InP/InGaAs on a semi-insulating substrate of InP doped with Fe*

Figure 4.19 shows the response coefficient of the phototransistor in its two modes of operation, transistor (Tr Mode) and photodiode (PIN mode), as a function of the modulation frequency of light. The response coefficient (R) is expressed in dB, R (dB) = 20 log $S$, with $S$ expressed in A/W. The PIN mode corresponds with the phototransistor operating as a photodiode ($V_{BE}$ = 0V) and the Tr mode operating with the transistor effect ($V_{BE}$ > 0V). As discussed in section 4.3.2, the optical gain $G_{opt}$, expressed in dB, is equal to the difference between the response coefficient in the Tr mode and that of the PIN mode. The response coefficient in the PIN mode, $S_0$, is 0.25 A/W. This response coefficient $S_0$ has been increased to 0.44 A/W in another phototransistor, while maintaining an equivalent optical gain (32 dB). However, as predicted by equation [4.18], the transition frequency drops to 56 GHz [THU 99a].

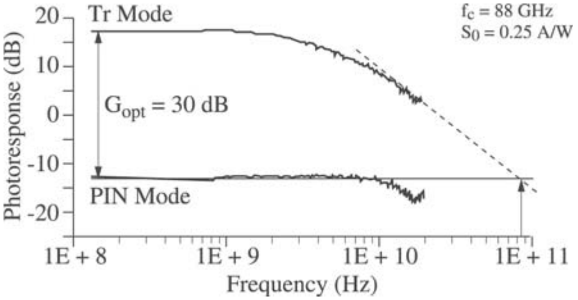


**Figure 4.19.** *Frequency response of an InP/InGaAs photodiode. The response coefficient (photoresponse) is expressed in dB*

Monolithic photoreceiver circuits for wideband amplification involving photo-HBTs and HBTs have been constructed. The amplifier circuit proposed and built by [KAM 00] exhibits a bandwidth of 40 GHz. The circuit diagram is shown in Figure 4.20. The circuit consists of a phototransistor and two HBT transistors. The two types of device have a double heterojunction and use the same epitaxial layers.

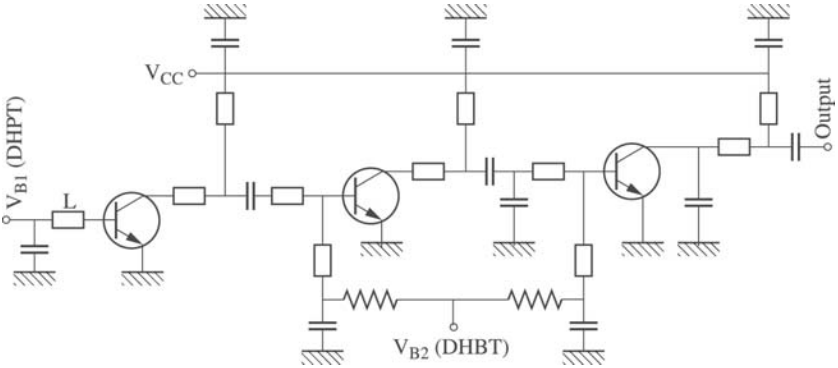


**Figure 4.20.** *Monolithic circuit diagram for wideband amplification consisting of a phototransistor and two double heterojunction transistors, using InP technology [KAM 00]. The structure of the epitaxial layers is the same for the two devices*

### 4.5.2.2. *InP/InGaAs phototransistors as an optoelectrical mixer*

The optoelectrical mixing transistor can be used in mixed radio/fiber network access, or in satellite communications, to achieve frequency conversion (up- or down-conversion). It can also be used in high bitrate communication systems as an integral part of timing-signal extraction circuits. Suematsu *et al.* [SUE 96] were among the first to use a photo-HBT based on GaAs, to achieve conversion from the modulation frequency of the light at 2 GHz to a frequency band at 50 GHz. Later, other investigations were reported [GON 98, GON 00], using InP/InGaAs technology, to achieve conversion from an intermediate frequency (≤2 GHz) to frequency bands at 30 and 40 GHz. An example of an experimental setup used to achieve this high frequency conversion is shown in Figure 4.21. The light emitted by a laser source at 1.55 µm is amplitude modulated at an intermediate frequency IF (via an external EOM modulator) in the frequency range of 200 MHz to 2 GHz. The modulated optical signal is injected into the optical window of the phototransistor via a single-mode optical fiber. The base contact is connected to a local oscillator *LO* at a frequency of 30 GHz. After demodulation of the optical signal, the phototransistor mixes the *OL* and *IF* signals and the result is visualized with the help of a spectrum analyzer.

Figure 4.22 shows the output electrical powers of the *IF* signal (in PIN mode) and the mixed signal *OL* + *IF*, as a function of the IF frequency. The conversion gain, using the definition given in [4.32], is the ratio of the output power of the *LO* + *IF* signal to that of the IF signal detected in PIN mode. In the frequency range of IF, from 0.2 to 2 GHz, the conversion gain for the *LO* + *IF* signal is relatively constant, of the order of 7 dB. This positive conversion gain is high because of the direct amplification of the mixed signal by the phototransistor.

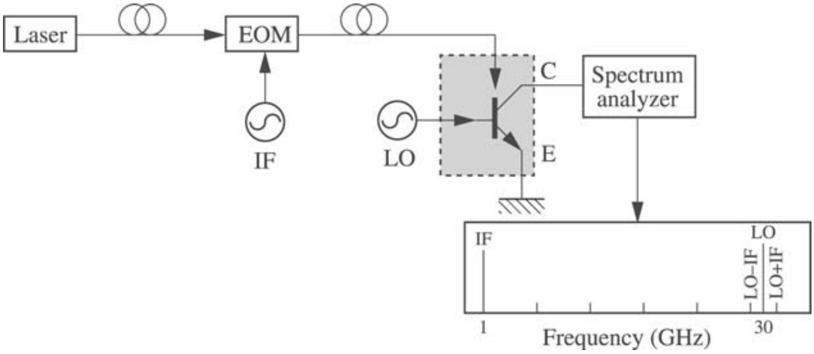


**Figure 4.21.** *Experimental setup for up-conversion from the intermediate frequency IF to a band at 30 GHz, using the HPT as an O/E mixer*

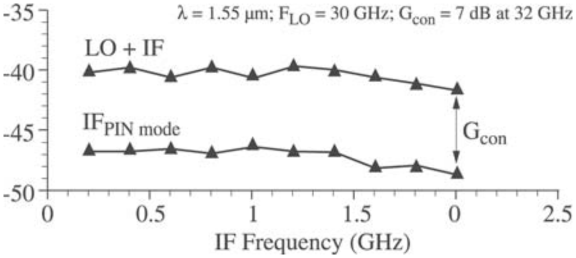


**Figure 4.22.** *Conversion gain for the IF signal transposed to a frequency of 30 + IF GHz, using the InP/InGaAs phototransistor as an O/E mixer*

## 4.6. Conclusion

In this chapter we have covered the main characteristics of the bipolar phototransistor. We have also presented its applications in the field of galvanic isolation for silicon phototransistors, and in the field of optical telecommunications for InP/InGaAs heterojunction phototransistors. The study of monolithic circuits incorporating InP/InGaAs phototransistors, and providing more complex functions such as optical-electrical mixing and auto-oscillation, is the particular aim of research intended to exploit the synergy between photonics and microwaves, with the goal of meeting the ever-growing requirements in sensitivity, speed and functionality of photodetectors.

## 4.7. Bibliography

[ALF 73] ALFAROV Z.I., AKHMEDOV F., KOROL'KOV V., NIKETIN V., "Phototransistor utilizing a GaAs-AlAs heterojunction", *Sov. Phys. Semicond.*, vol 7, no. 6, p. 780, 1973.

[ARN 92] ARNOULD J., MERLE P., *Dispositifs de l'électronique de puissance*, Hermes, Paris, 1992.

[BAA 77] BAAK C., ELZE G., WALF G., "GaAs MESFET: a high-speed optical detector", *Electron. Lett.,* vol. 13, p. 139, 1977.

[BAR 97] DE BARROS L.E., PAOLELLA A., FRANKEL M., ROMER M., HERCZFELD P., MADJAR A., "Photoreponse of microwave transistors to high-frequency modulated lightwave carrier signal", *IEEE Trans. on Microwave Theory Tech.,* vol. 45, no. 8, p. 1368-1374, 1997.

[BET 98] BETSER Y., RITTER D., LIU C.P., SEEDS A. J., MADJAR A., "A single stage three-terminal heterojunction bipolar transistor optoelectronic mixer", *J. Light. Tech.,* vol. 16, no. 4, p. 605-609, 1998.

[BLA 00] BLAYAC S., RIET M., BENCHIMOL J.L., BERDAGUER P., KAUFFMAN N., GODIN J., SCAVENNEC A., "Lateral design of InP/InGaAs DHBTs for 40 Gbit/s", *Proc. of the Indium Phosphide and Related Materials, IPRM'2000*, p. 481-484, 2000.

[CAM 82] CAMPBELL J.C., OGAWA K., "Heterojunction phototransistors for long-wavelength optical receivers", *J. Appl. Phys.*, vol. 53, no. 2, p. 1203-1208, 1982.

[CAM 85] CAMPBELL J.C., "Phototransistors for lightwave communications", *Semiconductors and Semimetals,* vol. 22D, p. 389-447, 1985.

[CAS 89] CASTAGNE R., DUCHEMIN J.P., GLOANEC M., RUMELHARD C., *Circuits intégrés en arséniure de gallium*, Masson, 1989.

[CHA 85] CHAND N., HOUSTON P.A., ROBSON P.N., "Gain of a heterojunction bipolar phototransistor", *IEEE Transactions on Electron Devices*, vol. ED-32, no. 3, p. 622-627, 1985.

[CHA 92] CHANDRASEKHAR S., LUNARDI L.M., GNAUCK A.H., RITTER D., HAMM R.A., PANISH G.L., QUA G.L., "A 10 Gbit/s OEIC photoreceiver using InP/InGaAs heterojunction bipolar transistors", *Electronics Letters,* vol. 28, no. 5, p. 466-468, 1992.

[ESC 95] ESCOTTE L., ROUX J. P., PLANA R., GRAFFEUIL J., GRUHLE A., "Noise modelling of microwave heterojunction bipolar transistors", *IEEE Transactions on Electron Devices*, vol. 42, no. 5, p. 883-888, 1995.

[FRE 96] FREEMAN P., ZHANG X., VURGAFTMAN I., SINGH J., BHATTACHARYA P., "Optical control of 14 GHz MMIC oscillators based on InAlAs: InGaAs HBTs with monolithically integrated optical wave guides", *IEEE Transactions Electrons Devices*, vol. 43, no. 3, p. 373-379, 1996.

[GAU 85] GAUTIER J.L., PASQUET D., POUVIL P., "Optical effects on the static and dynamic characteristics of a GaAs MESFET", *IEEE Trans. on Microwave Theory Tech.,* vol. 33, no. 9, p. 819-822, 1985.

[GON 98] GONZALEZ C., THURET J., RIET M., BENCHIMOL J.L., "Optoelectronic up-converter to millimetre-wave band using an heterojunction bipolar phototransistor", *Proc. of the European Conference on Optical Communication, ECOC'98*, vol. 1, p. 443-444, 1998.

[GON 00] GONZALEZ C., MULLER M., BENCHIMOL J.L., RIET M., JAFFRÉ P., LEGAUD P., "HBT phototransistor for remote upconversion in hybrid fibre radio distribution systems", *Proc. of the European Conference on Optical Communication, ECOC 2000*, vol. 2, p. 103-104, 2000.

[GON 00a] GONZALEZ C., MULLER M., BENCHIMOL J.L., RIET M., JAFFRÉ P., LEGAUD P., "A 28 GHz HPT/HBT monolithically integrated photoreceiver for hybrid fibre radio distribution systems"*, Proc. of the 8th IEEE International Symposium on Electron Devices for Microwave and Optoelectronic Applications, EDMO 2000*, p. 55-60, 2000.

[KAM 95] KAMITSUNA H., "Ultra-widebande monolithic photoreceivers using HBT compatible HPTs with novel base circuits and simultaneously integrated with an HBT amplifier", *J. Lightwave Tech.,* vol 13, no. 12, p. 2301-2307, 1995.

[KAM 00] KAMITSUNA H., MATSUOKA Y., YAMAHATA S. AND SHIGEKAWA N., "A 82 GHz-Optical-gain-cutoff-frequency InP/InGaAs double-hetero-structure phototransistor (DHPT) and its application to a 40 GHz band OEMMIC photoreceiver", *Proc. of the European Microwave week 2000*, *EuMC 2000*, *EuMC36*, 2000.

[KRO 57a] KROEMER H., "Quasi-electric and quasi-magnetic fields in non-uniform semiconductors", *RCA Rev.* vol. 18, p. 332, 1957.

[KRO 57b] KROEMER H., "Theory of a wide-gap-emitter for transistors", *Proc. IRE*, vol. 45, p. 1535, 1957.

[LAS 00] LASRI J., BILENCA A., EISENSTEIN G., RITTER D., ORENSTEIN M., SIDEROV V., GOLDGEIER S., COHEN S., "A two heterojunction bipolar photo-transistor configuration for millimeter wave generation and modulation", *Proc. of the International Topical Meeting on Microwave Photonics, MWP'00*, p. 62-65, 2000.

[LET 78] LETURCQ P., REY G., *Physique des composants actifs à semiconducteurs*, Dunod University, 1978.

[MAD 92] MADJAR A., PAOLELLA A., HERCZFELD P.R. "Analytical model for optically generated currents in GaAs MESFET", *IEEE Trans. on Microwave Theory Tech.,* vol. 40, p. 1681-1691, 1992.

[MON 71] de LA MONEDA F.H., CHENETTE E.R., VAN DER ZIEL A., "Noise in phototransistors", *IEEE Transactions on Electron Devices,* vol. ED-18, no. 6, p. 340-346, 1971.

[MOR 72] MORIIZUMI T., TAKAHASHI K., "Theoretical analysis of heterojunction phototransistors", *IEEE Transactions on Electron Devices,* vol. 19, no. 2, p. 152-159, 1972.

[POU 94] POUVIL P., *Composants semiconducteurs micro-ondes*, Masson, 1994.

[ROM 96] ROMERO M., MARTINEZ M., HERCZFELD, "An analytical model for the photodetection mechanisms in high-electron mobility transistors", *IEEE Trans. on Microwave Theory Tech.,* vol. 44, no. 12, p. 2279-2287, 1996.

[SAW 99] SAWADA H., IMAI N., "An experimental study on a self-oscillating optoelectronic up-converter that uses a heterojunction bipolar transistor", *IEEE Trans. on Microwave Theory Tech.*, vol. 47, no. 8, p.1515-1521, 1999.

[SEE 90] SEEDS A. J., DE SALLES A.A.A., "Optical control of microwave devices", *IEEE Trans. on Microwave Theory Tech.*, vol. 38, no. 5, p. 577-585, 1990.

[SHI 53] SHIVE J.N., "The properties of germanium phototransistors", *J. Opt. Soc. Am.*, vol. 43, p. 239, 1953.

[SHO 51] SHOOCKLEY W., SPARKS M., TEAL G.K., "p-n junctions transistors", *Phys. Rev.*, vol. 83, p. 151, 1951.

[SIM 86] SIMONS R., BHASIN K., "Analysis of optically controlled microwave/millimeter- wave device structures", *IEEE Trans. on Microwave Theory Tech.,* vol. MTT-34, no. 12, p. 1349-1355, 1986.

[SMI 80] SMITH R.D., PERSONICK S.D., *Semiconductor Devices for Optical Communications*, Springer, Berlin, Heidelberg, p. 89, 1980.

[SUE 96] SUEMATSU E., IMAI N., "A fiber optic: millimeter-wave radio transmission link using HBT as direct photodetector and an optoelectronic upconverter", *IEEE Trans. on Microwave Theory Tech.,*vol. 44, no. 1, p. 133-142, 1996.

[THU 99] THURET J., Phototransistor bipolaire à hétérojonction InP/InGaAs pour conversion optique/bande millimétrique dans les réseaux de distribution hybride radio sur fibre, Thesis, University of Paris VI, 1999.

[THU 99a] THURET J., GONZALEZ C., BENCHIMOL J.L., RIET M., BERDAGUER P., "High-speed InP/InGaAs heterojunction phototransistor for millimetre-wave fibre radio communications", *Proc. of the Indium Phosphide and Related Materials, IPRM'99*, p. 389-392.

[WAN 86] WANG H., Photorécepteur monolithique intégrant un phototransistor et des transistors bipolaires à heterojonction GaAlAs/GaAs pour transmission par fibre optique, Thesis, University of Paris XI, 1986.

[WAN 86a] WANG H., ANKRI D., "Monolithic integrated photoreceiver implemented with GaAs/GaAlAs heterojunction bipolar phototransistor and transistors", *Electronics Letters,* vol. 22, no. 7, p. 391-393, 1986.

# Chapter 5

# Metal-Semiconductor-Metal Photodiodes

## 5.1. Introduction

Among the different semiconductor photodetectors, the metal-semiconductor-metal (MSM) photodiode has a special place. It is a planar photodiode based on simple technology, it is easy to integrate, but it has not, up to now, gained the interest of the large telecommunication firms. Currently, to the best of our knowledge, only one device for millimeter applications has been commercialized. However, we will see that this photodetector is not without benefits, and indeed has very specific properties linked to its interdigitated structure. In this discussion, we will first consider the operation and structure of this photodiode before discussing its different characteristics: its response coefficient, dynamic behavior and noise. We will conclude with integration possibilities for the device.

---

Chapter written by Joseph HARARI and Vincent MAGNIN.

## 5.2. Operation and structure

### 5.2.1. *Fundamentals*

An MSM photodiode consists of two Schottky contacts on a semiconductor. Because of this, with each contact having a current-voltage characteristic *I(V)* similar to that of a traditional junction, the MSM photodiode behaves like two diodes in series, facing in opposite directions. Briefly, under bias, with one of the diodes necessarily being reverse-biased, the dark current crossing the structure is weak and, to the extent that the two Schottky contacts are identical, the symmetry of the device gives rise to a symmetric global *I(V)* characteristic, with each half (positive or negative bias) corresponding to the behavior of a reverse-biased metal-semiconductor junction.

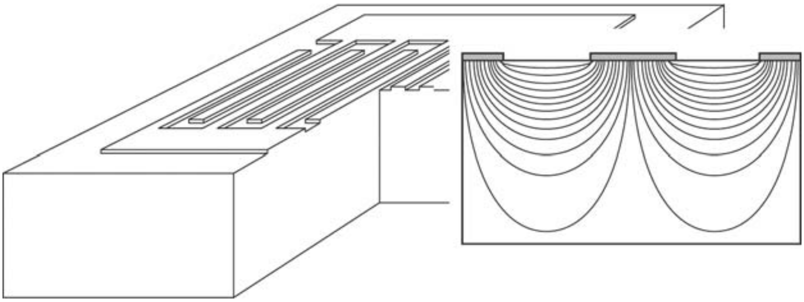


**Figure 5.1.** *MSM photodetector with its typical interdigitated structure. Inset: current lines in the structure*

If the bias voltage is sufficiently high, the region between the electrodes is completely depleted and this produces an electric field. The basic principle of this photodetector is that light is absorbed in a semiconductor layer forming part of the depletion region, such that the electron-hole pairs generated are separated under the effects of the electric field, with each carrier then being directly collected by an electrode. This effect, familiar in photodiodes, takes place in a planar structure whose geometry is shown in Figure 5.1. In general, the electrodes have an interdigitated shape, which allows the carriers to be collected over a large semiconductor area while maintaining a short inter-electrode distance. If, for example, illumination is from above, this allows carriers to be collected over the whole surface of the light beam emitted by a fiber in the near field. The inset in Figure 5.1 shows the current lines between several fingers of the electrode, and shows that, due to the planar structure, the deeper a carrier is photogenerated in the material, the longer it will take to reach the electrode. This leads to a very specific distribution of transit times for the carriers. As with all photodiodes, an MSM photodetector under illumination can be characterized by its response coefficient, its cutoff frequency and its photodetection noise. However, before we consider these different aspects, we will discuss the behavior of the device in the dark.

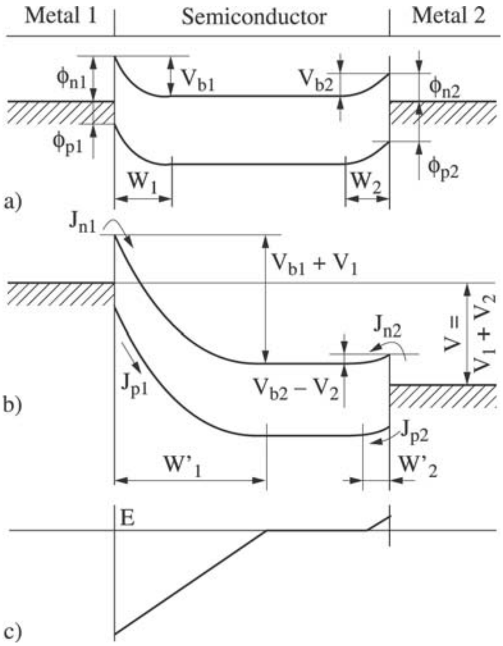


**Figure 5.2.** *Behavior of an MSM photodetector in the dark: a) band structure in thermodynamic equilibrium, b) band structure under bias, c) corresponding electric field under bias*

Figure 5.2 shows the evolution of the band structure of an MSM structure under bias. The two electrodes are assumed to be of different metals, so that the band structure of each metal-semiconductor junction differs. On each side, $\phi_n$ is the height of the barrier seen by the electrons in the metal, $\phi_p$ is the height for holes and $V_b$ is the *built-in* bias of each junction.

In thermodynamic equilibrium (see Figure 5.2a), the two depletion regions depend on the doping of the semiconductor and the work function which separates the Fermi level in the metal from the vacuum level in the semiconductor. Since the semiconductor is normally weakly doped, the carriers mostly cross the potential barriers through thermoelectric effects [SZE 96], in other words only those carriers with sufficient energy cross the barrier. The sum of all the currents across the barriers is zero. Under bias, as shown in Figure 5.2b, the situation is more complicated, with one of the junctions (number 1 in this case) being reverse-biased (voltage $V_1$), and the other being forward-biased (voltage $V_2$). It is the current densities of electrons $J_{n1}$ and $J_{n2}$, and of holes $J_{p1}$ and $J_{p2}$ which determine the behavior [SZE 71]. If we ignore recombination in the region between the electrodes, we have:

$$J_{n1} = J_{n2} \text{ and } J_{p1} = J_{p2} \qquad [5.1]$$

These relationships show that the forward bias voltage of junction 2 is such that the current flowing across it is equal to the current allowed across junction 1 in reverse. Under these conditions, it is the height of the barrier $\phi_{n1}$ which determines the value of $J_{n1}$ (and hence $J_{n2}$), and it is $\phi_{p2}$ which determines the value of $J_{p2}$ (and $J_{p1}$). The value of the total dark current thus depends on the heights of the barriers, which are linked to the nature of the metal used for each electrode. In parallel to this process, the depletion region of junction 1 expands, while that of junction 2 contracts slightly. Under these conditions we can define two particular bias voltages which depend on the structure of the MSM under consideration: the first is the total depletion voltage $V_{TD}$, which is reached when the depletion region of the reverse-biased junction reaches that of the forward-biased junction, that is, when:

$$W_1 + W_2 = s \quad [5.2]$$

where $s$ is the inter-electrode distance.

The second is the flat band voltage $V_{FB}$, reached when $W_1 = s$. The electric field present in the inter-electrode region is thus only zero at electrode 2 (the anode), and it is at a maximum at the other side (at the cathode). Clearly, an MSM photodetector is only useful when the bias voltage is greater than $V_{FB}$. If we assume a homogenous N-type semiconductor, with doping level $N_d$ and permittivity $\varepsilon_s$, we have:

$$V_{FB} = \frac{q \cdot N_d \cdot s}{2 \cdot \varepsilon_s} - V_{b1} + V_{b2} \quad [5.3]$$

When the applied bias is increased, the electric field increases at the cathode, which leads gradually to a breakdown process linked either to shock ionization in the semiconductor material or to tunneling through the potential barriers of the metal-semiconductor barriers from band to band. All this leads to the typical current-voltage characteristics displayed in Figure 5.3. For $V < V_{FB}$, the dark current crossing the photodiode is weak and grows slowly with the bias, until it reaches a saturation value which is maintained until breakdown, which can be more or less gradual with voltage. If the two electrodes are made of the same metal, the *I(V)* characteristics in the dark are, in principle, symmetric with respect to the $V = 0$ V axis. Generally, since a photodetector must operate with a weak dark current, the best bias conditions for a photodiode are:

– a bias voltage considerably below the breakdown voltage, in order to limit the dark current ($V < V_C$);

– an inter-electrode region that is completely depleted, with a sufficiently high internal electric field so that the velocities of the photogenerated charge carriers are in the saturation system (in any case, $V > V_{FB}$).

Of course, given the planar structure of the device, some elements must be added to this general description.

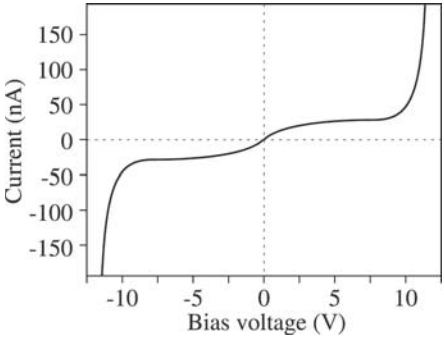


**Figure 5.3.** *Typical current-voltage characteristics of an MSM photodetector in the dark*

The biasing of the semiconductor, using electrodes placed above it, leads to a particular electric field distribution, as shown in Figure 5.4. Initially, the field decreases with depth in the device, and then there are local maxima in the electric field just under the edge of each electrode (or each finger for an interdigitated structure). Consequently, the breakdown of the device always begins at the edge of the electrode, and a very high electric field may exist close to the electrodes, but the field is weaker in the middle of the inter-electrode space and deep in the photodetector. Finally, as the field is weaker deeper in the structure, the carriers that are photogenerated deep in the structure not only have a great distance to travel to reach the electrode (as we have already seen in Figure 5.1), but also have a small drift speed, which increases the transit delay and decreases the dynamic performance of the photodetector. All this confirms that the optimal bias voltage of such a device must be selected with care in order to obtain good dynamic performances without degradation.

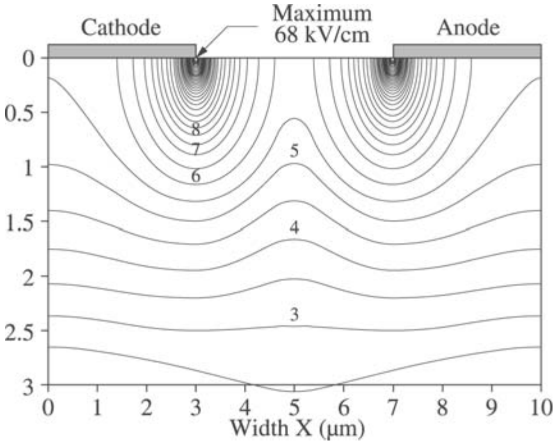


**Figure 5.4.** *Distribution of the modulus of the electric field in an MSM photodetector. The applied bias is 4 V, and the values of the field are in kV/cm. The semiconductor is weakly doped (Nd = $5 \times 10^{14}$ cm$^{-3}$)*

## 5.2.2. *Materials used*

### 5.2.2.1. *Electrode composition*

Since the composition of the electrodes has a direct influence on the band structure of the photodetector due to the magnitudes of $\phi_{n1}$, $\phi_{p1}$, $\phi_{n2}$ and $\phi_{p2}$, as shown in Figure 5.2, it will have an effect on the dark current resulting from the thermoelectric effect. If we bias far from the breakdown voltage, the phenomena limiting the dark current are the thermoelectric passage of electrons across junction 1, and the thermoelectric passage of holes across junction 2. The total dark current is thus commonly written as:

$$J = \left[ A_n^* T^2 \cdot \exp\left[ -\frac{q\left(\phi_{n1} - \Delta\phi_{n1}\right)}{kT} \right] + A_p^* T^2 \cdot \exp\left[ -\frac{q\left(\phi_{p2} - \Delta\phi_{p2}\right)}{kT} \right] \right] \cdot \left( 1 - \exp\left( -\frac{qV}{kT} \right) \right) \qquad [5.4]$$

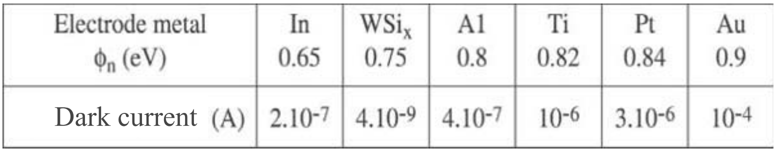
where $A_n^*$ and $A_p^*$ are the Richardson constants associated with electrons and holes, $T$ is the absolute temperature, $V$ is the applied voltage, $\phi_{n1}$ and $\phi_{p2}$ are the heights of the barriers seen by electrons and holes respectively, corrected with the terms introduced by the effects of the electric field. This results in [SZE 96]:

$$\Delta\phi_{n1,p2} = \sqrt{\frac{qE_{1\,2}}{4\pi\varepsilon_s}} \qquad [5.5]$$

where $\varepsilon_s$ is the permittivity of the semiconductor and $E_{1,2}$ is the value of the electric field at the metal-semiconductor junction. Greater barrier heights thus minimize the dark current. If we assume that the two electrodes are made of the same material, we have (see Figure 5.2):

$$\phi_{n1} = E_g - \phi_{p2} \qquad [5.6]$$

Simultaneously achieving maximum-height barriers for electrons and holes requires us to choose the contact metal for which $\phi_n$ is as close as possible to $E_g/2$. Measurements of the dark current [ITO 86], made for a gallium arsenide (GaAs) MSM photodetector for different electrode metals, are presented in Table 5.1.



| Electrode metal<br>$\phi_n$ (eV) | In<br>0.65 | $WSi_x$<br>0.75 | A1<br>0.8 | Ti<br>0.82 | Pt<br>0.84 | Au<br>0.9 |
|---|---|---|---|---|---|---|
| Dark current (A) | $2.10^{-7}$ | $4.10^{-9}$ | $4.10^{-7}$ | $10^{-6}$ | $3.10^{-6}$ | $10^{-4}$ |

**Table 5.1.** *Dark current measured for GaAs MSM photodetectors ($E_g = 1.42$ eV) for different electrode metals and photodetectors with the same surface*

Because of this, for this type of photodetector it is desirable to decide on the cathode and anode at the fabrication stage, choosing a metal with a high electron barrier $\phi_n$ for the first electrode and a high $\phi_p$ barrier for the second. Under these conditions, the values of the dark current are of the order of nano-amps, which are compatible with all possible uses of the device.

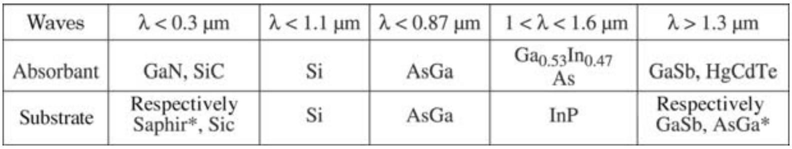
Clearly, all this assumes that no other effect, associated with fabrication defects, surpasses the thermoelectric effect which is normally the main cause of the dark current. This can occur if the band structure is perturbed close to the metal-semiconductor junction. In this case, a narrowing of the potential barrier favors the passage of carriers through the tunnel effect, to the detriment of the thermoelectric effect. The tunnel effect can be associated with the presence of a native oxide layer at the interface, which leads to an accumulation of electrons or holes; alternatively, the effect can be assisted by trapping states present in the semiconductor material, which create energy levels inside the bandgap. As a result, attention must be paid to the epitaxy, the cleanliness of the semiconductor surface before the metal deposition, and above all to the passivation of the structure. In the case of a photodetector illuminated from above, the dielectric layer used for this final operation can operate as an anti-reflection coating.

### 5.2.2.2. *Epitaxial structure*

The epitaxial structure of an MSM is designed starting with the absorbing layer, which itself depends on the wavelength to be absorbed according to the basic law:

$$\frac{h \cdot c}{\lambda} > E_g \tag{5.7}$$

where h is the Planck constant and c is the speed of light. The different materials already reported in this application are given in Table 5.2 along with the substrate they are grown on (an asterisk indicates growth with a lattice mismatch) and the associated wavelengths.



| Waves | $\lambda < 0.3$ µm | $\lambda < 1.1$ µm | $\lambda < 0.87$ µm | $1 < \lambda < 1.6$ µm | $\lambda > 1.3$ µm |
|---|---|---|---|---|---|
| Absorbant | GaN, SiC | Si | AsGa | $Ga_{0.53}In_{0.47}As$ | GaSb, HgCdTe |
| Substrate | Respectively Saphir*, Sic | Si | AsGa | InP | Respectively GaSb, AsGa* |

**Table 5.2.** *Absorbing materials already used to fabricate MSM photodetectors, along with their growth substrate and their associated wavelengths*

As Figure 5.5 shows, around the absorption layer there are normally two layers with a higher bandgap. The adaptation layer acts to isolate the absorbing layer from the substrate during growth, and the barrier layer acts to increase the potential barrier at the level of the metal-semiconductor junction [BUR 91, WOH 97]. A barrier layer is required, for example, when the absorbing material is N-type GaInAs, on which the Schottky contacts are poor quality. Furthermore, to improve the passage of carriers from the absorbing layer to the electrodes, we also introduce one or more transition layers which reduce the discrepancy in the band structure as seen by free carriers, between the barrier layer and the absorbing layer [WAD 89, ZHA 96]. These very thin layers allow a gentle transition in the band structure and prevent the accumulation of carriers in the upper part of the absorbing layer under the electrodes. This accumulation leads to a drop in the external quantum efficiency of the device, as a non-negligible number of free carriers recombine while they are blocked at the interface. It also simultaneously leads to an increase in the transit time for charge carriers, which has an adverse effect on the frequency response of the photodetector.

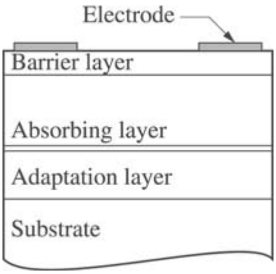


**Figure 5.5.** *Typical epitaxial structure of an MSM photodetector. The semiconductor is weakly doped (Nd = $5 \times 10^{14}$ cm$^{-3}$)*

The thickness of the absorbing layer depends mostly on its absorption coefficient because, as we will see, this directly influences the response coefficient of the photodetector when it is illuminated from above or below. The thickness of the adaptation layer is normally greater than one micron, while that of the barrier layer is as thin as possible (<0.2 µm in general). In some cases, for example, that of a gallium arsenide MSM or a silicon MSM, the substrate and the adaptation layer are also absorbing, but this requires the collection of carriers photogenerated deep in the structure which, as we have already seen, decreases the dynamic performance of the photodiode. To avoid this problem, the adaptation layer of the GaAs MSM can be made of non-absorbing AlGaAs matched to the lattice of the GaAs substrate. In the case of silicon, SIMOX wafers are used, which have a layer of silicon underneath.

## 5.3. Static and dynamic characteristics

### 5.3.1. *Response coefficient*

The response coefficient, which provides the photocurrent generated per Watt of incident light (in units of A/W as a result), initially depends on the mode of illumination. The photodetector can be illuminated from above, across the interdigitated metallic structure, from below or side-on, through an optical waveguide. The case of lateral illumination must be treated on a case-by-case basis because the response coefficient depends both on the optical coupling between the waveguide and the detector, and between the fiber and the waveguide. This requires a detailed study of the propagation of the light in the multi-layer structure of the device. Illumination from below is similar to illumination from above, apart from the absence of electrodes to partially occlude the light entering the device; in addition, the absorption associated with the substrate must be taken into account if it is not thin and not perfectly transparent. The absorption coefficient of the substrate is generally weak at the wavelengths under consideration, but the thickness of the material to be crossed can be sufficiently large to have an effect on the response coefficient. Illumination from above is common. In this case, the response coefficient is:

$$\mathrm{R} = \frac{q}{h\nu} \cdot \eta \qquad [5.8]$$

where $q$ is the electronic charge, $h\nu$ the energy of a photon of the incident light and $\eta$ the quantum efficiency, which is itself given by:

$$\eta = (1 - R) \cdot \frac{s}{(s + d)} \cdot \left[1 - \exp(-\alpha \cdot W_a)\right] \qquad [5.9]$$

where $R$ is the reflection coefficient at the surface of the device, $s$ is the distance between two fingers, $d$ the width of a finger, $\alpha$ the absorption coefficient and $W_a$ the thickness of the absorbing layer. The ratio $s/(s + d)$ introduces the shadowing effect of the electrodes which is typical of this photodetector. In brief, the response coefficient is directly linked to the characteristics of the absorbing layer and the geometry of the interdigitated electrodes.

As far as the characteristics of the absorbing layer are concerned, in the case of a semi-infinite layer, the absorption coefficient must be sufficiently high so that all the carriers are photogenerated in the depletion region. This is the case for GaInAs at wavelengths of 1.3 and 1.55 µm, and for GaAs around 0.8 µm. In contrast, for silicon, the absorption coefficient at 0.8 µm is of the order of 10 $\mu m^{-1}$. This reduces the quantum efficiency of the photodetector and causes a reduction in the bandwidth, because at least some of the carriers photogenerated far from the depletion region reach it through diffusion, which is much slower here than the transport under the effects of the electric field. This clearly depends on the inter-electrode distance, as we will see. Finally, we can see that optimizing the transport of the carriers across the planar structure implies a relationship between the thickness $W_a$ of the absorbing region and the inter-electrode distance $s$.

If we now consider the geometry of the electrodes, we find that attempts to reduce the shadowing effect leads to two approaches. The first consists of using opaque electrodes while reducing the finger width relative to the inter-electrode distance. However, this increases the resistance of the finger, which has a detrimental effect on the bandwidth. Measurements made on electrodes deposited on silica show that the effective resistance of a metallic finger is greater than that which would be calculated based on the volume of the material.

This effect is more important the narrower the finger, importance being obtained at around 0.1 µm. This may be due to metallic limitations that the structure places on electron transport becoming important. In any case, for fingers made of titanium (150 Å) and gold (350 Å), and of width 0.06 µm, the measured resistance is around 80 Ω/µm instead of 11 Ω/µm for $d$ = 0.5 µm. This gives an idea of the sizes to be considered when designing the electrodes. As a result, $s$ and $d$ normally take similar values, within the range $d \leq s \leq 4d$. The second approach consists of research into electrodes made of transparent materials. Several possibilities exist, notably thin layers of gold [MAT 96] and doped oxides: indium tin oxide (ITO) or cadmium tin oxide (CTO). In each case, there is a tradeoff between the transparency at the wavelength under consideration and the resistance of the material. For example, doped ITO is remarkably transparent at short wavelengths ($\lambda$ < 0.85 µm), but it requires a compromise between resistance and transparency at longer wavelengths as doping reduces its transparency [GAO 94, SEO 93]. The same is true for CTO. In any case, an optical transmission of 100% cannot be achieved at long wavelengths. Furthermore, attention must be paid to the conditions of the deposition, which for the oxides is normally through sputtering, which does not favor particularly good characteristics of the electrode-semiconductor interface. In short, while the response coefficient is an absolute priority, the deposition of transparent electrodes clearly reduces the shadowing effect, but it is not without consequences in terms of the dynamic behavior of the photodetector. Firstly, it leads to an increase in the finger resistance, the consequences of which can be seen in the equivalent circuit of the MSM in the dynamic system. Also, the electric field diagram (Figure 5.4) shows that this is weak in the medium under each finger. The drift velocity of the photocarriers will therefore be weak at these points. As a result, the electrons photogenerated just under the transparent cathode, and the holes photogenerated under the anode, will take a long time to reach the opposite electrode. Not only is their drift velocity low as they travel through areas of weak field, but they also have a long way to travel (see Figure 5.1). Thus, transparent electrodes increase the mean transit time and decrease the bandwidth. A similar effect on the mean transit time is seen when the device is illuminated from below. In this case, the majority of the carriers are photogenerated in the lower part of the absorbing layer, as much underneath the fingers as between them. Since this is a region of weak electric field, this also leads to a longer mean transit time.

Still on the subject of the electrode geometry, a property that cannot be ignored in the case of illumination from above is the transparency of the network of fingers when its period $p$ ($p = d + s$) becomes smaller than the wavelength of light. Then we observe typical optical behavior which is more complicated than a simple shadowing effect. The few studies carried out on this subject [KUT 94] show that the transmission of light through the metallic network depends on the ratios of $\lambda/p$ and $\lambda/d$, on the finger thickness, on the refractive index $n_s$ of the semiconductor, and on whether the optical polarization is parallel or perpendicular to the direction of the fingers. It is higher in the perpendicular case than in the parallel case, where only a few percent of the optical energy reaches the semiconductor. Furthermore, it is a minimum for $\lambda/p = 1$ and $\lambda/p = n_s$. The finger thickness becomes relevant when it is close to its width. All these complex properties, which must be modeled on a case-by-case basis, are sharp enough to make it possible to built a wavelength discriminator using two parallel MSMs with slightly different periods, illuminated from the same beam and measuring the ratio of the photocurrents [CHE 97].

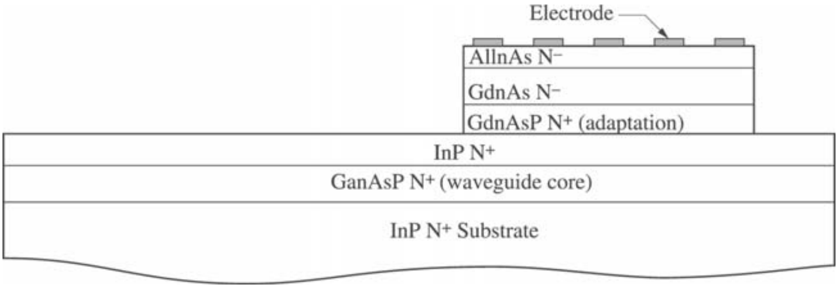


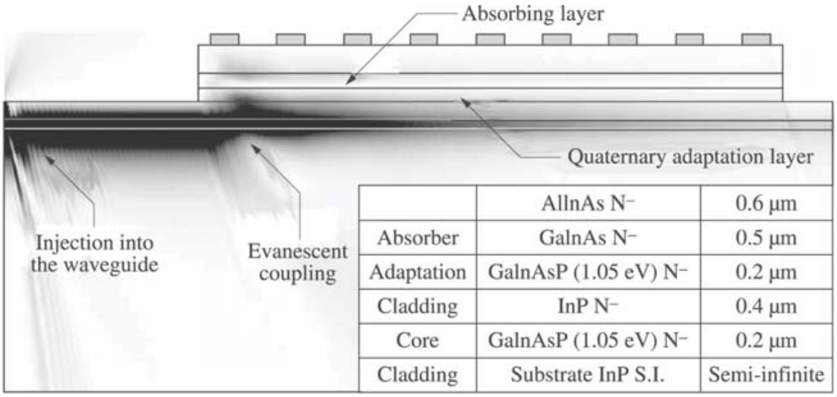
**Figure 5.6.** *Example of the epitaxial structure of an MSM photodetector on top of an optical waveguide*

In order to understand all the modes of illumination, we will now consider the case of lateral illumination. The electrode structure of the MSM is particularly favorable to the injection of light through an optical waveguide [SOO 88, VIN 89]. Indeed, the direct lateral illumination, for which absorption occurs in the first few microns after the cut face, is difficult to achieve, as is the case for a PIN photodiode. The example in Figure 5.6, of an MSM structure on top of a waveguide, is based on an N-type InP/GaInAsP/InP optical waveguide, the standard for this type of technology. The quaternary GaInAsP layer (with bandgap $E_g$ = 1.05 eV), formed into a lattice on InP, is what forms the guiding layer, with InP acting as the cladding. Coupling with the photodetector occurs through evanescent waves. Modeling of the optical propagation through such a structure (see Figure 5.7) shows how the injected beam couples with the absorbing layer through the intermediary of the thin quaternary adaptation layer whose characteristics (thickness and composition) determine the absorbed wavelength. Of course, the photocarriers are still generated in the lower part of the absorbing layer, but the thickness of this layer can be selected to be quite thin without significantly reducing the quantum efficiency. For a fixed adaptation layer, the effect of the thickness of GaInAs on the internal quantum efficiency (see Figure 5.8) shows that, for TE and TM optical polarizations at a wavelength of 1.55 µm, a certain number of optimum cases exist, even for small thicknesses.

At a conceptual level, to achieve the shortest possible absorption length, it is necessary to carefully choose both the thickness of the adaptation layer and that of the absorbing layer. The response coefficient is:

$$\mathrm{R} = \left(1 - R\right) \cdot K \cdot \frac{q}{h\nu} \cdot \eta_{\text{internal}} \qquad [5.10]$$

where $\eta_{\text{internal}}$ is the internal quantum efficiency determined by the epitaxial structure, $R$ is the reflection coefficient at the entrance of the waveguide and $K$ is the coupling coefficient between the beam leaving the optical fiber and the semiconductor waveguide. The internal quantum efficiency can only be calculated through detailed modeling of the optical transport in the material structure of the photodetector/waveguide. The dynamic behavior of such a structure is comparable to that seen with illumination from below; however, the thin width of the absorbing layer allows a short transit time to be retained. Thus, this type of structure allows a high response coefficient to be combined with a small absorber thickness.



| | | |
|---|---|---|
| | AlInAs N⁻ | 0.6 μm |
| Absorber | GaInAs N⁻ | 0.5 μm |
| Adaptation | GaInAsP (1.05 eV) N⁻ | 0.2 μm |
| Cladding | InP N⁻ | 0.4 μm |
| Core | GaInAsP (1.05 eV) N⁻ | 0.2 μm |
| Cladding | Substrate InP S.I. | Semi-infinite |

**Figure 5.7.** *Propagation and absorption of light in a MSM on a waveguide with evanescent coupling*

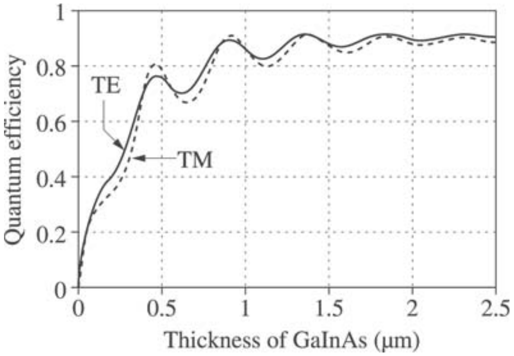


**Figure 5.8.** *Evolution of the quantum efficiency of the detector/waveguide as a function of the thickness of the absorbing material*

To conclude, we will discuss all the internal electronic effects in the device which affect the response coefficient. Typically, this is reduced under the effect of recombination and increases with internal gain. Recombination occurs either at the surface of the semiconductor between to fingers or at the interface between two epitaxial layers, wherever there is a potential barrier associated with the transition between different materials, which blocks the transport of carriers and leads to their accumulation. In the first case, careful passivation of the surface of the photodetector eliminates the problem. The second case is relevant when a barrier layer is introduced to improve the quality of the contact with the electrode. A transition layer between the absorber and the barrier is thus required, made of an epitaxial layer with variable bandgap or even a super-lattice [WAD 89, ZHA 96].

There are many effects which lead to gain [KLI 94]. They mostly consist of two types: those connected with the trapping of carriers, which operate at low frequencies (below a few tens of megahertz), and shock ionization close to the edges of the fingers, whose cutoff frequency is higher, as can be seen in the case of avalanche photodiodes. In the case of trapping, the gain effect is due to a photocarrier becoming trapped in the semiconductor for a time $\tau_p$ greater than the inter-electrode transit time $\tau_T$ which triggers the injection of additional carriers in order to maintain electrical neutrality. As long as this carrier remains trapped, a carrier with the opposite charge will cross the device to provide electrical compensation for it. The gain is then given by:

$$G_{trapping} = \frac{\tau_p}{\tau_T} \qquad [5.11]$$

This trapping can occur at the surface of the semiconductor, as in the case of GaAs photoconductors, or at the interface between two epitaxial layers. When it occurs at the surface, the holes are mostly trapped close to the cathode and electrons trapped close to the anode, this creates an asymmetric charge distribution. Hetero-interface trapping is associated with epitaxial growth defects, and normally a barrier layer is used to eliminate surface effects. Generally, attempts are made to limit both recombination and trapping in order to obtain a photodetector with constant dynamic behavior up to the high cutoff frequency, while ensuring a good reliability.

### 5.3.2. *Dynamic behavior*

As with all photodetectors, the dynamic behavior of MSMs is determined by their capacitance and their input and output resistances, in other words their equivalent electric circuit under modulated electric power, and the transit time of the photocarriers. The equivalent small-signal circuit (see Figure 5.9) shows:

– the intrinsic capacitance $C_{PD}$ and resistance $R_{PD}$ of the device;

– the finger resistance $R_F$, the capacitance and the resistance of the contact, $C_{CONTACT}$ and $R_{CONTACT}$;

– the connection wires with inductance $L_F$ and capacitance $C_W$.

The finger resistance is given by:

$$R_F = 2 \cdot R_0 \cdot \frac{L}{N} \qquad [5.12]$$

where $R_0$ is the resistance per unit length of each finger, L the length and N the number of fingers. In the equivalent circuit, $R_L$ is the load resistance and we see how the finger resistance can, in some cases, increase the time constant *RC* of the device. One of the important terms is the intrinsic capacitance of the MSM. A standard analysis of the planar structure [LIM 68] shows that:

$$C_{PD} = \frac{A}{(d+s)} \cdot \varepsilon_0 \cdot (1+\varepsilon_R) \cdot \frac{K(k)}{K(k')} \qquad [5.13]$$

where *A* is the surface area of the active region, $\varepsilon_0$ and $\varepsilon_R$ are the vacuum and relative permittivities, and finally:

$$K(k) = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1-k^2\sin^2(\varphi)}} \text{ with } k = \tan^2\left(\frac{\pi.d}{4\cdot(d+s)}\right) \text{ and } k' = \sqrt{1-k^2}$$

The calculation shows two details connected to the interdigitated planar structure of the MSM. Firstly, the capacitance of the device falls when the ratio *s/d* increases; the decrease in the shadow effect is thus inextricably linked to the capacitance of the photodetector. Secondly, for a photosensitive surface and an identical mean transit time, the capacitance of the MSM is 3 to 4 times weaker than that of a PIN photodiode. This aspect favors the MSM for high modulation frequencies (millimeter wavelengths), where the size of the devices is very small.

Typical values for the equivalent circuit can be given as a guide. The capacitance of the MSM lies between 10 and 1000 fF, its resistance between 1 and 100 Ω, and the capacitance of the contact is a significant parasitic element, generally less than 100 fF (and in all cases less than $C_{PD}$). Finally, the connection wires have an auto-inductance of the order of 10 pH and a capacitance of a few fF linking them to the body of the device. If we ignore all the parasitic elements, that is to say the connection wire, the capacitance of the contact, and the finger resistance, we find a standard time constant: $R_L C_{PD}$. The cutoff frequency associated with the capacitance effects is thus given by:

$$f_{CAPA} = \frac{1}{2 \cdot \pi \cdot R_L C_{PD}} \qquad [5.14]$$

Now turning to the transit time, a thin, two-dimensional physical model is required to account for the effects of the geometry of the electrode, the thickness of the absorbing layer, of the hetero-interfaces to be crossed, and of the electric field distribution produced by the bias voltage [ASH 95].

Nevertheless, a number of rules allow us to evaluate the cutoff linked to the transit time of the carriers. Firstly, the presence of high electric field areas just at the edges of a finger mean that it is difficult to appropriately bias a device with a large inter-electrode distance ($s >$ 3 μm). What happens in this case is that a local electric field is reached which is above the breakdown voltage, while the field in the middle of the inter-electrode space is still too weak to give the photocarriers their saturation velocity. This effect, visible in Figure

5.4, is more serious for holes, as they move more slowly than electrons. Even for small inter-electrode distances, the variation of the field in the device must be taken into account. The bias voltage is thus an important parameter that must be adjusted to limit the mean transit time while avoiding breakdown.

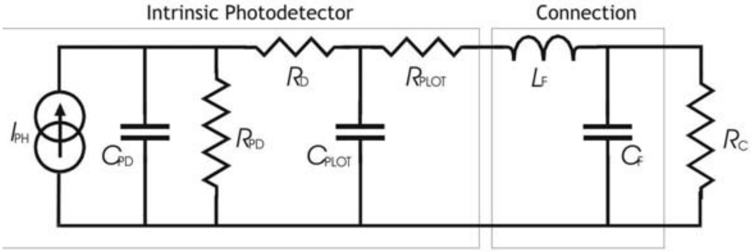


**Figure 5.9.** *Equivalent circuit for an MSM photodetector in the case of a small signal*

Secondly, the distribution of the current lines in the device favors a certain ratio between the thickness of the absorbing layer and the inter-electrode distance. The effect of transit time is smallest when this value is close to 2 ($W_a/s = 2$). If this is achieved, the mean transit time in the structure is:

$$\tau = \delta \frac{s}{2 \cdot v_{sat}} \qquad [5.15]$$

where $v_{sat}$ is the saturation velocity of the carriers (we will take the same velocity for holes as for electrons), $s$ is the inter-electrode distance and $\delta$ accounts for the effects of the planar structure ($1 \leq \delta \leq 2$). The saturation velocity used is taken as equal to that of the holes for the most pessimistic results. The cutoff frequency associated with the transit time is thus given by:

$$f_T = \frac{1}{2 \cdot \pi \cdot \tau} \qquad [5.16]$$

The cutoff frequency of the MSM photodiode under these conditions is:

$$f = \left( \frac{1}{f_{CAPA}^2} + \frac{1}{f_T^2} \right)^{1/2} \qquad [5.17]$$

An increase in the inter-electrode distance will decrease the capacitance and increase the transit time. For every structure, then, there is a compromise to be made.

### **5.3.3.** *Noise*

The different sources of noise in MSMs, as in other photodetectors, are thermal noise, $1/f^{\alpha}$ noise and shot noise. Thermal noise is linked to the variations in energy of the carriers in resistive regions. It is independent of frequency and its spectral density is linked to the absolute temperature $T$ and the resistance $R$ by:

$$\left\langle i_{TH}^2 \right\rangle = \frac{4kT}{R} \left( \text{in } A^2/Hz \right) \qquad [5.18]$$

The $1/f^{\alpha}$ noise is associated with all the random generation and recombination effects in the semiconductor material. These effects are themselves linked to the different imperfections, in terms of irregularities in the structure of the crystal lattice or in terms of foreign atoms. This generates a specific noise whose spectral power density varies as $1/f^{\alpha}$, where $\alpha$ is normally close to 1. This spectral density is often higher in horizontal transport devices, such as MSMs or field effect transistors. Finally, the shot noise, familiar from PIN and avalanche photodiodes, is linked to the photogeneration process and the charge transport. Its spectral power density is:

$$\left\langle i_{SHOT}^2 \right\rangle = 2 \cdot q \cdot \left( I_{DARK} - I_{PH} \right) \left( \text{in } A^2/Hz \right) \qquad [5.19]$$

In principle, this is white noise, i.e. its spectral density is independent of frequency; however, in practice it follows the frequency response curve of the device, making it possible to obtain the frequency response of a photodetector using a continuous signal, by measuring the variation of its shot noise as a function of frequency.

Having covered these items, we will now describe the variation of the spectral noise density of the MSM photodetector in two situations: in normal operating conditions and close to breakdown in the conditions where a gain in the photocurrent, linked to the shock ionization effect, is possible.

In general, if the finger and contact resistances are sufficiently low, the thermal noise is much weaker than the shot noise, so that we can ignore its effects. Under normal operating conditions [SOO 91, WAD 88], the $1/f^{\alpha}$ noise dominates for frequencies of the order of 1 to 10 MHz, after which it is the shot noise which dominates and which must be considered when performing calculations for photoreceivers. This shot noise increases with the dark current, which is why a low dark current is desirable, normally below about 100 nano-amps.

However, as breakdown is approached [SCH 90], two changes are observed. Firstly, the $1/f^{\alpha}$ noise increases, so much so that it dominates for frequencies up to the order of 100 MHz; secondly, the shot noise also increases, in association with the multiplication process connected with shock ionization. The spectral density of the noise becomes:

$$\left\langle i_{SHOT}^{2}\right\rangle = 2\cdot q\cdot\left(I_{DARK}-I_{PH0}\right)\cdot G^{2}F\ \left(\text{in } A^{2}/Hz\right)$$

where $G$ is the gain, $I_{PH0}$ the primary photocurrent and $F$ the noise factor, which can be expressed as a function of the gain $G$, as is done in the case of avalanche photodiodes: $F = G^{x}$. If, as is often the case, the part linked to the dark current is negligible compared to the part linked to the photocurrent, we have:

$$\left\langle i_{SHOT}^{2}\right\rangle = 2\cdot q\cdot I_{PH0}\cdot G^{2\ \ x}\ \left(\text{in } A^{2}/Hz\right)$$

However, values of $x$ measured up to now for MSMs are somewhat higher: between 1.5 and 3, which is not desirable. Due to these noise characteristics, MSM photodetectors are not used close to breakdown.

## 5.4. Integration possibilities and conclusion

The functionalities offered by MSM photodiodes are considerable. It is with these photodetectors that the highest current cutoff frequencies, above terahertz levels, have been achieved [CHO 92], while maintaining an acceptable response coefficient. Of course, the inter-electrode distance required for such a performance is small (25 nm) and requires the use of an electronic mask, but this shows the interest in a device whose main advantages are:

– a planar configuration of the electrodes allowing, in particular, the integration of the photodiode into a coplanar microwave transmission line;

– weakly doped active layers which remain consistent with technologies based on semi-insulating substrates;

– a capacitance significantly below that of all other photodetectors with equivalent transit time characteristics;

– electrode technologies allowing the dark current to be minimized;

– a photodetection noise equal to that of a PIN photodiode.

These properties allow the MSM to be integrated without difficulty into all integrated circuit technologies based on III-V materials, whether on GaAs [BUR 91, HAR 88, ITO 84] or on InP [HOR 96], with MESFET or HEMT transistors. In general, the weakly-doped epitaxial structure of the detector is grown before the various layers required to make the transistors and the metallic deposition of the interdigitated electrode is the same as that of the transistor gates.

In conclusion, the optical properties of thin metallic lattices, which are still poorly-understood, offer interesting and original characteristics, even if the power limitations are more restrictive than those for PIN photodiodes.

## 5.5. Bibliography

[ASH 95] Ashour I.S., Harari J., Vilcot J.P., Decoster D., “High optical power nonlinear dynamic response of AlInAs/GaInAs MSM photodiode”, *IEEE Trans. on Electron Devices*, vol. 42, no. 5, p. 828-834, May 1995.

[BUR 91] Burroughes J.H., “H-Mesfet compatible GaAs/AlGaAs MSM photodetector”, *IEEE Photonics Technology Letters,* vol. 3, no. 7, p. 660-662, July 1991.

[CHE 97] Chen E., Chou S.Y., “A wavelength detector using monolithically integrated subwavelength metal-semiconductor-metal photodetectors”, *Proceedings SPIE,* vol. 3006, p. 61-67, 1997.

[CHO 92] Chou S.Y., Liu Y., Fischer P.B., “Tera-hertz GaAs metal-semiconductor-metal photodetectors with 25 nm finger spacing and width”, *Applied Physics Letters*, vol. 61, no. 4, p. 477-479, July 1992.

[GAO 94] Gao W., Khan A., Burger P.R., Hunsperger R.G., Zydzik G., O’Brian H.M., Sivco D., Cho A.Y., “InGaAs metal-semiconductor-metal photodiodes with transparent cadmium tin oxide Schottky contacts”, *Applied Physics Letters*, vol. 65, no.15, p. 1930-1932, October 1994.

[HAR 88] Harder C.S., Van Zeghbroeck B.J., Meier H., Patrick W., Vettiger P., “5.2 GHz bandwidth monolithic GaAs optoelectronic receiver”, *IEEE Electron Device Letters*, vol. 9, no. 4, p. 171-173, April 1988.

[HOR 96] Horstmann M., Schimpf K., Marso M., Fox A., Kordos P., “16 GHz bandwidth MSM photodetector and 45/85GHz ft/fmax HEMT prepared on an identical InGaAs/InP layer structure”, *Electronics Letters*, vol. 32, no. 8, p. 763-764, April 1996.

[ITO 84] Ito M., Wada. O., Nakai K., Sakurai T., “Monolithic integration of a metal-semiconductor-metal photodiode and a GaAs preamplifier”, *IEEE Electron Device Letters*, vol. 5, no. 12, p. 531-532, December 1984.

[ITO 86] Ito M., Wada O., “Low dark current GaAs metal-semiconductor-metal photodiodes using Wsix contacts”, *IEEE Journal of Quantum Electronics*, vol. 22, no. 7, p. 1073-1077, July 1986.

[KLI 94] Klingenstein M., Kuhl J., Rosenzweig J., Moglestue C., Hülsmann A., Schneider J., Kohler K., “Photocurrent gain mechanisms in metal-semiconductor-metal photodetectors”, *Solid State Electronics*, vol. 37, no. 2, p. 333-340, 1994.

[KUT 94] Kuta J.J., Van Driel H.M., Landheer D., Adams J.A., “Polarization and wavelength dependence of metal-semiconductor-metal photodetector response”, *Applied Physics Letters*, vol. 64, no. 2, p. 140-142, January 1994.

[LIM 68] LIM Y.C., MOORE R.A., "Properties of alternately charged coplanar parallel strips by conformal mapping", *IEEE Transactions on Electron Devices*, vol. 15, p. 173-180, 1968.

[MAT 96] MATIN M.A., SONG K.C., ROBINSON B.J., SIMMONS J.G., THOMPSON D.A, GOUIN F., "Very low dark current InGaP/GaAs MSM Photodetector using semi transparent and opaque contacts", *Electronics Letters*, vol. 32, no. 8, p. 766-767, April 1996.

[SCH 90] SCHUMACHER H., SOOLE J.B.D., LEBLANC H.P., BHAT R., KOZA M.A., "Noise behaviour of InAlAs/GaInAs MSM photodetectors", *Electronics Letters*, vol. 26, no. 9, p. 612-613, April 1990.

[SEO 93] SEO J.W., CANEAU C., BHAT R., ADESIDA I., "Application of indium-tin-oxide with improved transmittance at 1.3µm for MSM photodetectors", *IEEE Photonics Technology Letters*, vol. 5, no. 11, p. 1313-1315, November 1993.

[SOO 88] SOOLE J.B.D., SCHUMACHER H., ESAGUI R., KOZA M.A., BHAT R., "Waveguide integrated MSM photodetector on InP", *Electronics Letters*, vol. 24, no. 24, p. 1478-1480, April 1988.

[SOO 91] SOOLE J.B.D., SCHUMACHER H., "InGaAs metal-semiconductor-metal photodetectors for long wavelength optical communications", *IEEE Journal of Quantum Electronics*, vol. 27, no. 3, p. 737-752, March 1997.

[SZE 71] SZE S.M., COLEMAN D.J., LOYA J.R.A., "Current transport in metal-semiconductor-metal structure", *Solid State Electronics*, Pergamon Press, New York, vol. 14, p. 1209-1218, 1971.

[SZE 96] SZE S.M., *Physics of Semiconductor Devices*, John Wiley & Sons Ltd., New York, 1996.

[VIN 89] VINCHANT J.F., VILCOT J.P., LORIAUX J.L., DECOSTER D., "Monolithic integration of a thin and short metal-semiconductor-metal photodetector with a GaAlAs optical inverted rib waveguide on a GaAs semi insulating substrate", *Applied Physics Letters*, vol. 55, no. 19, p. 1966-1968, May 1989.

[WAD 88] WADA O., HAMAGUCHI H., LE BELLER L., BOISROBERT C.Y., "Noise characteristics of GaAs metal-semiconductor-metal photodiodes", *Electronics Letters*, vol. 24, no. 25, p. 1574-1575, December 1988.

[WAD 89] WADA O., NOBUHARA H., HAMAGUCHI H., MIKAWA T., TACKEUCHI A., FUJII T., "Very high speed GaInAs metal-semiconductor-metal photodiode incorporating an AlInAs/GaInAs graded superlattice", *Applied Physics Letters*, vol. 54, no. 1, p. 16-17, January 1989.

[WOH 97] WOHLMUTH W., ARAFA M., MAHAJAN A., FAY P., ABESIDA I., "Engineering the Schottky barrier heights in InGaAs metal-semiconductor-metal photodetectors", *Proceedings SPIE*, vol. 3006, p. 52-60, 1997.

[ZHA 96] ZHANG Y.G., LI A.Z., CHEN J.X., "Improved performance of InAlAs-InGaAs-InP MSM photodetectors with graded superlattice structure grown by gas source MBE", *IEEE Photonics Technology Letters*, vol. 8, no. 6, p. 830-832, June 1996.

# Chapter 6

# Ultraviolet Photodetectors

## 6.1. Introduction

The first mention of the existence of ultraviolet radiation was in 1801 when J.W. Ritter observed that certain types of chemical reactions were accelerated by a radiation which was not visible to the eye and whose wavelength was shorter than that of violet light. It was quickly shown that this particular radiation was subject to the laws of interference just like visible light (T. Young, 1804). Thus this observation, along with various others, made it possible to establish at the start of the 19th century that visible and ultraviolet radiation exhibited the same electromagnetic properties, and differed only in their wavelength.

We know today that ultraviolet radiation (UV) covers a wide spectral range which extends from the limit of the visible, around 400 nm (3.1 eV) up to the X-ray boundary at 10 nm (124 eV) (see Figure 6.1), covering a total spectral range which in terms of energy is about 80 times as large as that of the visible spectrum. It is traditionally divided into four categories:

---

Chapter written by Franck OMNES and Eva MONROY.

– UVA, covering wavelengths between 400 and 320 nm (from 3.1 to 3.87 eV);

– UVB, whose wavelengths range from 320 to 280 nm (from 3.87 to 4.43 eV);

– UVC, for wavelengths between 280 and 200 nm (from 4.43 to 6.20 eV);

– the far UV, for wavelengths between 200 and 10 nm (form 6.2 to 124 eV).

Due to its own radiation absorption properties, the terrestrial atmosphere does not allow the free propagation of light with wavelengths below 200 nm: as a result, the spectral end of the UV between 200 and 10 nm cannot really propagate except in a vacuum, and because of this it is often known as "vacuum UV" (we will use the term VUV for the rest of the chapter when referring to this particular spectral range).

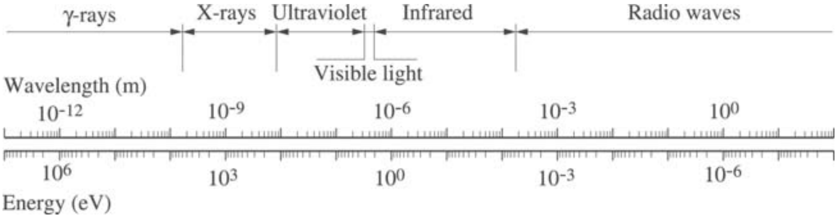


**Figure 6.1.** *Spectral ranges of the main types of electromagnetic radiation*

The fields of application for UV detectors, of which there is a great variety, mostly consist of:

– biological and chemical sensors: detection of ozone, pollutants and biological agents;

– fire detection: fire alarms, detection of missile exhausts, flame control;

– optical inter-satellite spatial communications: secure transmission of data in space;

– the calibration of emitters and UV imaging: instrumentation, measurement of solar UV, astronomical research.

In general, photodetectors suited to these areas of application must simultaneously possess a high response coefficient, a good linearity of the photocurrent as a function of the incident optical power, a large bandgap energy (in the case of semiconductors), a low level of noise and a high degree of spectral selectivity. The level of importance and the priority ascribed to these different requirements clearly depends on the type of application intended. A high response speed can be particularly useful in applications requiring rapid real-time processing of the signal, such as UV imaging.

The sun represents the most important natural source of UV radiation. It is therefore not surprising that most of the applications of UV detectors are connected with the measurement of solar radiation. The solar radiation spectrum is shown in Figure 6.2.

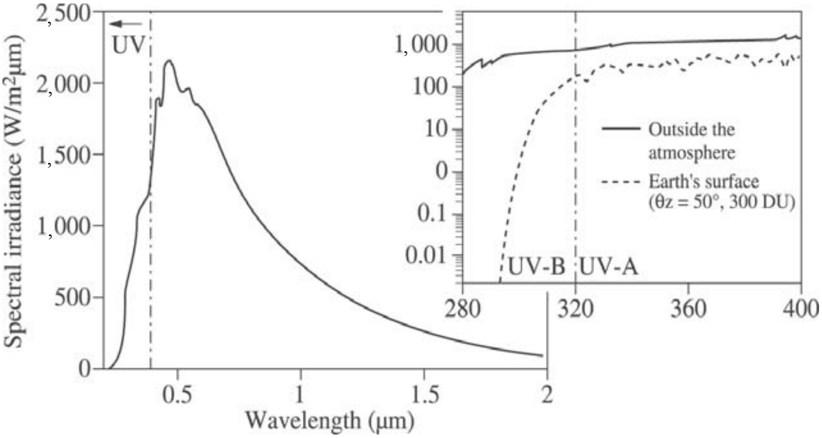


**Figure 6.2.** *Solar radiation spectrum. Inset: UVA and UVB solar irradiance, outside the Earth's atmosphere and at ground level*

For the vast majority of applications, it is helpful if the UV photodetectors are not sensitive to visible and infrared radiation. They basically operate as high-pass filters, and they are classed as *visible blind* detectors for cutoff wavelengths between 400 and 280 nm, or *solar blind*, assuming that the UVC radiation is blocked by atmospheric absorption for cutoff wavelengths typically less than or equal to 280 nm. The visible blind detectors are nowadays quite widely used for flame detection[1], and have potentially very promising growth possibilities for applications in space to optical inter-satellite telecommunications, for UV wavelengths below 280 nm which clearly cannot be detected at ground level. Outside the Earth's atmosphere, UV light represents 9% of the total luminous power emitted by the Sun. The stratospheric ozone layer completely absorbs the UVC part of the solar spectrum, and only allows UVA and UVB to be transmitted to the ground after attenuation (see the inset in Figure 6.2). It is generally considered that solar radiation of wavelengths below 280 nm does not reach the surface of the Earth. UVA and B radiation, on the other hand, are mostly responsible for the biological effects of solar radiation, since they reach the ground (see Figure 6.3).

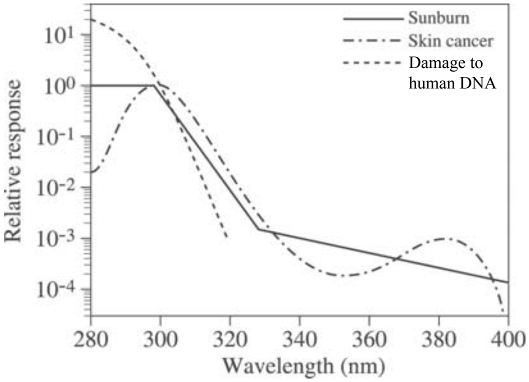


**Figure 6.3.** *Spectral dependence of the harmful effects of UVA and UVB radiation on human health*

---

[1] The applications of flame detection mostly cover the following areas: fire detector heads, forest fire detection systems, missile track detection systems, etc. Gas flames, for example, have an emission peak in the UV around a wavelength of 320 nm.

UVB radiation, being particularly energetic, is the direct cause of sunburn and skin cancers, and it has also been shown more recently that UVA radiation, although considered less dangerous, is able to penetrate the skin and has enough ionizing power to cause deep damage to living cells[2]. It has been found that UVA radiation can in particular alter the genetic structure of cells and, thus, also produce a carcinogenic effect comparable to that of UVB, particularly in the case of prolonged or repeated exposure. This represents a clear threat to human heath from UV radiation, but also allows it to play an extremely useful role in a number of different fields. Among others, UV radiations are involved in the synthesis of Vitamin D in the human body, which means that a moderate amount of exposure to sunlight is important for everybody, and the ionizing properties of these radiations make them useful for the activation of numerous chemical reactions, both organic and inorganic. Also, their bactericidal action makes them useful for many types of biological cleaning and sterilization procedures[3]. Finally, the ozone layer present in the upper atmosphere is naturally regenerated through the conversion of oxygen in the air into ozone, a reaction which is photo-assisted by UV radiation.

On the other hand, the rapid depletion of the stratospheric ozone layer, caused in particular by human activity[4], and the resultant increase in exposure of the Earth's surface to UV radiation has raised considerable concern among scientists and professionals interested in health and the environment. This has recently driven a significant

---

[2] This issue is particularly interesting since it has public health implications because of the popularity of tanning salons with the general public.

[3] The illumination of objects with high intensity UVB and C radiation can be used for this, particularly in techniques for the preservation of perishable foodstuffs, the sterilization of surgical instruments, etc. Housewives have also known since time immemorial that the drying of washing in the sun in fair weather makes it healthier, etc.

[4] The emission of aerosols and gases containing fluoro-, chloro- and bromo-carbons is largely responsible for this. Large-scale natural phenomena such as volcanic eruptions also contribute in a manner which can be very significant. One of the most significant "holes" in the ozone layer is currently situated above the continent of Antarctica, but the effect is very widespread and nowadays has a global effect.

increase in efforts to develop many specialized optoelectronic devices which in particular allow the detection and measurement of solar UV illumination. Thus, visible blind detectors have recently found a growing number of applications in UVA and B monitoring systems for the environmental study of ground-level solar illumination and UV dosimetry in fields which are mostly connected with biology, medical science or research into cosmetic products (the development of sun creams or other products which modify the photosensitivity of human skin to UV radiation, etc.). Personal solar dosimeters able to detect direct or diffuse UV light are also currently available to the general public for UVB dosimetry. The spectral response curve of these sensors reproduces as closely as possible (as we will see later) the erythema (sunburn) action spectrum on human skin.

There are a wide range of different UV photodetectors suitable for all the different applications listed above. Traditionally, photomultiplier tubes have always been used for UV detection and their use is still widespread, particularly in laboratories. However, they generally involve detection systems which are cumbersome, expensive and fragile, although they offer a sensitivity which is difficult to match with the other types of photodetectors applicable to these ranges of wavelength. The increasing need for miniaturization and reliability of these systems for the development of portable or remote applications has naturally led to the development of UV photodetectors based on semiconductors, whose active components can be photoconductors, Schottky junction photodiodes, p-i-n photodiodes, MSM photodiodes, phototransistors or avalanche photodiodes.

The use of semiconductors for UV photodetection historically began with the use of narrow bandgap materials such as silicon and certain III-V materials (GaP, GaAsP, etc.) whose bandgaps are mostly situated in the near-infrared (Si) or the red (GaAsP). In exceptional cases they may be in the yellow, as in the specific case of GaP, which represents their upper limit. The main drawback of silicon, and small bandgap semiconductors in general, is that it is impossible to use them directly for UV photodetection. Firstly, the direct exposure of such sensors to daylight clearly results in them being flooded with visible wavelengths and sometimes even infrared, so that it is simply impossible to extract information pertaining to the UV spectral component from this white light. Also, the quantum efficiency of a semiconductor photodetector (defined as the probability of electron-hole creation by the absorption of an incident photon) is greatest for energies above the bandgap of the material, but limited to a spectral range relatively close to that value. For energies much higher than the bandgap of materials, such as those that we are interested in for UV light, part of the energy of the incident light is effectively lost through heating of the semiconductor (excitation of vibration modes, or phonons in the crystal lattice), so that the quantum efficiency of the photodetector decreases in the short wavelength part of the optical spectrum. The use of semiconductor materials with a small bandgap energy for UV photodetection therefore rapidly drove the development of filtering devices, which normally use layers of selectively-absorbing material, and sometimes a phosphor-based high-pass filter which absorbs the UV light and re-emits towards the semiconductor photodetector a light whose spectral maximum is not far from the bandgap of the semiconductor material being used [GOL 99]. It still needs to be joined on to detection systems that are relatively complex and also expensive, whose behavior is also increasingly liable to drift in the long term, due to aging of the filters, resulting in a certain degree of performance loss.

A number of years ago, a new generation of UV photodetectors appeared, made from semiconductor materials with large bandgaps such as silicon carbide (SiC), diamond or gallium nitride (GaN) and AlGaN alloys. Silicon carbide, whose bandgap at ambient temperature (T = 300 K) is 2.86 eV (6H-SiC), has allowed sensitive detection of the spectral range in question. In particular, it has recently given rise to industrial applications such as the construction of flame detectors that are visible-blind, operating in the UVC range. The only issue is that even here the use of high-pass optical filters is standard in order to tune the photodetection system to the desired spectral window, meaning that this only offers a partial solution to the problem.

Thanks to the development of epitaxial crystal growth techniques, the recent use of gallium nitride (GaN), and still more recently aluminum gallium nitride ($Al_xGa_{1-x}N$, with $x$ being between 0 and 1), has helped reduce the problems listed earlier, enabling the fabrication of high performance and robust visible blind photodetectors. The direct bandgap of $Al_xGa_{1-x}N$ materials, whose size is an increasing function of the concentration of aluminum, is found at ambient temperature to extend from 3.42 eV ($\lambda$ = 362 nm) ($x$ = 0) to 62 eV ($\lambda$ = 200 nm) ($x$ = 1). Thus, the spectral range that is covered includes in particular the cutoff values for UVB (3.87 eV or $\lambda$ = 320 nm) and UVC (4.43 eV or $\lambda$ = 280 nm). It follows that it becomes possible to fabricate UV photodetectors using these materials which are free from any need for intermediate spectral filtering, opening up the possibility of simplifying photodetection systems through miniaturization, accompanied by an improved reliability of the devices. In addition, the high chemical stability of GaN and $Al_xGa_{1-x}N$ compounds, comparable to that of SiC, means that these materials are also suitable for use in the fabrication of photodetectors which are naturally resistant to extreme operating conditions (high temperatures and high values of inverse bias voltage in particular). Due to their direct bandgap at all levels of aluminum, the optoelectronic properties of $Al_xGa_{1-x}N$ alloys are also clearly superior to those of SiC, whose indirect bandgap does not so strongly favor the effects associated with photo-generation of carriers. For all these reasons, group III nitrides therefore offer a technical solution for the future, as we will see throughout the rest of this chapter, a solution which is at the same time flexible, robust and very well-suited for all visible-blind or solar-blind photodetection applications, because they allow relatively low-cost fabrication of effective and reliable UV photodetectors suitable for all the application windows listed above.

In this chapter, we will first introduce the UV-visible contrast and UV detectors based on silicon and SiC. We will then describe in detail several types of UV detectors based on nitrides: photoconductors, Schottky photodiodes, MSM photodiodes, and p-i-n photodiodes. The most sophisticated technological advances involved with the recent construction of avalanche photodiodes and phototransistors based on nitrides are also briefly described[5].

## 6.2. The UV-visible contrast

The absorption of light is the fundamental effect governing the dependence of the spectral response of a photodetector as a function of the energy of the incident photons. In direct-bandgap semiconductors, we therefore observe a dramatic transition in the value of the spectral response close to the bandgap of the semiconductor material, dropping all at once to a very low value when we drop below the bandgap. In the case of UV photodetectors, assigning an arbitrary limit to the wavelength of visible light of 400 nm, the UV-visible contrast is defined by the ratio $R_{imax}(\lambda = \lambda_G)/R_i(\lambda = 400\,\text{nm})$, where $R_{imax}$ is the maximum value of the response coefficient at the wavelength $\lambda_G$ of the bandgap and $R_i\,(\lambda = 400\,\text{nm})$ represents the value of the response coefficient at the limit of visible light.

The UV-visible contrast is mostly limited by the two following factors:

– the presence in the bandgap of the semiconductor of deep energy states which are electrically and optically active. These originate in particular from extended or localized defects in the crystal structure of the semiconductor, but can also be attributed to the presence of deep energy levels due to impurities, or deep levels caused by imperfections in the surface of the semiconductor;

– in the case of Schottky junction photodetectors with semi-transparent contacts, the effects of photo-emission of carriers by the metal.

---

[5] The work carried out by the authors on UV photodetectors, described in this chapter, was carried out with the support of the European Community, under the Environment and Climate grant number #ENV4-CT97-0539.

## 6.3. Si and SiC photodetectors for UV photodetection

Contemporary semiconductor-based commercial technical solutions proposed for UV photodetection mostly rely on Si and SiC-based technologies. Other applications, a little less widespread and well-known, use other semiconductor compounds such as GaAs, GaP and GaAsP. These are mainly used for Schottky barrier photodiodes. This particular family of devices is in particular very well-suited for the detection of VUV radiation, for which they offer a remarkable stability of operation. It is also worth mentioning the recent emergence of UV photodetectors based on diamond, which allows the construction of photoconductors with a short cutoff wavelength (225 nm) with a UV/visible contrast of $10^6$ [WHI 96]. Although diamond-based devices are potentially very interesting and well-suited for UV photodetection applications and the detection of high-energy particles, due in particular to diamond's very high resistance to the effects of irradiation, the development of this last type of photodetector has nevertheless been rather restricted by the current technical challenges of achieving epitaxial films with adequate homogenity and monocrystallinity: the best material obtained to date is a crystal mosaic with grain boundaries which have a fairly strong mismatch between one another. Clearly, this results in a high density of electrically and optically active defects which reduces the performance of the photoconductors, both in terms of response coefficient and UV/visible contrast and in terms of response time. Nevertheless, it is important to emphasize that UV photodetectors based on thin epitaxial films of diamond have recently become commercially available, which is the sign of a definite and growing interest in these devices for UV photodetection and its applications [MAI 00].

Next we will give, by way of illustration, a brief description of the main families of Si and SiC-based photodetectors.

### 6.3.1. *UV photodiodes based on silicon*

Silicon p-n junction photodiodes are currently by far the most economical and widespread commercial solution for UV photodetection. Silicon gives rise to low-cost detectors which are suitable for the vast majority of UV detection applications across a very wide spectral range extending from near-UV to VUV, and even extends to the high-energy system of soft X-rays. However, the general shape of their spectral response is not uniform, and in particular it reaches into the infrared, silicon's bandgap being 1.1 eV at ambient temperature (300 K). As a result, their use in selective UV photodetection applications requires, generally speaking, the use of absorbing filters between the light source and the photodetector, in order to select only the spectral window of direct interest. We note in passing that this often significantly increases the cost of the final detection system, the filters normally requiring complex technologies which make them very expensive.

UV photodiodes based on silicon can be subdivided into two large families: p-n junction photodiodes and metal-oxide-semiconductor photodiodes with a charge inversion layer [GOL 99, RAZ 96]. The first family of Si UV photodetectors comprises photodiodes whose p-n junction is situated at a shallow depth below the surface (typically of the order of 0.2 μm), and which are covered with a thin layer of $SiO_2$. This insulating layer plays the double role of passivation of the surface of the semiconductor and of antireflection coating. The depth of the junction is an extremely important parameter, because absorption of light is more superficial the more energy the photon has. In the case of p-n junction photodiodes, it is desirable, in order to optimize its behavior and obtain the maximum photocurrent, that as many UV-photogenerated carriers as possible reach the junction region without recombining. For the same reason, control of surface recombination effects is of fundamental importance for silicon UV photodetection. Detailed control and optimization of the $Si/SiO_2$ interface properties makes it possible to reduce the density of surface traps, while naturally producing an electric field at the surface of the silicon which limits, or even eliminates, recombination effects at the interface.

The p-n junction UV photodiode with record performance was initially developed by Korde *et al.* [CAN 89, KOR 87]. In order to form the junction, these authors used phosphor (P) diffusion onto an Si (p) substrate free from dislocations. A $SiO_2$ layer of thickness 60 nm was built on the surface of the device. Due to its thickness, this layer absorbs all radiation of wavelengths below 120 nm, which unfortunately makes this type of photodiode unusable in applications involving the detection of high energy UV (VUV in particular). Nevertheless, this type of photodiode is characterized in particular by an internal quantum efficiency of 100% in the spectral range of 350-600 nm. The internal quantum efficiency is greater than 1 in the short-wavelength UV region (<350 nm), due to secondary impact ionization effects: a high kinetic energy is imparted to the electron-hole pair photogenerated by high-energy photons, and this is sufficient to create other secondary electron-hole pairs through impacts with atoms in the crystal lattice. On the other hand, the efficiency becomes less than 1 beyond a wavelength of 600 nm, because this long-wavelength light is absorbed in the deeper layers of the material which are far from the p-n junction, creating electron-hole pairs which are unable to reach the junction since their diffusion lengths are too short. As a result, they recombine in the volume of the semiconductor and do not contribute to the photocurrent.

It is interesting to dwell for a moment on UV photodiodes with a p-n junction designed for the detection of VUV up to an energy of 124 eV. Although the base structure of the p-n junction remains the same as that which we have already described, conversely the oxide layer present on the surface is much thinner than in other types of UV Si photodiodes. Its characteristic thickness is in this case no larger than 4-5 nm, making it a factor of 10 or 20 times smaller than that normally found in more conventional applications. Transparent to short wavelengths of light, its useful volume is on the other hand reduced, in such a way that the possibilities for UV light and humidity to create high densities of recombinant traps in the oxide (which is normally one of the main sources of aging in Si UV photodiodes) is very significantly reduced compared to the normal situation. It follows that the stability of operation of these devices is very good when they are exposed to very high energy photons. The process of photogeneration produces carriers with very high kinetic energies, which multiply through secondary impact ionization. In the system of high photon energies, the internal quantum efficiency $\eta$ of Si VUV photodiodes typically follows a linear variation as a function of the energy $E$ of the photon, which can be written: $\eta = E/3.63$ [RAZ 96]. To give a better idea of the orders of magnitude, the internal quantum efficiency of a Si VUV photodiode is typically of the order of 30, for a photon energy of 124 eV. These photodiodes are commonly equipped with a bandpass filter consisting of thin metallic layers, picking out the 10-50 nm band (Al/C, Al/C/Sc, Ti, Sn, Ag, etc.) [RAZ 96].

Photodiodes with a charge inversion layer [GOL 99, RAZ 96] are similar in their architecture to metal-oxide-semiconductor structures used to build field effect transistors: they make use of effects associated with the presence of the two-dimensional charge induced through inversion at the Si/$SiO_2$ interface. Through the effects of the energy band curvature in the semiconductor, a thin deposited layer of silica on the surface of the Si (p) material leads to the creation of a charge bilayer with the opposite sign to the doping of the volume, which consists of electrons in the case we will describe. This very localized accumulation of charge results in a high electric field being established in the immediate vicinity of the semiconductor surface which, in the case of UV photodetection, offers the particular advantage that this is the exact region where the main absorption of the high-energy incident light takes place. As a result, this configuration is ideal for maximizing the quantum efficiency of the photodiodes, something which is also helped by the effectively complete absence of a “dead region” of recombination at the surface, which tended to be found in the first generations of p-n junction UV Si photodiodes. Si UV photodiodes with an inversion layer offer a high quantum efficiency in the spectral band from 250-500 nm and a cutoff wavelength of 120 nm, which is fundamentally linked to the absorption of the UV light by the oxide layer above this wavelength. It should, however, be noted that the behavior of this type of photodiode tends to change over time under UV illumination, since UV light has a tendency to alter the surface oxide layer. Additionally, the range of linearity of these photodetectors remains reduced compared to the possibilities offered by p-n junction UV Si photodiodes, an effect which seems to be mostly explained by the effects of the high electrical resistance presented by the charge inversion bilayer at the $SiO_2$/Si interface.

### 6.3.2. *SiC-based UV photodetectors*

The best UV photodetectors based on 6H-SiC are p-n junction photodiodes [RAZ 96]. The optimized structure consists of a layer of SiC (n+) with a thickness between 0.2 and 0.3 µm, sitting on a SiC (p) substrate. The doping levels are $5–10 \times 10^{18}$ $cm^{-3}$ for SiC (n+) and $5–8 \times 10^{17}$ $cm^{-3}$ for SiC (p). The structure is topped with a $SiO_2$ passivation layer on the surface. An ohmic contact based on nickel is present on the n+ side. The response coefficient of p-n SiC photodiodes made in this way lies in the range from 0.15 to 0.175 A/W at a wavelength of 270 nm, which corresponds to an internal quantum efficiency of 70 to 85%. At $\lambda$ = 200 nm, the response coefficient of these same structures is typically 0.05 A/W. The spectral response of these photodiodes contains a peak lying between $\lambda$ = 268 and $\lambda$ = 299 nm. The dark current of these devices is very weak. A dark current value of $10^{-11}$ A for a reverse-bias of –1 V has been measured for SiC photodiodes, at a temperature of 473 K [EDM 94].

Schottky junction 6H-SiC photodiodes are also currently being developed, operating in the 200–400 nm spectral range [RAZ 96]. The best performances have traditionally been obtained for Schottky devices made from SiC (p). The height of the Schottky barrier is about twice as large as that normally achieved for n-type SiC. In addition, the diffusion length for electrons is significantly larger than that for holes. Very low leakage currents can be obtained for Schottky diodes made using SiC (p). Leakage currents below 1 pA are currently measured for a reverse-bias of –10 V. However, Anikin *et al.* [ANI

92] report the fairly recent construction of a high quality Schottky photodiode in SiC (n). In this the Schottky junction is made through deposition of gold, which gives a Schottky barrier height in SiC which is typically between 1.4 and 1.63 eV. The behavior of these photodetectors is characterized by a very low leakage current, of the order of 100 pA for a reverse-bias voltage of –100 to –170 V, and a high value of the response coefficient, which reaches 150 mA/W at a wavelength of 215 nm [RAZ 96].

As we have already emphasized, one of the interests in SiC-based devices is their remarkable resilience under extreme operating conditions. This is why only a very small deterioration in their operational characteristics (spectral response and *I*(*V*) characteristics) is seen when these photodetectors are used at high temperatures, which can reach or even surpass 300°C without seeing a significant degradation of the devices.

## 6.4. UV detectors based on III-V nitrides

Among the various families of materials suitable for UV photodetection, III-V nitrides represent one of the most interesting and flexible technical solutions at present. The first work in characterizing the physical properties of polycrystalline gallium nitride (GaN) dates back more than 30 years. The development of epitaxial techniques for this material has for its part been held back for several decades, mainly due to the poor suitability of the techniques used and the absence of a substrate with a suitably matching lattice parameter, which for a long time made the development of applications impossible. Despite a strong lattice mismatch (it is 16% that of GaN) and despite a strong difference in thermal expansion coefficient compared to GaN, sapphire substrates are currently very widely employed for the growth of thin layers of III-V nitrides and the resultant applications, for which they are the substrate of choice. Organometallic vapor phase epitaxy (OMVPE) and, more recently introduced for nitride growth, molecular jet epitaxy with a plasma or gas source (MJE) currently allow the growth of thin layers of GaN and related compounds which, despite a high density of dislocations (of the order of $10^9$ to $10^{10}$ $cm^{-2}$ in GaN), exhibit structural, optical and electrical properties on sapphire which are good enough to make them suitable for many applications, mostly involving light-emitting diodes at short emission wavelengths (from green to blue, and even into the near-UV), blue lasers, field effect transistors and UV photodetectors.

### 6.4.1. *Photoconductors*

#### 6.4.1.1. *Spectral response*

Photoconductors based on AlGaN typically consist of an epitaxial layer of AlGaN on sapphire, doped with silicon, with a thickness of 1 µm, onto which two ohmic contacts are attached (see Figure 6.4).

The electrical bias circuit includes a DC voltage source connected to the photoconductor, and the device is in series with a low-value load resistance. The current induced in the device can be easily deduced from measurement of the voltage drop across the load resistance when the photoconductor is illuminated. Photoconductors based on GaN and AlGaN have a gain under continuous illumination which can reach $10^5$–$10^6$ and which strongly depends on the incident optical power. Its variation is a function which decreases in proportion to $P^{-\gamma}$ over a range of five decades, with $0.5 < \gamma < 0.95$, where $P$ is the incident optical power [MON 99a]. This behavior is independent of the wavelength of the light. The value of $\gamma$ is a decreasing function of the electrical resistance of the layer, and decreases with temperature.

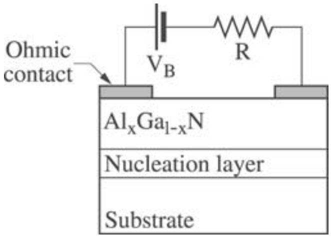


**Figure 6.4.** *Diagram of a photoconductor based on $Al_xGa_{1-x}N(Si)$*

The spectral response of photoconductors based on $Al_xGa_{1-x}N$ exhibits a cutoff at 298 nm for $x = 0.319$ nm for $x = 0.22$ and 368 nm for $x = 0.35$. The UV/visible contrast is significantly smaller than the value obtained from a simple analysis based on the absorption coefficient of these materials. The mechanism which explains the variation of the spectral response of the photoconductor must therefore involve more than just the optical absorption of the semiconductor. In addition, a sub-linear dependence of the photocurrent is observed as a function of the power of the incident light. In the model developed by Garrido *et al.* [GAR 98], the photoconductor's response coefficient $R_i$ is the combination of two terms, one due to the photogenerated free carriers, $\Delta n$, and the other due to the modification by the light of the effective conducting cross-section, $\Delta S$. The response coefficient can be written:

$$R_i = \frac{\Delta I}{P_{opt}} = \frac{qV_B\mu_e}{LP_{opt}}\left(S\Delta n + n\Delta S\right) \qquad [6.1]$$

where $q$ is the elementary charge of the electron, $\mu_e$ is the electron mobility, $L$ is the distance between the contacts, $S$ is the conductive cross-section, $V_B$ is the bias voltage and $n$ is the concentration of free carriers.

The transverse conduction cross-section in the model of Garrido *et al.* is not the same as the geometric cross-section of the device, due to the presence of space charge regions (SCRs) [HAN 98] around lattice discontinuities (dislocations that reach the surface, grain boundaries and interface). The absorption of light produces a reduction in these depletion regions, leading to a modification of the conduction cross-section. Given that $\Delta n$ can be expressed in the form [RAZ 96]:

$$\Delta_n = \eta g\tau \frac{P_{opt}}{hc/\lambda} \qquad [6.2]$$

where $\eta$ is the quantum efficiency, $g$ the photoconduction gain, $\tau$ the lifetime of the free carriers and $hc/\lambda$ the photon energy; by substituting [6.2] into [6.1], it follows that:

$$R_i = \frac{\Delta I}{P_{opt}} = \frac{qV_B\mu_e}{L}\left(\frac{\eta g\tau S}{hc/\lambda} + \frac{n}{P_{opt}}\Delta S\right) \quad [6.3]$$

The spectral variation of the first term in this equation is given by $\lambda.\eta(\lambda)$. Since the quantum efficiency is a direct function of the absorption, this first term leads to a significant UV/visible contrast. However, experiments show that the dominant mechanism in these photodetectors is represented by the second term of equation [6.3], in other words, it depends strongly on the modification to the effective transverse conduction cross-section. This mechanism can also explain the high value of the spectral response for photon energies above the bandgap energy: the energy levels responsible for the absorption of light in the visible can originate either from point defects distributed homogenously within the semiconductor, either dopants or vacancies, or in extended defects associated with lattice discontinuities. If these defects are electrically charged, they create a depletion region around them, which reduces the effective conduction cross-section of the device. Complete absorption of light through these defects can be treated as negligible, so that their presence has little effect on the absorption coefficient, and hence on the first term of equation [6.3]. However, their effect on the photoconduction spectral response is large, since the electrical charge concentrated at these lattice discontinuities is modified by the light, which alters the effective transverse conduction cross-section.

#### 6.4.1.2. *Response time*

AlGaN-based photoconductors display significant *persistent photoconductivity* (PPC), effects. This means they have a very slow transitory temporal response (several thousandths of a second) and a non-exponential tail, which is consistent with the frequency dependence of their spectral response [SHE 99].

The slow drop-off in the photocurrent can be explained by a model based on the modulation of the transverse conduction cross-section [MON 99a]. The curvature of the bands around defects and dislocations leads to a spatial separation of electrons and holes, which can be the cause of the PPC. When the light is extinguished, the electrons are inclined to recombine, but they must first cross the potential barrier separating them from their recombination sites. The height of this barrier depends on the charge present at the defect, so that it will change over time as recombination occurs. Through this mechanism, the SCR returns slowly to its original extent in the dark, and as a result the resistance of the device also slowly returns, in a non-exponential manner, towards its dark value. The dominant defect type may be that associated with dislocations at the AlGaN-air interface or the AlGaN-substrate interface [SHE 99]. However, it is difficult to discriminate between these two regions, and their relative significance will depend on the crystalline quality of the material.

Quantitatively, the temporal response of photoconductors can be calculated in the following manner, if $\Delta Q_{SS}$ is the increase in charge at the defects, connected to the variation in the width of the SCR, $\Delta x_{SCR}$. $\Delta Q_{SS}$ can be expressed in the form:

$$\Delta QSS = qN_D \Delta x_{SCR} = \sqrt{2N_D \varepsilon_S \varepsilon_0}\left(\sqrt{\Psi_0} - \sqrt{\Psi_{il}}\right) \qquad [6.4]$$

where $N_D$ is the level of n-type doping, $\varepsilon_S$ and $\varepsilon_0$ are the relative dielectric constant of the semiconductor and the vacuum permittivity respectively, and $\Psi_0$ and $\Psi_{il}$ are the heights of the barrier around the defect, in the dark and under illumination respectively. It follows that the response coefficient of the photoconductor is given by:

$$R_i = \frac{A_{opt} V_B \mu_e}{P_{opt} L^2} \Delta QSS \qquad [6.5]$$

where $A_{opt}$ is the optical surface area of the device.

In the dark, $\Delta Q_{SS}$ originates from the current crossing the potential barriers around the defects. Including the thermionic emission current $J_{th}$, as well as the tunneling current $J_{tunnel}$, it follows that:

$$\frac{dQ_{SS}}{dt} = J_{th} + J_{tunnel} \qquad [6.6]$$

The thermionic and tunneling currents can be expressed respectively in the following forms:

$$J_{th} = A^* T^2 \exp\left(\frac{-q\Psi_0}{kT}\right)\left[\exp\left(\frac{qV_{ph}}{kT}\right) - 1\right] \qquad [6.7]$$

$$J_{tunnel} = A^* T^2 \frac{\sqrt{\pi E_{00}\left(\Psi_0 - V_{ph}\right)}}{\frac{kT}{q}\cosh\left(\frac{qE_{00}}{kT}\right)} \exp\left(\frac{-q\left(\Psi_0 - V_{ph}\right)}{E_0}\right)\left[1 - \exp\left(\frac{qV_{ph}}{kT}\right)\right] \qquad [6.8]$$

with $E_{00} = (h/4\pi).(N_D/m^*_e)$ and $E_0 = coth(qE_{00}/kT)$. $V_{ph}$ is the voltage induced by the light. The drop in voltage across the barrier is $V_{ph}(t)$, which can be expressed as:

$$V_{ph}\left(t\right) = \Psi_0 - \frac{Q^2 SS\left(t\right)}{2q\varepsilon_S\varepsilon_0 N_D} \qquad [6.9]$$

To illustrate this, Figure 6.5 shows the temporal response of AlGaN photoconductors, both measured, and calculated by numerical solutions of equations [6.4] to [6.9].

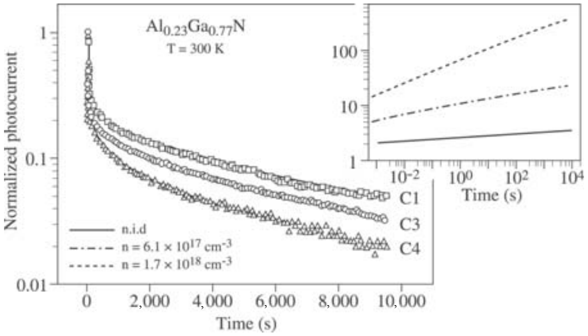


**Figure 6.5.** *Decrease in the normalized photocurrent in $Al_{0.23}Ga_{0.77}N$ photoconductors, along with numerical simulation using equations [6.4]–[6.10]*

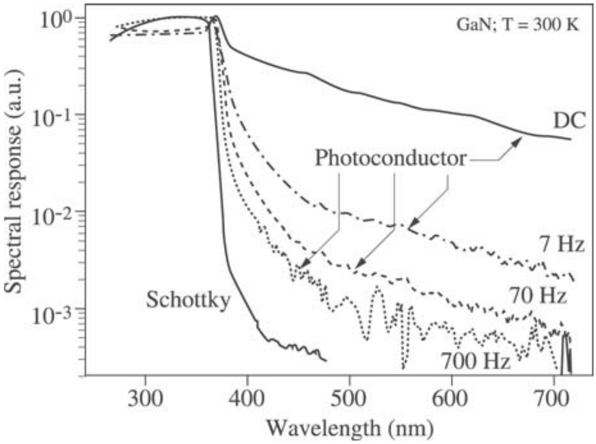


**Figure 6.6.** *Normalized spectral response of a photoconductor, measured at different modulation frequencies of the incident optical signal. The spectral response of a Schottky photodiode made from the same batch is also shown for comparison*

### 6.4.1.3. *Effects of frequency modulation of the incident optical signal*

The effect of frequency on the spectral response of photoconductors is shown in Figure 6.6. When the frequency increases, the SCRs associated with the defects do not have time to respond. The result is that the cutoff slope becomes sharper and the spectrum of the response coefficient tends to approach that of a photovoltaic detector. As a result, the UV/visible contrast is much better at higher frequencies. The mechanism at the root of the response of III-V nitride-based photoconductors therefore makes them of little interest for applications which requires speed or a high UV/visible contrast.

## 6.4.2. *Schottky barrier photodiodes based on AlGaN*

### 6.4.2.1. *Electrical properties*

Work published on Schottky barrier photodiodes involves planar or vertical structures (Figure 6.7), which in theory both give extremely similar operating characteristics. Although the vertical structure is particularly interesting for the construction of fast devices with a high response coefficient, the current limitations in III-V nitride technologies (connected in particular with damage to the material during the engraving of the *mesa*) result in a performance degradation, particularly in terms of the bandwidth and the noise level.

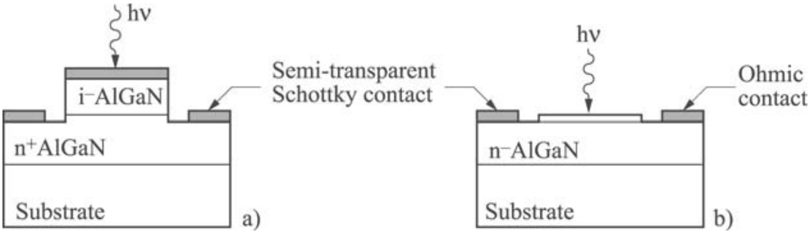


**Figure 6.7.** *Schematic diagram of AlGaN Schottky photodiodes in (a) vertical and (b) planar configurations*

Figure 6.8 shows the typical current-voltage (*I-V*) characteristics of planar Schottky photodiodes. In GaN-based devices, the ideality factor is of the order of 1.2, with a series resistance of 20-50 Ω and a leakage resistance greater than 1 GΩ. The leakage current increases as a function of the fraction of aluminum, and the ideality factor also increases and can reach values close to 4. Consequently, the measured *I-V* characteristic of AlGaN-based diodes does not make it possible to deduce precise and reliable information about the height of the Schottky barrier, due to the high values of the ideality factor. The variation of the capacitance as a function of the bias voltage leads to barrier heights of 0.84 eV and 1.02 eV respectively for semi-transparent Ni and Au contacts on GaN. The height of the barrier as a function of the molar fraction of aluminum for Schottky diodes with nickel and gold contacts is shown in Figure 6.9.

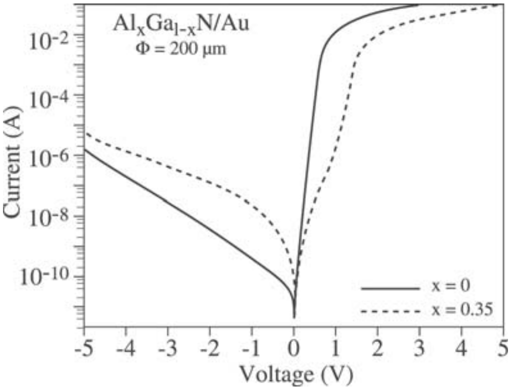


**Figure 6.8.** *Typical current-voltage characteristics for Schottky diodes made of* Al$_x$Ga$_{1-x}$N *on sapphire*

### 6.4.2.2. *Spectral response*

In contrast to photoconductors, the photocurrent of Schottky barrier devices depends linearly on the incident light power (inset in Figure 6.10). This linearity is independent of the size of the diode, the nature of the metal used to fabricate the Schottky diode and the excitation wavelength, for values situated both above and below the bandgap energy of the semiconductor. The decrease in the response coefficient observed when the fraction $x$ of aluminum increases results from two effects. Firstly, at a constant quantum efficiency, the response coefficient decreases at the shortest wavelengths. Also, the penetration depth of the light is larger than the SCR, and so the response coefficient depends on the diffusion length of the minority charge carriers, which decreases with the concentration of aluminum.

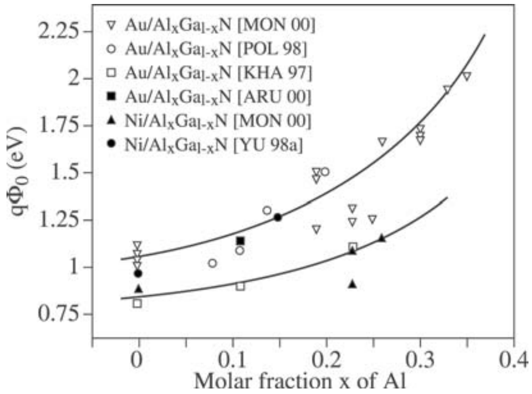


**Figure 6.9.** *Variation of the height of the Schottky barrier as a function of the molar fraction of Al*

Figure 6.10 shows that the UV/visible contrast is more than three orders of magnitude. The cutoff wavelength develops from 362 nm to 293 nm. The steepness of the cutoff slope of Schottky photodetectors shows that the only limitation in the UV/visible contrast is that caused by absorption by deep defects, in contrast to the results obtained for photoconductors. The spectral response is flat for wavelengths smaller than those of the bandgap, which is an advantage that Schottky photodiodes have over other types of photodetectors, particularly p-n junction photodetectors. Since the SCR lies just below the surface of the semiconductor, there is no decrease in the quantum efficiency at short wavelengths as is normally seen in p-n and p-i-n photodiodes, where the photogenerated carriers must diffuse across the layer above before reaching the SCR at the level of the junction. In junction devices based on GaN, this problem is accentuated by the high density of dopants required to achieve p-type conductivity, which results in a very thin SCR and a material which has a very low mobility. From this point of view, the choice of Schottky photodiodes based on $Al_xGa_{1-x}N$ has more potential than p-n or p-i-n devices for constructing wideband UV detectors.

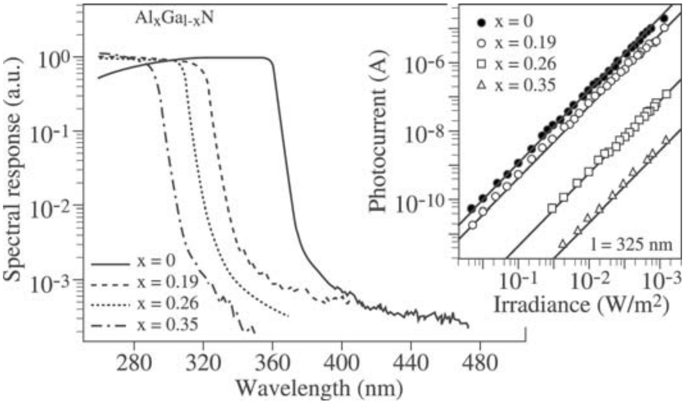


**Figure 6.10.** *Normalized spectral response at ambient temperature of Schottky photodiodes for different values of the molar fraction of Al. Inset:: variation of the photocurrent as a function of the incident light power*

The maximum value of the response coefficient obtained in AlGaN Schottky photodiodes is limited by the optical transmission of the semi-transparent Schottky contact, and by the diffusion length $L_h$ of the minority carriers in this material. For radiation with an energy close to the bandgap of the semiconductor, a certain proportion of the photons may cross the SCR without absorption and penetrate deep into the volume of the material. In order to be collected, the electron-hole pairs must therefore diffuse towards the SCR at the surface, and a low value of $L_h$ results in a decrease in the response coefficient. $L_h$ is a function of the mobility and the lifetime of the minority charge carriers, and is an increasing function of the quality of the material. It follows that the use of thick AlGaN layers (thickness greater than 1 μm) is preferred because they allow a reduction in the density of dislocations, and hence an increase in the diffusion length and the response coefficient; it also results in a reduction of the dark current.

### 6.4.2.3. *Response time*

The response time of these photodetectors is mainly limited by the *RC* time constant of the device, where *C* is the sum of the internal capacitance of the diode and the load capacitance, and *R* is the sum of the load resistance and the series resistance of the device. Response times around 50 nanoseconds are typical for diodes of diameters between 200 and 300 µm. The capacitance of the depletion region is given by the expression:

$$C_{SCR} = A\sqrt{\frac{qN_D\varepsilon_S\varepsilon_0}{2\left(\Psi_0 + V\right)}} \qquad [6.10]$$

where $A$ is the area of the Schottky contact, $N_D$ is the density of dopants, $\varepsilon_s$ is the dielectric constant of the semiconductor, $\varepsilon_0$ is the vacuum permittivity, $\Psi_0$ is the height of the Schottky barrier seen from the semiconductor side, and $V$ is the bias voltage. Since the series resistance does not depend on the bias voltage, the response time of the device decreases in proportion to $(\Psi_0+V)^{-1/2}$.

For planar devices, $R_S$ and $C_{SCR}$ are proportional to $(\mu N_D)^{-1}$ and $N_D{}^{1/2}$ respectively, where $\mu$ is the electronic mobility. The maximum bandwidth of these devices is therefore proportional to $\mu N_D{}^{1/2}$. Since the mobility increases as the doping level falls, there exists an optimum doping level, of the order of $10^{18}$ cm$^{-3}$, for which the response time is minimized. This effect is accentuated for AlGaN alloys with high concentrations of aluminum, because of the difficulty in achieving good quality ohmic contacts.

### 6.4.2.4. *Noise and detectivity*

The *1/f* noise dominates at low frequencies. Its spectral noise power density, $S_n$, satisfies the relationship:

$$S_n = s_0\frac{I_{dark}^{\alpha}}{f^{\gamma}} \qquad [6.11]$$

where $I_{dark}$ is the dark current, $f$ is the frequency, and $s_0$, $\gamma$ and $\alpha$ are dimensionless parameters whose values are close to 1 and 2 respectively. As well as the 1/*f* noise, the contribution of the shot noise must be included. The noise equivalent power (NEP) can therefore be written in the form:

$$NEP = \frac{\sqrt{\left\langle i_{shot}^2 \right\rangle + \left\langle i_{1/f}^2 \right\rangle}}{R_i} \qquad [6.12]$$

where $i_{shot}$ is the shot noise current, $i_{1/f}$ is the $1/f$ noise current and $R_i$ is the response coefficient of the device. The value of $<i_{shot}^2>$ is given by:

$$\left\langle i_{shot}^2 \right\rangle = 2qI_{dark} \cdot \Delta f \qquad [6.13]$$

where $\Delta f$ is the bandwidth of the photodiode, and $I_{dark}$ represents the dark current. NEPs of 8 pW/Hz$^{1/2}$ and 41 pW/Hz$^{1/2}$ have been obtained for Schottky GaN/Au and $Al_{0.22}Ga_{0.78}N$/Au photodiodes respectively, at a reverse-bias voltage of –2 V. The corresponding detectivity values are 6.1 × $10^9$ W$^{-1}$H$^{-1/2}$ cm and 1.2 × $10^9$ W$^{-1}$Hz$^{1/2}$ cm respectively.

### 6.4.2.5. *Schottky barrier photodiodes made from GaN grown using epitaxial lateral overgrowth (ELOG)*

Despite these promising characteristics, heteroepitaxial growth of GaN results in a high density of dislocations (~ $10^8$–$10^9$ cm$^{-3}$), which limits the UV/visible contrast in GaN-based photodetectors. The recent development of the technique of epitaxial lateral overgrowth (ELOG) has reduced this density of dislocations by at least two orders of magnitude [NAM 97, USU 97].

Schottky barrier photodiodes made with epitaxial GaN layers grown using the ELOG technique offer an order of magnitude increase in the UV/visible contrast compared to GaN-based devices grown on sapphire using standard methods. The dark current, whose value lies below 1 nA/cm$^2$ for a reverse-bias voltage of –1 V, is significantly smaller than that of Schottky photodiodes based on GaN/sapphire. As the response time of the device is limited by the time constant *RC*, the weaker residual doping in these epitaxial layers makes it possible to obtain bandwidths greater than 30 MHz, 12 MHz and 8 MHz for devices with diameters equal to 200 µm, 400 µm and 600 µm respectively. A detectivity as high as $5 \times 10^{11}$ $W^{-1}Hz^{1/2}$ cm has been measured for GaN ELOG photodiodes with a diameter of 400 µm, with a bias voltage of –3.4 V.

### 6.4.2.6. *Application of AlGaN Schottky barrier photodetectors for the simulation of the biological effects of UV light*

UV radiation produces a wide range of biological effects [DEG 94, MCK 87]:

– pigmentation: maximum effect at $\lambda = 360$–$400$ nm;

– erythema (sunburn): maximum sensitivity at $\lambda < 297$ nm;

– synthesis of vitamins D2 and D3: $\lambda = 249 - 315$ nm. Maximum yield at 290 nm;

– damage to plants: $\lambda < 317$ nm;

– bactericidal action: $\lambda = 210$–$310$ nm. Maximum effects at 254 nm;

– carcinogenic effects: UVB and UVC. Maximum at 310 nm;

– DNA damage: $\lambda < 320$ nm. Effect grows rapidly as the wavelength decreases.

Simple, accurate, reliable and low-cost devices are hence required to evaluate the biological effects of UV radiation. UV detectors with a wide spectral band have been developed to study the effects of UVA and UVB on the skin [MCK 87]. Commercial UV dosimeters use photodiodes with a small bandgap (Si, GaAs, GaP) combined with a series of filters inserted into the path of the incident light. Muñoz *et al.* [MUN 00] showed that the spectral response of AlGaN Schottky photodiodes can match the UVA and B erythema action spectrum, with the appropriate selection of an appropriate molar fraction and carefully selected growth conditions (Figure 6.11). In this way, the response coefficients of these devices can give direct information on the biological effects of UV light. For this application, AlGaN photodetectors therefore offer a flexible solution with a low cost and minimum footprint, since they require neither filters nor temperature stabilization.

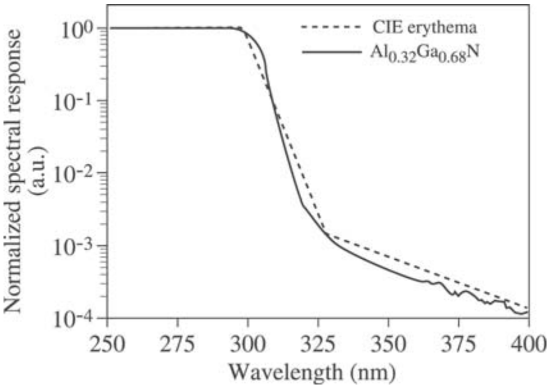


**Figure 6.11.** *Normalized spectral response of an AlGaN Schottky photodiode compared with the sunburn action spectrum (international standard set by the International Commission on Illumination (CIE))*

### 6.4.3. *MSM photodiodes*

MSM photodiodes consist of two interdigitated Schottky contacts deposited onto a planar surface (see Chapter 5). The main advantage of these structures, based on AlGaN, is their ease of fabrication. The typical structure of such an MSM diode involves two interdigitated Ni-Au Schottky contacts deposited on an epitaxial layer of AlGaN, not intentionally doped, of micron-scale thickness. Their dark current is very weak and follows the thermionic emission model with a value of the Richardson constant equal to 26 $Acm^{-2}K^{-2}$ and a barrier height of 1.04 eV for GaN.

Their spectral response is relatively flat above the bandgap, with a steep cutoff slope which shifts towards shorter wavelengths as the proportion of aluminum increases. The photocurrent varies in a quasi-linear manner with the incident optical power. The UV-visible contrast varies from around three orders of magnitude for the weakest bias voltages (typically less than 5V) to four or five orders for the highest bias voltages. This accentuation of the contrast with the bias voltage can be explained by the appearance of a gain. At present, there is no satisfactory explanation available for the origin of this gain. Despite the existence of this gain, the dynamic behavior of these photodetectors is for the most part limited by the *RC* time constant and the transit time of the photocarriers [MON 99c].

Their NEP lies between a few pW/Hz$^{1/2}$ and a few tens of pW/Hz$^{1/2}$ depending on the geometry of the device and the bias voltage [MON 99b].

### 6.4.4. *p-n and p-i-n photodiodes*

The performance limitations of the first GaN p-n junction photodiodes were mainly caused by a high resistivity of the p-type layer, connected with difficulties associated with magnesium p doping, and the high resistivity of the associated ohmic contact. Progress made in understanding the growth conditions of the materials, and in the technology of the ohmic contacts, then made it possible to make p-n junction GaN photodiodes with a fast temporal response (response time of the order of 100 ns) and a very low noise level. The performance of these devices was further improved by the introduction of a region that was not intentionally doped (p-i-n photodiodes) and through the use of AlGaN/GaN heterostructures. Arrays of heterostructure $Al_{0.20}Ga_{0.80}N$/GaN p-i-n photodiodes with dimensions 32 × 32 and 128 × 128 pixels [BRO 00] have been made with good performance in terms of their response coefficient and detectivity. p-i-n photodiodes in $Al_xGa_{1-x}N$, sometimes with high levels of aluminum (up to 76% [SAN 00]), have also been described. The optimization of these last devices is however limited by the intrinsic difficulties in achieving a high level of *p* doping in $Al_xGa_{1-x}N$ materials, which can only be made with difficulty to exceed $10^{17}$ cm$^{-3}$, and by the associated ohmic contact whose resistivity remains high. The behavior of the device is thus doubly penalized: firstly, the response time of a $Al_xGa_{1-x}N$ p-i-n photodetector cannot be made small enough, because the series resistance component due to the resistance of the electrical contacts remains large, and the device also contributes a significant level of noise associated with an ohmic contact of insufficient quality, which leads to detectivity values which are notably smaller than those that can be obtained with p-n and p-i-n photodiodes based on GaN.

### 6.4.4.1. *Static characteristics*

In p-n and p-i-n photodiodes based on $Al_xGa_{1-x}N$, the linearity of the photocurrent as a function of the incident light power extends over a scale of more than five orders of magnitude [WAL 99]. The response coefficient of p-n and p-i-n homojunction GaN photodiodes typically lies in the range 0.10–0.15 A/W, which corresponds to external quantum efficiencies of between 30 and 44%. These results are improved by the use of a $Al_xGa_{1-x}N$ layer in the illuminated region (which can be n- or p-doped), in order to avoid the loss of carriers through diffusion. A response coefficient of 0.2 A/W at $\lambda = 365$ nm has been measured for a sapphire-$Al_{0.28}Ga_{0.72}N$(n)-GaN(i)-GaN(p) heterostructure illuminated by the rear face through the sapphire [YAN 98]. A maximum in the response coefficient of 57 mA/W was reported at $\lambda = 287$ nm, which corresponds to a quantum efficiency of 25% [PAR 99]. An increase in this coefficient is observed in all cases as a function of the bias voltage [WAL 99]. This result confirms that the sensitivity is limited by the diffusion length of the charge carriers. As the bias voltage increases, the width of the SCR increases, such that the carriers created through photo-ionization processes in the regions further from the junction are collected.

A UV/visible contrast of more than four orders of magnitude is observed for all these devices. The small diffusion length of the carriers can affect the spectral response above the bandgap of p-i-n photodiodes based on $Al_xGa_{1-x}N$ [XU 97]. The absorption coefficient increases as the wavelength decreases, meaning that the absorption of the light occurs closer to the surface. The result is that electron-hole pairs are generated above the junction in the upper layer and must diffuse to the junction in order to be collected. A short diffusion length therefore leads to a decrease in the response coefficient for shorter wavelengths, which is all the more marked when the surface layers are thicker (of the order of 0.2 µm). p-i-n detectors based on group III nitrides are characterized by a sharp cutoff at long wavelengths, whose energy value is very closely linked to the bandgap of the material forming the active region. For certain applications, however, an additional cutoff wavelength is also required at short wavelengths, defining a spectral window for UV detection. In their theoretical study, Pulfrey and Nener [PUL 98] investigated the possibility of using p-i-n $Al_xGa_{1-x}N$/GaN heterostructures as bandpass UV detectors, with a high energy cutoff wavelength at the bandgap of $Al_xGa_{1-x}N$. In this case, the $Al_xGa_{1-x}N$ layer must be thick enough not only to accentuate the absolute response peak, but also to act as an integral high-pass filter. Devices equipped with a 1 µm thick $Al_{0.10}Ga_{0.90}N$ layer [KRI 98] have demonstrated that it is possible to achieve rejection on a level of more than two orders of magnitude in the short wavelength region.

### 6.4.4.2. *Response time*

GaN-based p-n and p-i-n photodiodes have a temporal response which is in general limited by the *RC* time constant, and show an exponential fall in the photocurrent. Response times of the order of 25 nm have been measured with zero bias voltage on devices whose optical surface area is 200 × 200 $\mu m^2$ [OSI 97]. This response time drops to 10 ns for a bias voltage of –6 V, which can be explained by a reduction in the capacitance of the junction associated with the applied voltage. Faster response times can be obtained with AlGaN/GaN heterojunctions [XU 97]. However, the response time of these devices is often non-exponential, with characteristic fall times that are more complicated than a simple *RC* effect [CHE 95, VAN 97, WAL 98]. This behavior can be attributed to the presence of a defect in the semiconductor, with an activation energy of 99 meV and a high concentration (of the order of $10^{18}$ $cm^{-3}$) which can probably be explained by substituted magnesium atoms.

GaN p-i-n photodiodes with response times below 1 ns have been constructed [CAR 99]. This result is obtained by increasing the thickness of the "i" layer to 1 µm, which makes it possible to build a device which has a low capacitance. In this case, the factor limiting the temporal response is the transit time of the carriers. The only drawback of this particular type of device is its low response coefficient, which is around 0.03 A/W. Such devices are of course ideal for the detection of rapid signals.

#### 6.4.4.3. *Noise*

Under illumination, the spectral density of noise under reverse-bias is well described by the equation $S_n=2.q.I_{ph}$ (shot noise). A very low normalized noise equivalent power, of the order of 7 fW/Hz$^{1/2}$, has been obtained in p-n junction photodiodes for a reverse-bias voltage of –3 V [OSI 97]. Since these devices have an optical area of 200 × 200 µm, their corresponding normalized detectivity is high, of the order of $3\times10^{12}$ W$^{-1}$Hz$^{1/2}$cm.

#### 6.4.4.4. *p-i-n photodiodes on laterally-grown GaN*

$Al_{0.33}Ga_{0.77}N$ p-i-n photodiodes on GaN have been built using lateral growth, by means of a $SiO_2$ mask consisting of 35 µm widebands with 5 µm openings [PAR 99]. Diode mesas, with dimensions of 10 × 100 µm$^2$, have been made in the regions of the material corresponding to the "wings" of the lateral growth (outside the coalescence region, and outside the openings in the $SiO_2$ mask). Additional devices with mesa dimensions of 30 × 100 µm$^2$ have also been built by way of comparison. These particular devices also cover the coalescence region situated at the junction of the "wings" of lateral growth. Finally, devices with a similar form with square 300 × 300 µm$^2$ mesas have been made from GaN on standard sapphire. Low dark current densities, of the order of 10 nA/cm$^2$, have been measured for the smallest devices lying on the "wings" of lateral growth. This leakage current density is an order of magnitude smaller than that measured for devices which include the coalescence regions, and more than 6 orders of magnitude smaller than that of devices made from GaN on standard sapphire. This low value of the leakage current can be linked to the strong reduction in the density of dislocations, since these defects lead to a strengthening of the transport mechanism, assisted by the tunneling effect [YU 98b]. In addition, the spectral response has a more abrupt cutoff in the case of devices built on the "wings" of the growth, in a similar way both when the coalescence region is covered and when it is not. All these responses are consistent with similar observations made for Schottky photodiodes made with laterally-grown GaN [MON 99d].

#### 6.4.4.5. *Avalanche photodiodes based on GaN*

The mechanism of avalanche photodiodes is analyzed in detail in Chapter 3. The unusual requirement for GaN-based materials is that very high fields are required to achieve multiplication, as calculations of the ionization coefficients for electrons and holes in GaN show [KOL 95, OGU 97]. Although a homogenous avalanche multiplication of carriers has recently been shown for PIN photodiodes based on GaN [MCI 00], with a response time between 100 and 500 ns, the high density of defects in these materials makes it difficult to achieve a homogenous multiplication across the whole active area of the device. Also, to overcome the limitations in noise due to ionization coefficients for electrons and holes that are too close together, Ruden [RUD 99] proposed a hybrid structure consisting of an AlGaN/Si heterostructure, with multiplication taking place in the silicon.

### **6.4.5. *Phototransistors***

The operation of phototransistors is described in Chapter 4. As far as GaN-based devices are concerned, the literature refers to bipolar and field effect transistors [KHA 95, YAN 98].

#### 6.4.5.1. *Bipolar phototransistors*

In a bipolar transistor, the reverse-biased base-collector junction acts like a p-n junction photodiode, and its photocurrent is amplified by the transistor effect. In its most up to date configuration, the base contact is not connected (floating-base mode of operation). Yang *et al.* [YAN 98] report the fabrication of a GaN(n)/GaN(p)/GaN(i)/$Al_{0.20}Ga_{0.80}N$(n) heterojunction phototransistor. Electrical contacts were present on the collector and the emitter, and the base was left floating. The UV light enters the device through the sapphire substrate, crosses the $Al_{0.20}Ga_{0.80}N$(n) layer, and is absorbed in the GaN(i) layer. The photogenerated electron-hole pairs are separated by the electric field in the i region, and the electrons and holes travel to the collector and the base respectively. The accumulation of holes in the floating base increases the injection of electrons from the emitter, which results in a current gain. A gain of more than $10^5$ has been demonstrated.

GaN-based phototransistors exhibit a sub-linear behavior of the photocurrent as a function of the incident light power as well as showing persistent photoconductivity (PPC), analogous to that also recently observed in photoconductors. In their normal operating regime, holes ought to recombine in the base with electrons injected from the emitter. However, the holes are trapped at defects which reduces their recombination rate and leads to persistent charge effects. Yang *et al.* [YAN 98] showed that this recombination is increased by the application of a bias to the detector in order to force the holes to enter the emitter. The voltage impulse acts as a sort of electrical "reset to zero".

In order to avoid persistent effects, the phototransistor is subjected to a "reset to zero" voltage impulse before each measurement of the response coefficient (a voltage between 7 and 10 V is used for this), and the photocurrent is then measured at a bias voltage of 3 to 4 V. The gain is strongly dependent on the frequency, so that the gain-bandwidth product remains constant. An excellent UV/visible contrast of 8 orders of magnitude was obtained [YAN 98].

### 6.4.5.2. *Field effect phototransistors*

AlGaN(n)/GaN(n)/GaN(i) heterostructure field effect transistors have also been used in photodetection mode [KHA 95], with illumination through the substrate (sapphire). The photogenerated holes travel towards the channel, where they are rapidly drawn towards the drain by the high electric field. These devices display high response coefficients which can reach 3,000 A/W, with a steep cutoff and a temporal response of the order of 200 µs.

## 6.5. Conclusion

The development of visible blind ultraviolet photodetectors is at present largely motivated by the large number of possible applications in fields as diverse as UV astronomy, studies of the ozone layer, motor control, missile plume detection, flame detection, secure communications in space, etc. With these goals in mind, photoconductors, Schottky photodiodes, MSM photodiodes, p-n and p-i-n photodetectors and phototransistors based on $Al_xGa_{1-x}N$ materials have been developed in recent years, complementing older widely-available technologies for Si and SiC-based UV photodiodes. It is important to emphasize that this range of nitride-based sensors is currently bringing a new degree of improved flexibility to UV detector fabrication in general, due mainly to the large bandgaps available with $Al_xGa_{1-x}N$ materials which make it possible to build detection systems which are entirely free from intermediate spectral filters, with performances that are at least equivalent in terms of reliability and robustness to other currently-available technologies that are themselves already very high performance.

$Al_xGa_{1-x}N$ photoconductors offer a high internal gain at ambient temperatures (~100 for $P_{opt} = 1$ W/m$^2$). However, this gain is associated with a sub-linear behavior as a function of the incident optical power, a low UV/visible contrast and significant and undesirable persistent photoconductivity effects. These drawbacks make them less suitable for most applications. The use of a modulated optical signal makes it possible to considerably improve the linearity and UV/visible contrast. However, these photoconductors lose much of their advantage in this configuration, above all because the response coefficients of the devices is significantly reduced in such a configuration, and the measurement system becomes much more complex and as a result much more expensive.

Schottky photodiodes have the advantage of a uniform and flat spectral response when they are excited at energies above the bandgap, independent of the incident optical power and the temperature. They also offer a steep cutoff slope, with a UV/visible contrast which is of the order of $10^3$. Their temporal response is limited by the *RC* product, with minimum response times typically of the order of nanoseconds. This clearly demonstrates that these devices are suited to environmental applications and the construction of UV photodetector arrays.

MSM photodiodes with very low dark currents have been built. These devices have a photocurrent which varies linearly as a function of the incident light power and makes it possible to achieve a UV/visible contrast of $10^4$. Given their significant bandwidth and their low noise level, these devices can be a judicious choice to meet the needs of detection in the domain of visible blind optical communications. Furthermore, the possibility could be considered of integrating $Al_xGa_{1-x}N$ MSM photodiodes with field effect transistors based on group III nitrides in the context of the monolithic fabrication of optical receivers.

p-n and p-i-n junction photodiodes are linear as a function of the optical power, and can achieve a UV/visible contrast of $10^4$. However, their response time is normally limited by effects associated with the presence of trapping levels linked to magnesium, to which a degree of degradation in the spectral response can also be attributed. The minimum cutoff wavelength is currently limited by the difficulty of achieving a high p doping in $Al_xGa_{1-x}N$ alloys at high concentrations of aluminum. In spite of very promising published results, improvements are still currently needed in the p doping of these materials, in order to increase the performance of these devices, as well as their reliability.

Phototransistors offer the combination of a very high gain with a record UV/visible contrast of $10^8$. As a result, these devices are promising for applications requiring a high spectral resolution, in spite of the narrowness of their bandwidth which makes them poorly-suited to high-frequency operation.

To summarize, current results confirm that $Al_xGa_{1-x}N$ alloys are the best choice in terms of semiconductors suited to UV photodetection. However, the performance of these devices is still currently limited by the high density of defects present in the heteroepitaxial layers. In consequence, very significant improvements in photodetector devices are anticipated from work currently in progress in the fields of epitaxial growth and the technology of group III nitrides, such as the technique of lateral growth.

## 6.6. Bibliography

[ANI 92] ANIKIN M.M., ANDREEV A.N., PYATKO S.N., SAVKINA N.S., STRELCHUK A.M., SYRKIN A.L., CHELNOKOV V.E., "UV photodetectors in 6HSiC", *Sensors and Actuators: A. Physical* 33 (1-2), pp. 91-93, 1992.

[ARU 00] ARULKUMARAN S., EGAWA T., ZHAO G.Y., ISHIKAWA H., JIMBO T., UMENO M., "Electrical characteristics of Schottky contacts on GaN and Al. 0:11. Ga. 0:89. N", *Japanese Journal of Applied Physics*. 39, L351, 2000.

[BRO 00] BROWN J.D., BONEY J., MATTHEWS J., SRINIVASAN P., SCHETZINA J.F., NOHAVA T., YANG W., KRISHNANKUTTY S., "UV-specific digital camera based on a 128x128 focal plane array of GaNAlGaN PIN photodiodes", *MRS Internet Journal of Nitride Semiconductor Research,* 5, article 6, 2000.

[CAN 89] CANFIELD L.R., KERNER J., KORDE R., "Stability and quantum efficiency performance of silicon photodiode detectors in the far ultraviolet", *Applied Optics,* 28, 3940, 1989.

[CAR 99] CARRANO J.C., LI T., EITING C.J., DUPUIS R.D., CAMPBELL J.C., "Very fast metal-semiconductor-metal ultraviolet photodetectors on GaN with submicron finger width", *J. of Electronic Materials,* 28, 325, 1999.

[CHE 95] CHEN Q., KHAN M.A., SUN C.J., YANG J.W., "Visible-blind ultraviolet photodetectors based on GaNp-n junctions", *Electronics Lett.* 31, 1995.

[DEG 94] DE GRUJI F.R., VAN DER LEUN J.C., "Estimate of wavelength dependency of ultraviolet carcinogenesis and its relevance to the risk assessment of a stratospheric ozone depletion", *Health Physics*, 67, 314, 1994.

[EDM 94] EDMOND J.A., KONG H.S., CARTER JR C.H.," Blue LEDs, photodiodes and high-temperature rectifiers in 6H–SiC", *Physica B* 185, 453, 1993.

[GAR 98] GARRIDO J.A., MONROY E., IZPURA I., MUÑOZ E., "Photoconductive gain modelling of GaN photodetectors", *Semiconductor Science and Technology,* 13, 563, 1998.

[GOL 99] GOLDBERG Y.A., “Semiconductor near-ultraviolet photoelectronics”, *Semiconductor Science and Technology,* 14, R41, 1999.

[HAN 98] HANSEN P.J., STRAUSSER Y.E, ERICKSON A.N., TARSA E.J., KOZODOY P., BRAZEL E.G., IBBETSON J.P., MISHRA U., NARAYAMURTI V., DEN BAARS S.P., SPECK J.S., “Scanning capacitance microscopy imaging of threading dislocations in GaN films grown on (0001) sapphire by metalorganic chemical vapor deposition”, *Applied Physics Lett.*, 72, 2247, 1998.

[KHA 95] KHAN M.A., SHUR M.S., CHEN Q., KUZNIA J.N., SUN C.J., “Gate photodetector based on GaN/AlGaN heterostructure field effect transistor”, *Electronics Lett.*, 31, 398, 1995.

[KHA 97] KHAN M.R.H., NAKAYAMA H., DETCHPROHM T., HIRAMATSU K., SAWAKI N., “A study on barrier height of Au-$Al_xGa_{1-x}N$ Schottky diodes in the range 0”, *Solid State Electronics*, 41, 287, 1997.

[KOL 95] KOLNIK J., H.OGUZMAN I., BRENNAN K.F., WANG R., RUDEN P.P., WANG Y., “Electronic transport studies of bulk Zincblende and Wurzite phases of GaN based on an ensemble Monte Carlo calculation including a full zone band structure”, *J. of Applied Physics*, 78, 1033, 1995.

[KOR 87] KORDE R., GEIST J., “Quantum efficiency stability of silicon photodiodes”, *Applied Optics*, 26, 5284, 1987.

[KRI 98] KRISHNANKUTTY S., YANG W., NOHAVA T., RUDEN P.P., “Fabrication and characterization of GaN/AlGaN ultraviolet-band heterojunction photodiodes”, *MRS Internet Journal of Nitride Semiconductor Research*, 3, article 7, 1998.

[KUK 98] KUKSENKOV D.W., TEMKIN H., OSINSKY A., GASKA R., KHAN M.A., “Origin of conductivity and low-frequency noise in reverse-biased GaN p-n junction”, *Applied Physics Lett.*, 72, (11), 1365-1367, 1998.

[MAI 00] MAINWOOD A., “Recent developments of diamond detectors for particles and UV radiation”, *Semiconductor Science and Technology*, 15, R55, 2000.

[MCI 00] MCINTOSH K.A., MOLNAR R.J., MAHONEY L.J., MOLVAR K.M., EFREMOW N. JR., VERGHESE S., “Ultraviolet photon counting with GaN avalanche photodiodes”, *Applied Physics Lett.*, 76, 3938, 2000.

[MCK 87] MCKINLEY A.F., DIFFEY B.L., “A reference action spectrum for ultraviolet induced erythema in human skin”, *CIE Journal* 6, 17, 1987.

[MON 99a] MONROY E., CALLE F., MUNOZ E., OMNES F., BEAUMONT B., GIBART P., “Visible-blindness in photoconductive and photovoltaic AlGaN ultraviolet detectors”, *J. of Electronic Materials*, 28, 238, 1999.

[MON 99b] MONROY E., CALLE F., MUNOZ E., OMNES F., GIBART P., "High UV/visible contrast photodiodes based on epitaxial lateral overgrown GaN layers", *Electronics Lett.*, 35, 240, 1999.

[MON 99c] MONROY E., CALLE F., MUNOZ E., OMNES F., "AlGaN metal-semiconductor-metal photodiodes", *Applied Physics Lett.*, 74, 3401,1999.

[MON 99d] MONROY E., CALLE F., MUNOZ E., BEAUMONT B., OMNES F., GIBART P., "High UV/visible contrast photodiodes based on epitaxial lateral overgrown GaN layers", *Electronics Lett.*, 35, 1488, 1999.

[MON 00] MONROY E., CALLE F., PAU J.L., SANCHEZ F.J., MUÑOZ E., OMNES F., BEAUMONT B., GIBART P., "Analysis and modeling of $Al_xGa_{1-x}N$-based Schottky barrier photodiodes", *J. of Applied Physics*, 88, 2081, 2000.

[MUN 00] MUÑOZ E., MONROY E., CALLE F., OMNES F., GIBART P., "AlGaN photodiodes for monitoring the solar UV radiation", *Journal of Geophysical Research*, 105, 4865, 2000.

[NAM 97] NAM O.H., BREMSER M.D., ZHELEVA T., DAVIS R.F., "Lateral epitaxy of low defect density GaN layers via organometallic vapor phase epitaxy", *Applied Physics Lett.*, 71, 2638, 1997.

[OGU 97] OGUZMAN I.H., BELLOTTI E., BRENNAN K.F., KOLNIK J., WANG R., RUDEN P.P., "Theory of hole initiated impact ionization in bulk zincblende and Wurtzite GaN", *J. of Applied Physics*, 81, 7827, 1997.

[OSI 97] OSINSKY A., GANGOPADHYAY S., GASKA R., WILLIAMS B., KHAN M.A., KUKSENKOV D., TEMKIN H., "Low noise *p*-π-*n* ultraviolet photodetectors", *Appl. Phys. Lett.*, 71, 2334, 1997.

[PAR 99] PARISH G., KELLER S., KOZODOY P., IBBETSON J.P., MARCHAND H., FINI P.T., FLEISCHER S.B., DEN BAARS S.P., MISHRA U.K., TARSA E.J., "High-performance (Al,Ga)N-based solar-blind ultraviolet p-i-n detectors on laterally epitaxially overgrown GaN", *Applied Physics Lett.* 75, 247, 1999.

[POL 98] POLYAKOV A.Y., SMIRNOV N.B., USIKOV A.S., GOVORKOV A.V., PUSHNIY B., "Studies of the origin of the yellow luminescence band, the nature of nonradiative recombination and the origin of persistent photoconductivity in *n*-GaN films", *Solid State Electronics*, 42, 1959, 1998.

[PUL 98] PULFREY D.L., NENER B.D., "Suggestions for the development of GaN-based photodiodes", *Solid State Electronics*, 42, 1731, 1998.

[RAZ 96] RAZEGHI M., ROGALSKI A., "Semiconductor ultraviolet detectors", *J. of Applied Physics*, 79, 7433, 1996.

[RHO 88] RHODERICK E.H., WILLIAMS R.H., *Metal-Semiconductor Contacts*, Clarendon Press, Oxford, 1988.

[RUD 99] RUDEN P.P., "Materials theory based modeling for III-nitride devices", *SPIE Proceedings*, 3629, 174, 1999.

[SAN 00] SANDVIK P., WALKER D., KUNG P., K.MI, SHAHEDIPOUR F., KUMAR V., ZHANG X., J.DIAZ, C.JELEN, M.RAZEGHI, "Solar-blind AlxGa1-xN p-i-n photodetectors grown on LEO and non-LEO GaN", *SPIE Proceedings*, 3948, 265, 2000.

[SHE 99] SHEN B., YANG K., ZHAN L., CHEN Z., ZHOU Y., CHEN P., ZHAN R., HUANG Z., ZHOU H., ZHENG Y., "Study of photocurrent properties of GaN ultraviolet photoconductor grown on 6H-SiC substrate", *Japanese Journal of Applied Physics*, 38, 767, 1999.

[USU 97] USUI A., SUNAKAWA H., SAKAI A., YAMAGUCHI A., "Thick GaN epitaxial growth with low dislocation density by hydride vapor phase epitaxy", *Japanese Journal of Applied Physics*, 36, L899, 1997.

[VAN 97] VAN HOVE J.M., HICKMAN R., KLAASSEN J.J., CHOW P.P., RUDEN P.P., "Ultraviolet-sensitive, visible blind GaN photodiodes fabricated by molecular beam epitaxy", *Applied Physics Lett.*, 70, 2282, 1997.

[WAL 98] WALKER D., SAXLER A., KUNG P., ZHANG X., HAMILTON M., DIAZ J., RAZEGHI M., "Visible blind GaN p-i-n photodiodes", *Applied Physics Lett.*, 72, 3303, 1998.

[WAL 99] WALKER D., KUNG P., SANDVIK P., WU J., HAMILTON M., LEE I.H., DIAZ J., RAZEGHI M., "AlxGa1-xN p-i-n photodiodes on sapphire substrates", *Proc. SPIE*, 3629, 193, 1999.

[WHI 96] WHITFIELD M.D., MCKEAG R.D., CHAN S.S.M., JACKMAN R.B., "A thin film diamond photodiode for ultra-violet light detection", *Applied Physics Lett.*, 68, 290,1996.

[XU 97] XU G.Y., SALVADOR A., KIM W., FAN Z., LU C., TANG H., MORKOC H., SMITH G., ESTES M., GOLDENBERG B., YANG W., KRISHNANKUTTY S., "High speed, low noise ultraviolet photodetectors based on GaN *p-i-n* and AlGaN (*p*)-GaN(*i*)-GaN(*n*) structures", *Applied Physics Lett.*, 71, 2154, 1997.

[YAN 98] YANG W., NOHAVA T., KRISHNANKUTTY S., TORREANO R., MCPHERSON S., MARSH H., "Back-illuminated GaN/AlGaN heterojunction photodiodes with high quantum efficiency and low noise", *Applied Physics Lett.*, 73, 978,1998.

[YU 98a] YU L.S., QIAO D.J., XING Q.J., LAU S.S., BOUTROS K.S., REDWING J.M., "Analysis and modeling of $Al_xGa_{1-x}N$-based Schottky barrier photodiodes", *Applied Physics Lett.*, 73, 238, 1998.

[YU 98b] YU L.S., LIU Q.Z., XING Q.J., QIAO D.J., LAU S.S., REDWING J.D., "The role of the tunneling component in the current-voltage characteristics of metal-GaN Schottky diodes", *J. of Applied Physics*, 84, 2099, 1998.

# Chapter 7

# Noise in Photodiodes and Photoreceiver Systems

In systems transmitting information by means of photons, the recovery of a useful signal must necessarily involve an optical-electrical (photon-electron) transformation. This transformation, which makes use of an optoelectronic system, is, as with any system, fundamentally limited by the background noise. The aim of this chapter is to discuss the origin of this noise and also to study various means of minimizing it in order to obtain the best signal-to-noise ratio, a value which influences the quality of the optical detection. In section 7.1 we will recall the basics of signal analysis, leading to the frequency analysis of the power of random signals. After recalling the different sorts of noise associated with the most relevant physical processes (sections7.2 and 7.3), in section 7.4 we will present the main results concerned with the analysis of noisy electrical circuits. Section 7.5 is dedicated to the study of noise in photodetectors, which we choose to limit to PIN and avalanche photodiodes. Section 7.6 discusses the problems of background noise associated with a photodiode-amplifier pair forming the photoreceiver, and lays out the methods of obtaining the best signal-to-noise ratio. Finally, in order to

---

Chapter written by Robert ALABEDRA and Dominique RIGAUD.

illustrate these techniques, we present the complete background noise calculation for a PIN-FET photoreceiver.

## 7.1. Mathematical tools for noise

Before discussing the treatment of random signals, we will recall a summary of the treatment of known signals in an attempt to justify a frequency approach as compared to a temporal approach for the treatment of the power or the energy of random signals.

### 7.1.1. *Known signals with finite energy or power*

All systems are governed by a differential equation in the time domain, and we will characterize them by their impulse response *h*(*t*). The output signal *s*(*t*) for an input signal *i*(*t*) can be calculated in the time domain either by solving this differential equation or by calculating the convolution of *h(t)* with *i*(*t*):

$$s(t) = i(t) * h(t) = \int_{\infty}^{+\infty} e(\tau) h(t-\tau) d\tau = \int_{\infty}^{+\infty} h(\tau) e(t-\tau) d\tau \qquad [7.1]$$

The signal *s*(*t*) has the same dimensions as *i*(*t*) which implies that *h*(*t*) has the dimensions of inverse time. Thus, in the case of an *R*–*C* circuit, the impulse response is:

$$h(t) = \frac{1}{RC} \exp\left(-\frac{t}{RC}\right) \text{in } s^{\;1} \qquad [7.2]$$

In order to avoid calculating the convolution, we use Fourier analysis which transforms the convolution in the time domain to a simple product in the frequency domain, and we can recover the original *s*(*t*) through the inverse transform. On the other hand, if we want to analyze the same system from the point of view of energy or power, where phase information no longer has any meaning, we will then use the autocorrelation function $R_s(\tau)$ of the signal. The relationship between $R_i(\tau)$, the autocorrelation of the input signal, and $R_s(\tau)$, that of the output signal, is:

$$R_s(\tau) = h^*(-\tau) * h(\tau) * Re(\tau) \qquad [7.3]$$

where $h^*(\tau)$ is the complex conjugate of $h(\tau)$.

This expression will only occasionally be used, given its double convolution product. We will repeat this same analysis in terms of frequency, using the concept of spectral density and the Wiener-Khinchin theorem which shows that the spectral density $G_x(f)$ of a signal $x(t)$ is the Fourier transform of the autocorrelation $R_x(\tau)$ of $x(t)$ such that:

$$G_x(f) = T \cdot F \cdot \left[R_x(\tau)\right] = \int_{\infty}^{+\infty} R_x(\tau)\exp(-j2\pi f\tau)\,d\tau \qquad [7.4]$$

and its inverse

$$R_x(\tau) = T \cdot F^{\;1} \cdot \left[G_x(f)\right] = \int_{\infty}^{+\infty} G_x(f)\exp(j2\pi f\tau)\,df \qquad [7.5]$$

If we take $\tau = 0$, which is the maximum of the autocorrelation function, $R_x(0)$ represents the energy or the power of the signal. We find:

$$R_x(0) = \int_{\infty}^{+\infty} G_x(f)\,df \qquad [7.6]$$

$G_x(f)$ represents the energy or power distribution of the signal in the frequency domain. We take the Fourier transform of equations [7.1] to [7.3]:

$$F \cdot T \cdot \left[ R_s(\tau) = h^*(-\tau) * h(\tau) * Re(\tau) \right] \qquad [7.7]$$

which can be written as:

$$G_s(f) = H(f) H^*(f) G_i(f) \qquad [7.8]$$

or:

$$G_s(f) = \left| H(f) \right|^2 G_i(f) \qquad [7.9]$$

and which is much more easy to use than equation [7.3].

What we need to do now is to work out how to transfer what we have already set out for deterministic signals to the random signals which represent the sources of noise.

### 7.1.2. *Random signals and background noise*

In the case of random signals, we have no choice but to treat them in terms of power or energy, and of course without any consideration of their phase. As we did for the known signals in section 7.1.1, we will use the autocorrelation function, the spectral power density and the same Wiener-Khinchin theorem. This will lead us to define a class of random signals which, fortunately for us, will cover the majority of practical situations. This class is characterized by ergodicity and time-stability.

DEFINITION OF ERGODICITY – The average resulting from many simultaneous experiments is the same as that obtained by the temporal mean across a single experiment.

DEFINITION OF TIME STABILITY – In its strict sense, this means that the mean value is independent of time and the autocorrelation function only depends on $\tau = t_1 - t_2$.

In this case, we obtain for the random signal $x(t)$:

$$R_x(\tau) = E\left[x(t)x(t-\tau)\right] \qquad [7.10]$$

where $E$ is the expected value associated with the probability density of the process under study (Poisson, Gaussian systems, etc.). Under these conditions the autocorrelation function is practically analogous to that for known signals which have the same properties. Now, using this formalism, we can treat the random signals appearing in the expression in terms of energy or power, and calculate the signal-to-noise ratio which characterizes the receiving quality.

## 7.2. Fundamental noise sources

Here we are interested in noise generated by solid devices in electronic circuits. As we have seen in section 7.1, we will remain in the frequency domain in order to avoid convolving the input and output autocorrelation functions. We will use the spectral densities as our measurement of scale, and in order to recover a power or an energy we will integrate these spectral densities over the useful bandwidth. These spectral densities are expressed in $A^2$/Hz or $V^2$/Hz.

### 7.2.1. *Thermal noise*

At the terminals of a metallic resistance of value $R$ at temperature $T$ in degrees Kelvin, the Brownian motion of electrons produces a randomly-varying voltage $v(t)$ at its open-circuit terminals, with a mean value of zero. Bearing in mind the large number of electrons (of the order of $10^{22}$ cm$^{-3}$), the voltage $v(t)$ obeys a centered Gaussian law. For frequencies below the inverse of the mean time between two collisions ($\tau_r \approx 10^{-12}$ $s$), the spectral density associated with this random signal can be written in the basis of positive frequencies as:

$$S_v(f) = 4\,kTR \qquad \left(V^2/Hz\right) \tag{7.11}$$

where $k$ is the Boltzmann constant ($1.38 \times 10^{-23}$ J.K$^{-1}$). We note that this spectral density is independent of the frequency (white noise).

### 7.2.2. *Shot noise*

Shot noise stems from the discrete nature of electrons and their random emission and collection over time. Thus, it takes the form of a random variable governed by the Poisson distribution law. Frequency domain analysis of this process leads to a spectral noise current density (short circuit current) of:

$$S_i(f) = 2q\,I \qquad \left(A^2/Hz\right) \tag{7.12}$$

where $q$ is the elementary charge ($q = 1.6 \times 10^{-19}$ $C$). Here also we find a white noise as long as we ignore the transit time of the carriers. Shot noise is always encountered when the current is caused by carriers emitted from a reservoir or across a potential energy barrier greater than $kT$. This is the case for reverse-biased junctions. For PIN photodiodes, for example, we have:

$$S_{i\,total}(f) = 2\,q\left[I_{dark} - I_{ph}\right] \tag{7.13}$$

where $I_{dark}$ is the intrinsic dark current of the photodiode and $I_{ph}$ the photocurrent due to absorbed photons and representing the useful information. For modern devices, $I_{dark} << I_{ph}$ and the noise is mainly due to the photocurrent. In contrast, for detection of low light intensities, the dark current will make a significant contribution to the signal-to-noise ratio.

### 7.2.3. *Multiplication noise*

In a similar way to vacuum photomultipliers, avalanche photodiodes with internal gain are based on the process of multiplication through the impact of carriers in the space charge region of the photodiode. In the case of vacuum photomultipliers, there is only one type of carrier: electrons locally multiplied from one dynode to the next. If we ignore the multiplication inertia and the transit time of the electrons, the noise generated by this photomultiplier is amplified shot noise. The variance in the total multiplication factor is:

$$\overline{M^2} = \overline{M}^2 \tag{7.14}$$

and for large enough $M$ we have:

$$S_i(f) = 2qI_{pho}\,\overline{M}^2 \tag{7.15}$$

where $\overline{\mathrm{M}}$ is the mean value of the multiplication factor and $I_{pho}$ is the primary photocurrent. Setting:

$$\overline{M} = \frac{I_{total}}{I_{pho}} \tag{7.16}$$

we have:

$$S_i(f) = 2qI_{pho}\overline{M}^2 = 2qI_{pho}\frac{I_{total}^2}{I_{pho}^2} \tag{7.17}$$

in other words, we have a noise proportional to the total current squared. In the case of semiconductor avalanche photodiodes, the multiplication process involves both types of carrier across the whole multiplication region, i.e. in the high electric field region in the space charge region. Numerous theoretical approaches have been proposed in the literature. To summarize these approaches, we recall that the two types of carrier are characterized by their ionization coefficient (in cm$^{-1}$) $\alpha$ for electrons and $\beta$ for holes, and that these depend on the physical characteristics of the semiconductor materials. The important parameter is the ratio $k = \dfrac{\alpha}{\beta}$. The ideal material would be one in which only one of the two types of carrier could be ionized: either electrons, in which case $\beta = 0$ and $k = \infty$, or holes, in which case $\alpha = 0$ and $k = 0$. The noise characteristics would then approach that of a vacuum photomultiplier.

For avalanche photodiodes, the expression for the spectral noise density can be written in the general form:

$$S_i\left(f\right) = 2qI_{pho}\,\overline{M}^{x} \qquad [7.18]$$

where $x$ lies between 2 and 4.

Taking the noise of a vacuum photomultiplier as our reference, we define the excess noise factor $F$ as:

$$F\left(\overline{M}\right) = \frac{\overline{M^2}}{\overline{M}^2} = \frac{\overline{M}^{x}}{\overline{M}^2} = \overline{M}^{x\quad 2} \qquad [7.19]$$

in other words as the ratio of the variance of the multiplication process in the semiconductor to the variance of the same process in the vacuum tube. In the case of an initial injection of electrons, and for values of $k$ close to 1, we can approximate $F\left(\overline{M}\right)$ by the expression:

$$F\left(\overline{M}\right) = k\overline{M} + \left(2 - \overline{M}^{\quad 1}\right)\left(1 - k\right) \qquad [7.20]$$

In particular, for $k = 1$ ($\alpha = \beta$), as in germanium for example, we have $F\left(\overline{M}\right) = \overline{M}$ and:

$$S_i\left(f\right) = 2qI_{pho}\overline{M}^{3} \qquad [7.21]$$

For silicon avalanche photodiodes, $\alpha$ is 50 times larger than $\beta$. For multiplication initiated by electrons, we arrive at the expression:

$$S_i\left(f\right) = 2qI_{pho}\overline{M}^{2\ 2} \qquad [7.22]$$

which leads to an excess noise factor of:

$$F\left(\overline{M}\right) = \overline{M}^{0\ 2} \qquad [7.23]$$

In the limiting case where only electrons are responsible for multiplication ($\beta = 0$), the spectral density must approach that of a vacuum photomultiplier and:

$$F\left(\overline{M}\right) = \overline{M}^{0} = 1 \qquad [7.24]$$

However, multiplication initiated by the less ionizing carriers is of no practical interest. In this case, we find excess noise factors $F(M)$ of the order of $\overline{M}^{2}$, which gives:

$$S_i\left(f\right) = 2qI_{pho}\overline{M}^{2}F\left(\overline{M}\right) = 2qI_{pho}\overline{M}^{4} \qquad [7.25]$$

These results are all depicted in Figure 7.1.

In order to achieve behavior for solid avalanche devices approaching that of vacuum photomultipliers, several devices have been conceived based on engineering of the band structure with the help of heterojunctions. The aim was to obtain impact ionization by electrons only. Despite the sophistication of the technologies used, the results did not live up to the performances initially promised, because the interfaces resulted in an additional $1/f$ noise.

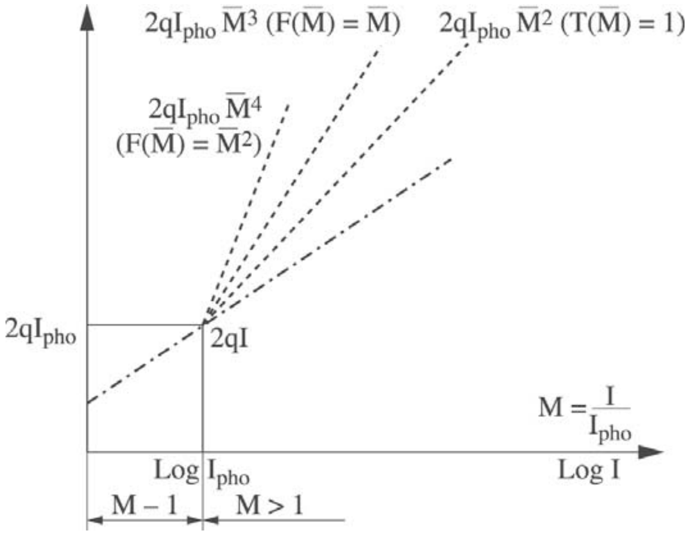


**Figure 7.1.** *Characteristic plots of the variations in the spectral noise density for multiplication in avalanche photodiodes as a function of the multiplied current*

## 7.3. Excess noise

In addition to the sources of white noise, there are two other sources of noise which depend on frequency. These are generation-recombination (GR) noise and scintillation noise, which is referred to as the 1/*f* noise. These two sources of noise are referred to together as excess noise, as compared to the fundamental sources which are thermal noise and shot noise.

### 7.3.1. *Generation-recombination noise*

This noise originates in the fluctuations of the number of carriers, linked to the presence of different trapping states in the bandgap of semiconductors. The spectral density associated with these fluctuations can be written as:

$$S_N(f) = \overline{\Delta N^2}\,\frac{4\tau}{1+\omega^2\tau^2} \quad [7.26]$$

where $\overline{\Delta N^2}$ is the second order moment of the random variable $N$ which represents the number of carriers under consideration. $\tau$ is the time constant of the generation-recombination process taking place. This spectrum, of Lorentzian form, displays a plateau at low frequencies followed by a $1/f^2$ dropoff at high frequencies. These fluctuations lead to current fluctuations proportional to the square of the mean current:

$$S_i(f) = \frac{\overline{\Delta N^2}}{\overline{N}^2} I^2 \frac{4\tau}{1+\omega^2\tau^2} \quad \left(A^2/Hz\right) \quad [7.27]$$

Generally, the observed time constants lie between $10^{-3}$ and $10^{6}$ $s$. However, it should be noted that in some specific cases, these values can be smaller.

### 7.3.2. *1/f noise*

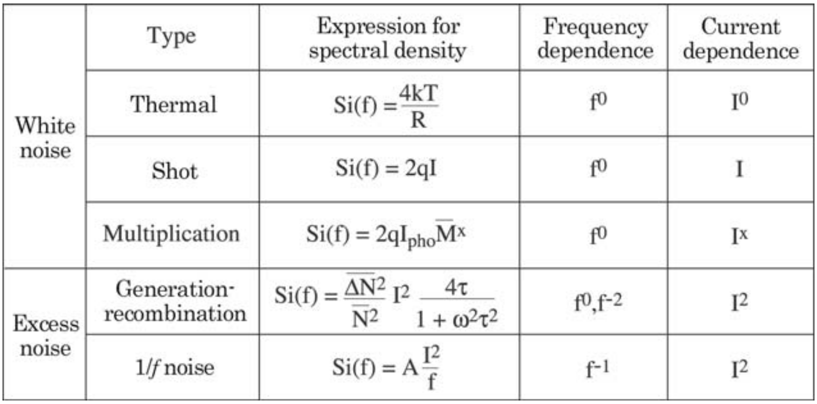
The origin of this noise is poorly understood. Its spectral density takes the form:

$$S_i(f) = \frac{A}{f^{\gamma}} I^2 \quad \left(A^2/Hz\right) \quad [7.28]$$

where $\gamma$ usually takes a value close to 1, which leads to the term $1/f$ noise, and $A$ is an experimentally-determined constant which varies from one device to another. We note that the spectral density varies as the square of the bias current. It is not possible to calculate the autocorrelation function of the process by the use of the inverse Fourier transform of $S_i(f)$ because the integral diverges. There are two interpretations of the origin of this noise. The first was proposed by

McWhorter [MCW 57] who attributed the origin of this noise to the surfaces and interfaces of the material where a continuum of trapping states for carriers can be found (surface effects), a situation which produces an overlapping series of generation-recombination events. The second interpretation is due to Hooge [HOO 76], who attributes this noise source to a volume effect in materials, due to fluctuations in the mobility of the carriers. Whichever of these is correct, it is almost certain that the amplitude of this 1/*f* noise is affected by the quality of the device and its fabrication technology. The frequency $f_c$ is defined as that for which the spectral density of the 1/*f* noise is equal to that of the fundamental noise (shot noise or thermal noise).

For the rest of this chapter, the main sources of noise that we will take into account when evaluating the noise characteristics of a photoreceiver are thermal noise, shot noise and its potential amplification through multiplication processes, and the 1/*f* noise when the optical signal requires a treatment that includes low frequencies. In practical terms, Table 7.1 summarizes the expressions for the spectral density for the different sources of noise involved with devices and systems.



| | Type | Expression for spectral density | Frequency dependence | Current dependence |
|---|---|---|---|---|
| White noise | Thermal | $Si(f) = \frac{4kT}{R}$ | $f^0$ | $I^0$ |
| | Shot | $Si(f) = 2qI$ | $f^0$ | $I$ |
| | Multiplication | $Si(f) = 2qI_{pho}\overline{M}^x$ | $f^0$ | $I^x$ |
| Excess noise | Generation-recombination | $Si(f) = \frac{\overline{\Delta N^2}}{\overline{N}^2} I^2 \frac{4\tau}{1 + \omega^2\tau^2}$ | $f^0, f^{-2}$ | $I^2$ |
| | 1/*f* noise | $Si(f) = A\frac{I^2}{f}$ | $f^{-1}$ | $I^2$ |

**Table 7.1.** *Expressions and characteristics of the main noise sources*

## 7.4. Analysis of noise electrical circuits

As is the case for deterministic signals, the behavior of a system is described in terms of its equivalent circuit, with which a certain number of noise sources must be associated. These sources are represented using the Thevenin or Norton formalisms. In contrast to generators of deterministic signals, noise generators will be characterized by the spectral density associated with the quantities that they represent. In this way we can avoid all the problems of temporal analysis while retaining the concept of impendence or admittance as traditionally defined in circuit analysis. The concepts presented below will be used later in the evaluation of the noise and the signal-to-noise ratio of photoreceiver systems.

### 7.4.1. *Representation of noise in bipoles*

A real bipole (for example a photodetector) is thus represented by a random voltage generator (Thevenin), which has a spectral density $S_v(f)$ in series with its impedance $Z(f)$, which is assumed not to be noisy; or by a random current generator (Norton) with spectral density $S_i(f)$ in parallel with its admittance $Y(f)=\dfrac{1}{Z(f)}$.

These representations are illustrated in terms of electrical circuits in Figure 7.2.

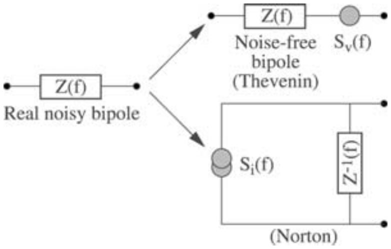


**Figure 7.2.** *Representation of a noisy bipole using a frequency spectrum approach for random processes*

Henceforth, all noise generators will be marked with hatching in order to distinguish them from normal signal generators. The equivalence of these two representations can be expressed in the following equation:

$$S_V(f) = |Z(f)|^2 S_i(f) \qquad [7.29]$$

which connects the two spectral densities that we have used.

Other parameters can be used to characterize the noise of a bipole. If we take the example of shot noise, we can define the equivalent noise current $I_{eq}$ as:

$$I_{eq} = \frac{S_i(f)}{2q} \qquad [7.30]$$

In the case of thermal noise, we can define:

– either the equivalent resistance $R_{eq}$ of the noise:

$$R_{eq} = \frac{S_v(f)}{4kT} \qquad [7.31]$$

– or the equivalent conductance $G_{eq}$ of the noise:

$$G_{eq} = \frac{S_i(f)}{4kT} \qquad [7.32]$$

– or the equivalent temperature $T_{eq}$ of the noise:

$$T_{eq} = \frac{S_v(f)}{4kRe\{Z(f)\}} = \frac{S_i(f)}{4kRe\{Y(f)\}} \qquad [7.33]$$

The representations of the noise given in Figure 7.2 make it possible to directly treat series or parallel combinations of noisy bipoles: two bipoles in series are equivalent to a single bipole whose spectral density of voltage noise can be written:

$$Sv_t(f) = Sv_1(f) + Sv_2(f) + 2\,\mathrm{Re}\left\{Sv_1v_2(f)\right\} \quad [7.34]$$

where $Sv_1v_2(f)$ represents the cross-spectrum which encapsulates the statistical dependence of the two random processes associated with the two noise sources. This dependence can be represented in terms of a correlation coefficient defined by:

$$C(f) = \frac{Sv_1v_2(f)}{\sqrt{Sv_1(f)Sv_2(f)}} \quad [7.35]$$

with $0 \leq |C| \leq 1$. Thus, $Sv_T(f)$ can then be written:

$$Sv_T(f) = Sv_1(f) + Sv_2(f) + 2\,\mathrm{Re}\left\{C(f)\sqrt{Sv_1(f)Sv_2(f)}\right\} \quad [7.36]$$

In the normal case where the two sources of noise are uncorrelated ($C = 0$), the resultant spectral density is then simply the sum of the spectral densities associated with each bipole. In a similar manner, in the case of a parallel combination of two bipoles, we have, using the parallel representation:

$$Si_T(f) = Si_1(f) + Si_2(f) + 2\,\mathrm{Re}\left\{C(f)\sqrt{Si_1(f)Si_2(f)}\right\} \quad [7.37]$$

### 7.4.2. *Representation of noise in quadripoles*

For quadripoles (the case of amplifiers placed downstream of photodetectors), the noise is generally represented by a minimum of two partially-correlated noise sources. These sources, of voltage and/or current, can be placed at the input and/or output of the quadripole. This leads to six equivalent representations, each one associated with one of six matrices describing the properties of the quadripole. In practice two of these representations are the most commonly used (see Figure 7.3).

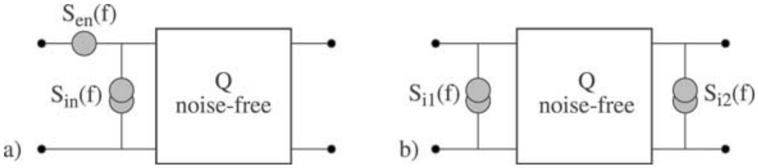


**Figure 7.3.** *The two most commonly-used configurations for representation of a noisy quadripole*

In the first case, the noise is represented by a voltage generator and a current generator both present at the input of the quadripole (which is itself assumed noise-free), with spectral densities $S_{en}(f)$ and $S_{in}(f)$. This representation is well suited to describe noise in amplifiers. In the second case, two current generators are used, one at the input and one at the output of the quadripole, with spectral densities $Si_1(f)$ and $Si_2(f)$. This representation is useful for describing the intrinsic noise of bipolar and field effect transistors.

An important quantity for characterizing the noise of a system is its noise factor, which is defined as the quotient of the input and output signal-to-noise ratios:

$$F(f)=\frac{(S/B)e}{(S/B)s} \qquad [7.38]$$

This equation shows that the noise factor can also be written as the ratio of the total output noise power to what the output noise power would be if the quadripole was noise-free. In this way it can be shown that the noise factor is always greater than 1, and that the presence of the quadripole amplifier degrades the signal-to-noise ratio. Bearing in mind the first representation given in Figure 7.3a, and assuming a resistive load $R_g$ at the input to the quadripole which only generates thermal noise, we have, for a given frequency *f*,

$$F(f) = 1 + \frac{S_{en}(f) + R_g^2 S_{in}(f) + 2R_g \operatorname{Re}\left\{S_{enin}(f)\right\}}{4kTR_g} \qquad [7.39]$$

As $R_g$ varies, the noise factor passes through a minimum for the value:

$$R_{gopt} = \sqrt{\frac{S_{en}(f)}{S_{in}(f)}}$$

Finally, for a series of quadripoles, it can be shown that under certain conditions the noise factor of the resultant quadripole can be written as a function of the individual noise factors $F_i$ and the power gains $G_i$:

$$F - 1 = F_1 - 1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \ldots \qquad [7.40]$$

This equation (the Friis formula) shows that the first quadripole controls the noise of the whole system as long as it is not very noisy and its gain is significant. Thus, when designing a photoreceiver, it will be necessary to match the photodetector to the first amplifying component in order to minimize its noise factor and give it a sufficient gain that the noise of the successive stages can be ignored. These issues will be covered when we study the PIN-FET system.

## 7.5. Noise in photodetectors

In order to describe the noise of a photodetector in the presence of a weak signal, we will need to evaluate a signal-to-noise ratio for the information carried by the photons. In order to do this, we need to define a number of optoelectronic quantities and characteristic parameters of photodetectors, which will allow us to express the noise associated with the photocurrent generated by the conversion of photons into electron-hole pairs.

### 7.5.1. *Characteristic parameters*

In all optoelectronic transducers, there is an electric field present which separates the two types of carriers generated by absorption of the energy $E_{ph}=h\nu$ of the incident photons, of wavelength $\lambda$. For this generation to take place, we require that:

$$E_{ph} = h\nu = \frac{hc}{\lambda} \geq E_g \qquad [7.41]$$

where $h$ is the Planck constant, $\nu = \frac{c}{\lambda}$ is the optical frequency and $E_g$ the size of the bandgap of the semiconductor material. For $h\nu=E_g$ we define the cutoff wavelength $\lambda_C$ given by:

$$\lambda_c = \frac{1.24}{E_g} \left(\text{in } \mu m \text{ with } E_g \text{ in } eV\right) \qquad [7.42]$$

The response coefficient $R$ of the detector allows us to relate the generated photocurrent $I_{ph}$ to the incident optical power $P_{opt}$:

$$R = \frac{I_{ph}}{P_{opt}} \left(A/W\right)$$

The calculation of $R$ involves the external quantum efficiency $\eta_{ext}$ and the wavelength of the optical signal. We have:

$$R = \eta_{ext} \frac{q\lambda}{hc} = 0.8\eta_{ext}\lambda \left(A/W\right) \qquad [7.43]$$

if we express $\lambda$ in µm and $c$ in m/s. In calculations of noise in photodetectors, only knowledge of the photoresponse $R$ is required. The photodetector exhibits a basic noise $S_i(f)$ with which the dark current is associated:

$$I_{obs} = \frac{S_i(f)}{2q}$$

We will now consider the incident optical power which would produce a photocurrent $I_{\mathrm{ph}}$ such that the associated shot noise $2qI_{ph}$ would be equal to the basic noise referred to above (a signal-to-noise ratio of 1). This optical power characterizes the photodetector and is designated the NEP (Noise Equivalent Power). The NEP represents the smallest optical power which can be detected for a given noise equivalent bandwidth associated with the electronic system. Based on the definition of the response coefficient, we then have, for linear photodetectors and a unit bandwidth:

$$R = \frac{I_{ph}}{P_{opt}} = \frac{I_{obs}}{NEP} \tag{7.44}$$

For a given noise equivalent bandwidth, an identical equation can be written in terms of rms values. Thus, we can directly express the signal-to-noise ratio in the form:

$$\left(\frac{S}{B}\right)_{dB} = 10\log\frac{P_{opt}(rms)}{NEP} \tag{7.45}$$

We can also define another characteristic parameter of photodetectors, the detectivity, which is the inverse of the NEP:

$$D = \frac{1}{NEP}\left(W^{\ 1}\right)$$

In the case of a photoreceiver, we will keep in mind when defining the NEP that the basic noise used as a reference consists of the basic noise of the photodetector and the basic noise of the amplification system, brought down to the level of the photodetector. For a bandwidth $\Delta f$, we therefore have:

$$NEP = \frac{\sqrt{S_i\left(f\right)}_{total}\Delta f}{R} \qquad [7.46]$$

where $S_i(f)_{total} = S_i(f)_{\text{photodetector}} + S_i(f)_{amplifier}$.

### 7.5.2. *PIN photodiodes*

These are diodes, of *mesa* or planar structure, in which the n and p regions are separated by an intrinsic semiconductor region where the photocarriers are generated and separated (see Figure 7.4).

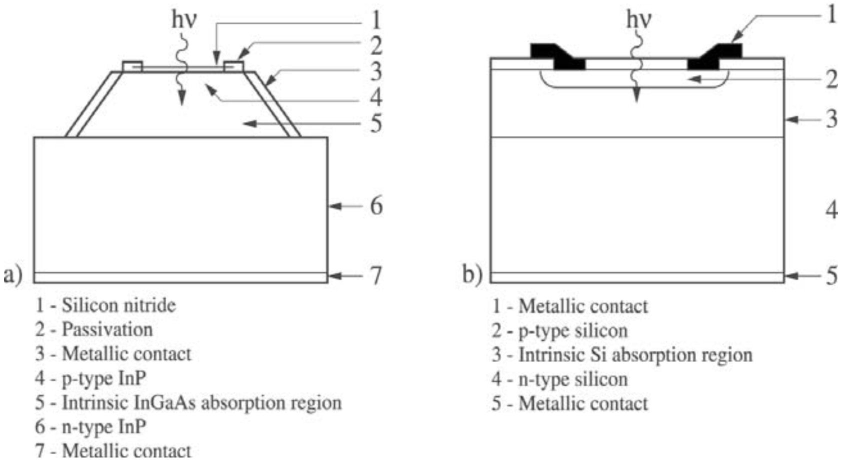


**Figure 7.4.** *Schematic cross-sections of PIN photodiodes using (a) mesa and (b) planar technology*

The most widely used semiconductor materials are shown in Table 7.2 along with their main characteristics.

| Material | Bandgap (eV) | Cutoff wavelength (μm) | Dark current at 25°C* | Dark current doubles every |
|---|---|---|---|---|
| Si | 1.11 | 1.12 | 5 nA | 7°C |
| InGaAsP | 0.89 | 1.4 | 6 μA | 10°C |
| InGaAs | 0.77 | 1.6 | 12 μA | 10°C |
| Ge | 0.67 | 1.5 | 12 μA | 8°C |

**Table 7.2.** *Main characteristics of the most widely used semiconductor materials*

A theoretical approach to these devices must take into account the band structure (direct or indirect) and optimize its geometric dimensions as a function of the bandwidths, operating temperature, dark currents, carrier lifetimes, leakage currents, series resistances, etc.

Figure 7.5 shows the ideal $I = f(V)$ characteristics of a photodiode at different incident light powers.

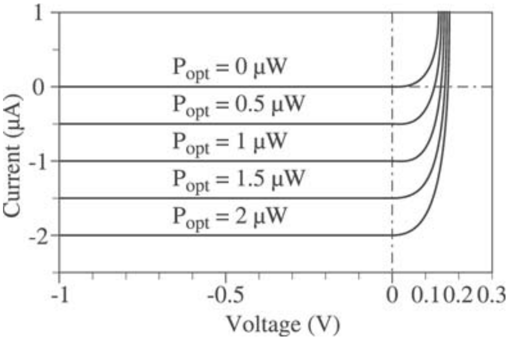


**Figure 7.5.** *Typical characteristics of a PIN photodiode under different illuminations*

Under reverse-bias, which is the normal mode of operation of a PIN photodiode, the conduction consists of two components: the dark current $I_{dark}$ and the photocurrent $I_{ph} = RP_{opt}$. Two shot noise generators are associated with these two currents, and taking into account the series resistance $R_S$ which is the source of the thermal noise, the equivalent noise circuit is shown in Figure 7.6.

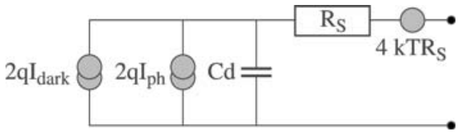


**Figure 7.6.** *Equivalent noise circuit for a reverse-biased PIN photodiode under illumination*

Starting from this equivalent circuit, we can calculate the intrinsic NEP of the device from equation [7.46]. For the large bandgap materials used to detect short wavelengths, the dark current is negligible, but technical issues are associated with the metal-semiconductor contacts due to the very high resistivity of these materials. For small bandgap materials, which are used for infrared detection, the dark current will be the main limitation on the operation of the device at ambient temperature.

### 7.5.3. *Avalanche photodiodes*

Avalanche photodiodes are PIN photodiodes which make use of the physics of impact ionization to achieve multiplication through the effect of the electric field on the charge carriers, with a mean gain *M*. This process generates noise as discussed in section 7.2.3. Figure 7.7 shows a cross-section through two avalanche photodiodes.

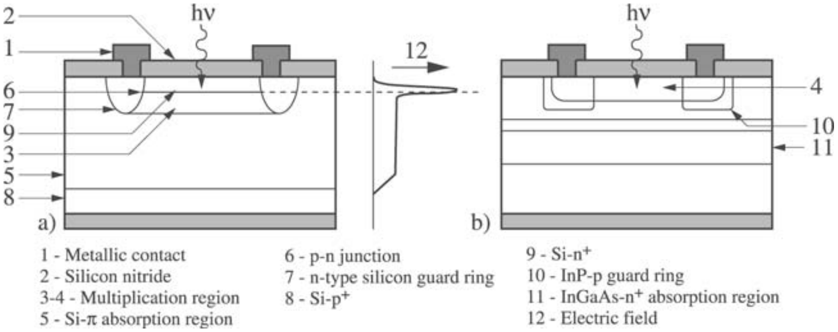


**Figure 7.7.** *(a) Structure of a $n^+p\pi p^+$ silicon avalanche photodiode for detection at $\lambda = 0.9$ µm. (b) Heterostructure avalanche photodiode with separate absorption and multiplication regions, for detection at $\lambda = 1.6$ µm*

The equivalent noise circuit is the same as for a PIN photodiode, but the photocurrent is written as:

$$I_{ph} = \overline{M} R P_{opt} \left(A\right)$$

For the shot noise, the spectral density of the noise current can at first approximation be written as:

$$S_i\left(f\right) = 2q\left\{RP_{opt} + I_{dark}\right\}\overline{M}^2 F\left(\overline{M}\right)\left(A^2/H_2\right) \qquad [7.47]$$

In general, the $I_{dark}$ component is multiplied by a factor $\overline{\mathrm{M}}$, which is different from that for the photocurrent because we do not have control over the injection into the multiplication region of the carriers originating from the dark current.

## 7.6. Noise optimization of photodetectors

The main aim when designing a photoreceiver is to minimize the optical power that needs to be detected while maintaining a specified signal-to-noise ratio. Achieving this optimal sensitivity depends on the choice of photodetector and the amplification sequence. Specifically, at the level of the detector, the signal-to-noise ratio can be written:

$$(S/B)_{detector} = \frac{\overline{i_{signal}^2}}{\int_{\Delta f} Si(f)_{detector}\, df} \qquad [7.48]$$

which in the presence of a preamplifier becomes:

$$(S/B)_{receiver} = \frac{\overline{i_{signal}^2}}{\int_{\Delta f} \left\{ Si(f)_{detector} + Si(f)_{preamp} \right\} df} \qquad [7.49]$$

Thus, the signal-to-noise ratios of the receiver and the detector are connected by the relationship:

$$(S/B)_{receiver} = \frac{1}{F}(S/B)_{detector} \qquad [7.50]$$

where $F$ is the noise factor of the preamplifier, whose input load is the photodetector. In order to achieve the minimum degradation in the signal-to-noise ratio of the photoreceiver, the amplifier must be matched to the detector in order to minimize $F$.

### 7.6.1. *Formulation of the problem*

Bearing in mind section 7.4, Figure 7.8 shows the equivalent noise circuit of the detector-preamplifier system.

The photodetector is represented by its admittance $Y_S$ in parallel with a noise current generator $I_S$. The preamplifier noise is represented at its input by a voltage generator $e_n$ and a current generator $i_n$. These two generators are correlated, and we write $i_\mathrm{n}$ in the form of one part which is correlated with $e_n$ (which we will call $i_c$), and another uncorrelated part ($i_{nc}$):

$$i_n(f) = i_c(f) + i_{nc}(f) \qquad [7.51]$$

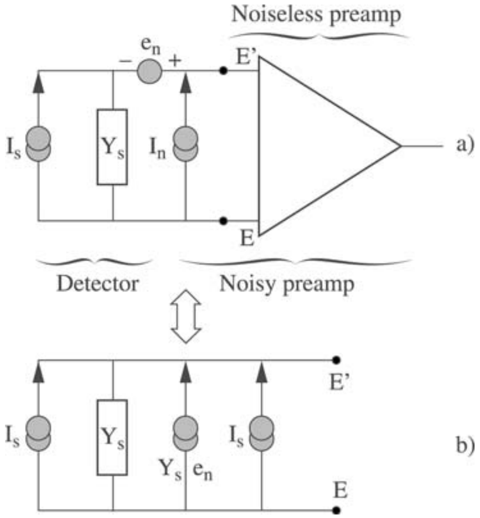


**Figure 7.8.** *Equivalent noise circuit for a photoreceiver, positioned at the system input*

Since the correlated part is proportional to $e_n$ at a given frequency, we will write it in terms of a correlation admittance $Y_{cor}$:

$$i_c(f) = Y_{cor}(f)e_n(f) \qquad [7.52]$$

The equivalent circuit in Figure 7.8b can then be drawn in the form shown in Figure 7.9.

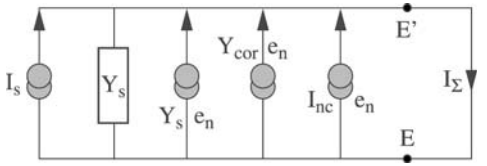


**Figure 7.9.** *Equivalent circuit used to calculate the spectral density of the total noise current of the photoreceiver*

Bearing in mind our frequency-based approach to random phenomena, we have in terms of spectral densities:

$$S_{i_\Sigma}(f) = S_{i_s}(f) + \left|Y_s + Y_{cor}\right|^2 S_{en}(f) + S_{i_{nc}}(f) \qquad [7.53]$$

The different spectral densities can be expressed as a function of their equivalent noise conductances (see section 7.4):

$$\left.\begin{aligned} S_{i_s}(f) &= 4kTG_s(f) \\ S_{en}(f) &= 4kTR_n(f) = 4kTG_n{}^{1}(f) \\ S_{i_{nc}}(f) &= 4kTG_{nc}(f) \\ S_{i_n}(f) &= 4kTG_{i_n}(f) \end{aligned}\right\} \qquad [7.54]$$

To avoid encumbering the notation, from hereon in we will not bother indicating that the conductances are a function of frequency. The noise factor of the photoreceiver now be written:

$$F = \frac{S_{i_\Sigma}(f)}{S_{i_s}(f)} = 1 + \frac{1}{G_s}\left\{G_{nc} + R_n\left[\left(G_s + G_{cor}\right)^2 + \left(B_s + B_{cor}\right)^2\right]\right\} \qquad [7.55]$$

Bearing in mind equations [7.51] and [7.52], we have:

$$G_{nc} = G_{in} - R_n\left|Y_{cor}\right|^2 \qquad [7.56]$$

and the noise factor can be written as:

$$F = 1 + \frac{1}{G_s}\left\{G_{in} - R_n\left\{G_s^2 + B_s^2 + 2\left(G_s G_{cor} + B_s B_{cor}\right)\right\}\right\} \quad [7.57]$$

In order to minimize the noise factor, we set to zero the derivative of $F$ with respect to the variables $G_s$ and $B_s$. We find:

$$B_{sopt} = -B_{cor} \quad [7.58]$$

then:

$$G_{sopt} = \sqrt{G_{cor}^2 + G_n G_{nc}} = \sqrt{G_{in} G_n - B_{cor}^2} \quad [7.59]$$

Under these conditions, the minimum noise factor can be written:

$$F_{min} = 1 + 2\frac{G_{sopt} + G_{cor}}{G_n} \quad [7.60]$$

It follows that:

$$F = F_{min} + \frac{1}{G_s G_n}\left\{\left(G_s - G_{sopt}\right)^2 + \left(B_s - B_{sopt}\right)^2\right\} \quad [7.61]$$

Hence, if we constrain $F$ to be greater than or equal to $F_{min}$ by modifying the source admittance $Y_s = G_s + jB_s$, we find we must solve the following equation in terms of $G_s$ and $B_s$:

$$G_s^2 + B_s^2 - 2G_s\left\{G_{sopt} + \left(F - F_{min}\right)G_n\right\} - 2B_s B_{sopt} + G_{sopt}^2 + B_{sopt}^2 = 0 \quad [7.62]$$

which in the basis $G_s$, $B_s$ represents a circle with center:

$$\begin{cases} G_{so} = G_{sopt} + \left(F - F_{min}\right)G_n \\ B_{so} = B_{sopt} \end{cases} \qquad [7.63]$$

and radius:

$$R = \sqrt{\left(F - F_{min}\right)G_n}\left\{\left(F - F_{min}\right)G_n + 2G_{sopt}\right\} \qquad [7.64]$$

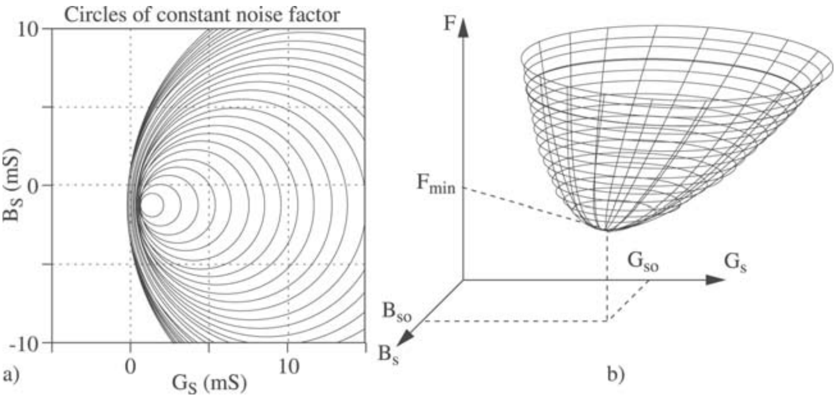


**Figure 7.10.** *(a) Circles of constant noise factor in the basis $G_S$, $B_S$. (b) Plot of the surface defined by the noise factor*

We note that for $F = F_{min}$ the circle shrinks to its center such that:

$$G_{so} = G_{sopt} B_{so} = G_{sopt}$$

The results obtained are shown in Figures 7.10a and 7.10b.

It must be remembered that the series of circles and their resultant surface is only representative of a single frequency. The matching of the detector and preamplifier must therefore be as independent as possible of this parameter.

### 7.6.2. *Concepts for photodetector-transistor matching*

A conceptual diagram for the matching is shown in Figure 7.11.

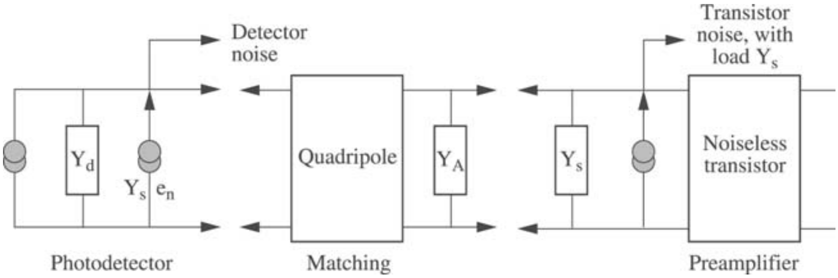


**Figure 7.11.** *Conceptual diagram for matching a photodetector to its preamplifier in order to minimize the noise factor*

$Y_s$ is the admittance presented by the matching quadripole at the input of the preamplifier. In order to have $F = F_{min}$, it is necessary that $Y_s = Y_{opt}$. Matching is achieved using a quadripole or matching transformer and/or through a matching admittance $Y_A$. $Y_d$ is the admittance of the photodetector. For example, a photodiode with a capacitance $C_d$ of 0.1 pF and a series resistance $R_s$ of 20 Ω presents an admittance of $Y_d = 0.03$ mS + $j$ 1.2 mS at 2 GHz and $Y_d = 0.78$ mS + $j$ 6.3 mS at 10 GHz.

Similarly, for a field effect transistor with a gate of width 200 µm we have, at 2 GHz: $Y_{sopt} = 2.5$ mS − $j$ 2.1 mS, and at 10 GHz: $Y_{opt} =$ 12.8 mS − $j$ 10.8 mS. Thus, in the complex plane of admittances, the matching of the photodetector to its transistor move from the point $Y_D$ to the point $Y_{opt}$ (see Figure 7.12a).

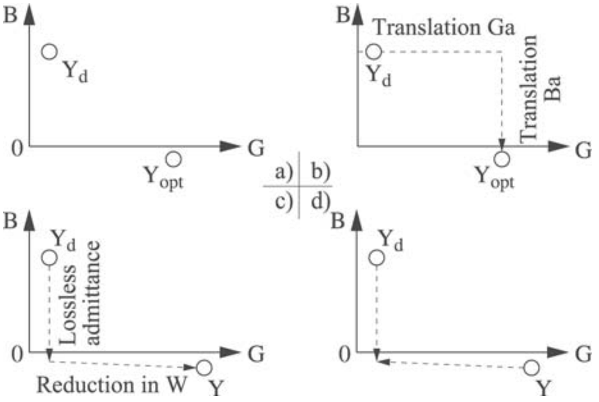


**Figure 7.12.** *(a) Representation in the admittance plane of the photodetector's $Y_d$ and of $Y_{opt}$ which leads to a minimum of the noise factor, (b) movement due to the matching admittance $Y_a = G_a - jB_a$, (c) matching by lossless admittance and reduction in the width W of the gate, (d) matching by use of a transformer*

In order to do this, we can imagine the following points:

a) connect the photodetector directly to the transistor. The noise factor will not be minimal. However, we will be close enough to the axis *G* of the conductances that $C_d$ will be weak;

b) place an admittance $Y_a = G_a + jb_a$ in parallel with $Y_d$ so that $Y_a = Y_{opt} + Y_d$ (see Figure 7.12b). The noise factor will be minimized, but at the expense of increasing the source noise. It follows that the signal-to-noise ratio is degraded: increasing the real part of the admittance seen by the transistor increases the noise and so is not a worthwhile solution;

c) reduce the width *W* of the gate of the field effect transistor in order to have $R_e\{Y_{opt}\} = R_e\{Y_d\}$ [I think that $R_e$ should read *Re* [i.e. "the real part of"]] (Figure 7.12c). The signal-to-noise ratio of the preamplifier is optimized. It should however be noted that the reduction in *W* leads to a reduction in the transconductance of the gain. This may then not be large enough for it to be possible to neglect the noise of the subsequent stage;

d) use an impedance transformer (Figure 7.12d). $Y_d$ becomes $K_e\{Y_d + jB_a\} = Y_{opt}$. In this case, it is also necessary to match the imaginary part by using a lossless admittance $jB_a$. The real part of $Y_s$ has been amplified along with the signal, and the signal-to-noise ratio of the source remains unchanged. In practice, this requires being able to build a lossless (and hence noiseless) matching circuit with a significant $K$ factor. It also requires that the lossless load $Y_a$ has a negative susceptance (inductance). Finally, the system must operate in the useful frequency range, bearing in mind that the admittance of an inductance decreases with frequency when $I_m\{Y_{opt}\}$ grows. This makes this option challenging to achieve in practice.

## 7.7. Calculation of the noise of a photoreceiver

### 7.7.1. *Basic equations*

We need to express the equivalent noise circuit in Figure 7.8 in terms of the basic noise associated with the transistor which forms the preamplifier. The behavior of the transistor will be described here in terms of its admittance matrix $Y_{ija}$, and its background noise will be represented by two correlated current generators $i_{b1}$ and $i_{b2}$ at the input and output of the device respectively (section 7.5.2). This leads to the arrangement depicted in Figure 7.13.

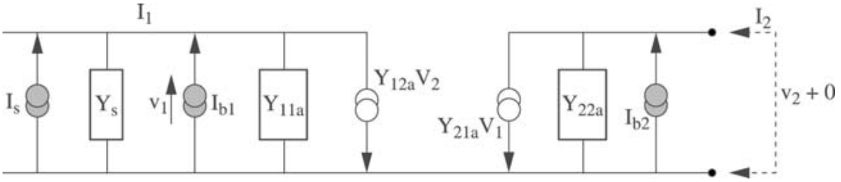


**Figure 7.13.** *Equivalent circuit for the preamplifier, highlighting the different sources of noise*

The calculation of the noise present at the input is made by considering the short-circuit output, and this does not detract from the generality of the problem since the current gain will also be calculated under these same conditions.

We set:

$$Y_{11} = Y_s + Y_{11a} \qquad [7.65]$$

and for $V_2 = 0$ we have:

$$\left. \begin{aligned} i_2 &= i_{b2} + Y_{21a} V_1 \\ V_1 &= \frac{1}{Y_{11}} \left( i_s + i_{b1} \right) \end{aligned} \right\} \qquad [7.66]$$

given that the output short-circuit current gain can be written:

$$\left. G_{i_{cc}} \right)_{V_2\ 0} = \frac{y_{21a}}{Y_{11}} \qquad [7.67]$$

we find that the spectral density of the noise current associated with current $i_2$ is:

$$S_{i_2}(f) = S_{i_{b2}}(f) + \left| G_{icc} \right|^2 \left\{ S_{is}(f) + S_{ib1}(f) + 2\,\mathrm{Re}\left\{ G_{icc} S_{i_{b1} i_{b2}}(f) \right\} \right\} \qquad [7.68]$$

and so the total noise at the input to the transistor can be written:

$$S_{i_e}(f) = S_{i_s}(f) + S_{i_{b1}}(f) + \frac{1}{\left| G_{i_{cc}} \right|^2} S_{i_{b2}}(f) + 2\,\mathrm{Re}\left\{ \frac{1}{G^*_{i_{cc}}} S_{i_{b1} i_{b2}}(f) \right\} \qquad [7.69]$$

Setting $Y_s = G_s + jB_s$ and $S_{i_{b1} i_{b2}}(f) / Y_{21_a} = G_{cor} + jB_{cor}$, we have:

$$S_{ie}(f) = S_{is}(f) + S_{i_{b1}}(f) + \frac{(G_{11a}+G_s)^2 + (B_{11a}+B_s)^2}{|Y_{21a}|^2} S_{i_{b2}}(f) \qquad [7.70]$$

$$+2\left\{G_{cor}(G_s+G_{11a}) - B_{cor}(B_s+B_{11a})\right\}$$

Insofar as $S_{is}(f)$ can be identified as the thermal noise of $G_s$, $S_{ie}(f)$ takes the form:

$$S_{ie}(f) = C + GG_s + BB_s + A\left(G_s^2 + B_s^2\right) \qquad [7.71]$$

which represents the basic equation for the input noise of a photoreceiver, with:

$$\left.\begin{array}{l} C = S_{i_{b1}}(f) + S_{i_{b2}}(f)\left|\dfrac{Y_{11a}}{Y_{21a}}\right|^2 + 2\left\{G_{cor}G_{11a} - B_{cor}B_{11a}\right\} \\ G = 2\left\{\dfrac{G_{11a}}{|Y_{21a}|^2} S_{i_{b2}}(f) + G_{cor}\right\} + 4kT \\ B = 2\left\{\dfrac{B_{11a}}{|Y_{21a}|^2} S_{i_{b2}}(f) - B_{cor}\right\} \\ A = S_{i_{b2}}(f) \Big/ |Y_{21a}|^2 \end{array}\right\} \qquad [7.72]$$

For a given value of $S_{ie}(f)$, equation [7.71] generates a circle in the admittance plane, similar to that which we saw previously.

### 7.7.2. *Models of transistor noise*

In this section, we will elucidate the two noise generators $i_{b1}$ and $i_{b2}$ as well as their correlation.

### 7.7.2.1. *Case of a field effect transistor*

$i_{b1}$ represents the noise associated with the gate and $i_{b2}$ the noise from the channel (common-source configuration of the transistor). In general terms, we have:

$$S_{i_{b1}}(f) = \frac{A_g}{f} + 2qIg + 16\pi^2 kT \frac{C_{gs}^2 f^2}{g_m} R \qquad [7.73]$$

There is a $1/f$ noise component, a shot noise component associated with the gate current, and a component produced by noise in the channel through capacitative coupling. Here $R$ is a dimensionless parameter which depends on the bias point of the device:

$$S_{i_{b2}}(f) = \frac{A_c}{f} + 4kTg_m P \qquad [7.74]$$

Here we have a $1/f$ noise contribution and a component associated with the diffusion noise (thermal noise) of the channel. $P$ is a parameter which also depends on the bias point. $A_g$, along with $A_c$, has units of $A^2$ and represents the value of the spectral density of the $1/f$ noise at 1 Hz. Working from the definition of the correlation coefficient $C_r$ between two noise sources (here the channel and gate noises):

$$C_r = C_{re} + jC_{im} = \frac{S_{i_{b1}i_{b2}}(f)}{\sqrt{S_{i_{b1}}(f) S_{i_{b2}}(f)}} \qquad [7.75]$$

we have from equations [7.73] and [7.74], where it is assumed that only the thermal noise contribution leads to correlation:

$$S_{i_{b1}i_{b2}}(f) = 4kTCgs\omega\sqrt{RP}\left\{C_{re} + jC_{im}\right\} \qquad [7.76]$$

The small-signal equivalent circuit for a field effect transistor, along with the two noise generators $i_{b1}$ and $i_{b2}$, is shown in Figure 7.14. The access points $R_g$, $R_s$ and $R_D$ may be noisy and contribute thermal noise in the form of $1/f$ noise.

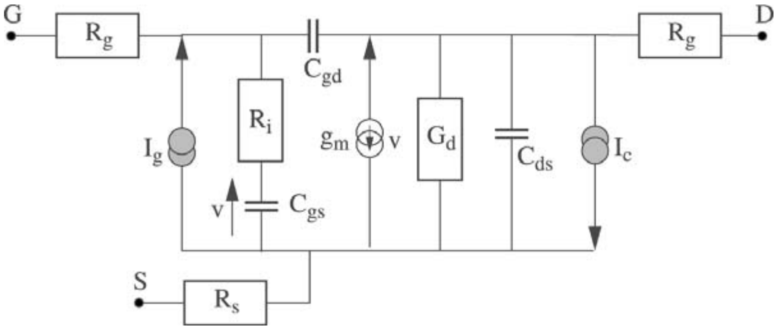


**Figure 7.14.** *Equivalent circuit of a field effect transistor, with noise current generators at its input and output*

### 7.7.2.2. *Case of a bipolar transistor*

The two noise sources model the shot noise associated with the base current:

$$S_{i_{b1}}(f) = 2qI_B = 2q\frac{Ic}{\beta} \qquad [7.77]$$

and the collector current:

$$S_{i_{b2}}(f) = 2qI_C \qquad [7.78]$$

On top of the shot noise, it is useful at low frequencies to add the $1/f$ noise sources (at least at the level of the base) and maybe account for the thermal noise of the terminal resistances (particularly for $r_{bb'}$). At high frequencies, bearing in mind that, from the point of view of the detector, the transistor is equivalent to its input resistance $r_\pi$ in parallel with a capacitance $C_T$, the total input noise to the device can be written:

$$S_{i_T}\left(f\right)=\frac{2qI_C}{\beta}+\left\{\frac{1}{r_\pi^2}+C_T^2\omega^2\right\}\frac{2qI_C}{g_m^2} \qquad [7.79]$$

Given that $r_\pi=\beta\frac{kT}{qI_c}$ and $g_m r_\pi=\beta$, we find:

$$S_{i_T}\left(f\right)=2qI_c\left\{\frac{1}{\beta}+\frac{1}{\beta^2}\right\}+2\frac{k^2T^2}{qI_c}C_T^2\omega^2 \qquad [7.80]$$

An increase in the capacitance can be counteracted by an increase in the current $I_C$. The minimum $S_{iT}(f)$ is obtained for the value of the current $I_C$ which cancels out the derivative of $S_{iT}(f)$, so that:

$$I_c=\frac{kT}{q}C_T\omega\frac{1}{\sqrt{\beta^{\;1}+\beta^{\;2}}} \qquad [7.81]$$

The input spectral noise density becomes:

$$S_{i_{opt}}\left(f\right)=4kTC_T\omega\sqrt{\frac{1+\beta^{\;1}}{\beta}} \qquad [7.82]$$

showing that this optimum varies linearly with frequency. The small-signal equivalent circuit for the bipolar transistor, along with the two noise generators $i_{b1}$ and $i_{b2}$ are shown in Figure 7.15.

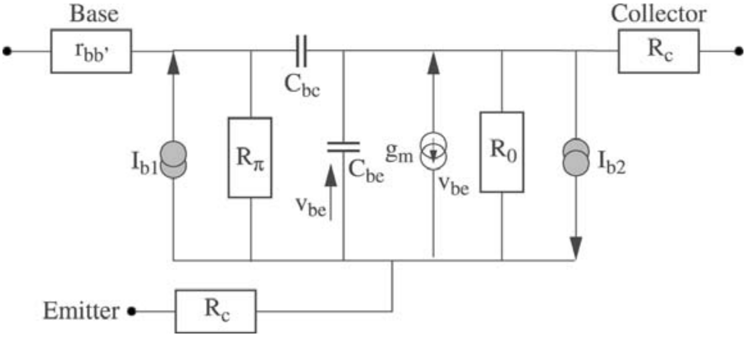


**Figure 7.15.** *Equivalent circuit for a bipolar transistor, with noise generators at its input and output*

### 7.7.3. *Example calculation: a PIN-FET photoreceiver*

Figure 7.16 shows the structure of a PIN-FET photoreceiver for direct detection of the high impedance type (without $R_f$) or the transimpedance type (with $R_f$).

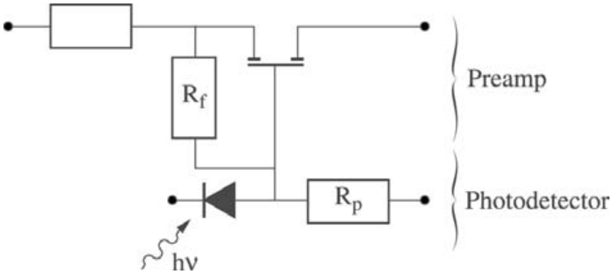


**Figure 7.16.** *Circuit diagram for a PIN-FET photoreceiver*

From this circuit diagram, the elements of the admittance matrix of the preamplifier can be written (ignoring parasitic elements):

$$\left.\begin{aligned}
Y_{11a} &= \frac{1}{R_f} + j\omega\left(C_{gs} + C_{gd}\right) \\
Y_{12a} &= -\frac{1}{R_f} - j\omega C_{gd} \\
Y_{21a} &= g_m - \frac{1}{R_f} - j\omega C_{gd} \neq g_m - \frac{1}{R_f} \\
Y_{22a} &= \frac{1}{R_f} + j\omega C_{gd}
\end{aligned}\right\} \qquad [7.83]$$

In addition, combining the PIN diode with its capacitance $C_d$, the source admittance takes the form:

$$Y_s = \frac{1}{R_{pol}} + jC_d\omega \qquad [7.84]$$

Designating the equivalent resistance of $R_{pol}$ and $R_f$ in parallel as $R_T$, and the equivalent capacitance of $C_d$, $C_{gs}$ and $C_{gd}$ in parallel as $C_T$ (there may also be parasitic capacitances $C_p$ to take into account), the equation $Y_{11} = Y_{11a} + Y_s$ can then be written:

$$Y_{11} = \frac{1}{R_T} + jC_T\omega \qquad [7.85]$$

The photodetector noise is accounted for in the shot noise that stems from the dark current $I_{dark}$ and the signal current $I_{signal}$, with a further contribution from thermal noise in the series resistance $R_{series}$ of the photodiode, which we can correct by taking account of the access resistance $R_g$ at the gate of the transistor:

$$S_{i_s}\left(f\right) = 2q\left(I_{obs} + I_{signal}\right) + 4kT\left(R_{series} + R_g\right)Cd^2\omega^2 \qquad [7.86]$$

The total noise entering the input of the system can be written (see equations [7.67] and [7.69]):

$$S_{i_\Sigma}(f) = S_{i_s}(f) + S_{i_{b1}}(f) + \left|\frac{Y_{11}}{Y_{21a}}\right|^2 S_{i_{b2}}(f) + 2\,\mathrm{Re}\left\{\frac{Y_{11}^*}{Y_{21a}^*} S_{i_{b1}i_{b2}}(f)\right\} \qquad [7.87]$$

Thus, we have:

$$\begin{aligned} S_{i_\Sigma}(f) = {} & 2q\left\{I_{dark} + I_{signal} + I_g\right\} + \frac{A_g}{f} + \frac{A_c}{f}\frac{\frac{1}{R_T^2} + C_T^2\omega^2}{g_m^2 + C_{gd}^2\omega^2} \\ & +4kT\Bigg\{\frac{1}{R_T} + \left(R_{series} + R_g\right)C_d^2\omega^2 + \frac{Cg_s^2\omega^2}{g_m}R + g_m P\frac{1/R_T^2 + C_T^2\omega^2}{g_m^2 + C_{gd}^2\omega^2} \\ & +2C_{gs}\omega\sqrt{RPRe}\left[\left(C_{re} + jC_{im}\right)\frac{1/R_T - jC_T\omega}{g_m + jC_{gd}\omega}\right]\Bigg\} \end{aligned} \qquad [7.88]$$

At this point, we can reorganize this equation to make clear the frequency behavior of the noise at the input to the preamplifier. For instance, where $g_m >> Cg_d\omega$, we have:

$$\begin{aligned} S_{i_\Sigma}(f) = {} & \frac{1}{f}\left\{Ag + \frac{Ac}{R_T^2 g_m^2}\right\} \\ & +2q\left\{Ig = I_D\right\} + 4kT\left\{\frac{1}{R_T} + \frac{P}{R_T^2 g_m}\right\} \\ & +f\left\{4f\pi^2 Ac\frac{C_T^2}{g_m^2} + 16\pi\frac{kT}{R_T}\sqrt{RP}\frac{C_{gs}}{g_m}\right\} \\ & +f^2\left\{16\pi^2\frac{kT}{g_m}\left(C_T^2 P + C_{gs}^2 R + 2C_{im}C_T\sqrt{RPC_{gs}}\right)\right\} \end{aligned} \qquad [7.89]$$

As an example, Figure 7.17 shows the variation of $pA/\sqrt{Hz}$ in $\sqrt{S_{i_\Sigma}(f)}$ as a function of frequency, for a system whose characteristics are listed in the caption.

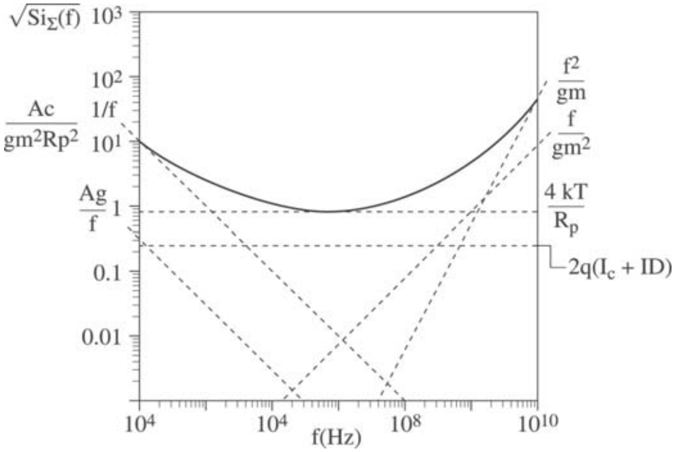


**Figure 7.17.** *Calculation of* $\sqrt{S_{i_\Sigma}(f)}$ *as a function of frequency, using the appropriate values for the characteristic parameters of the photoreceiver. Dotted lines indicate the contributions from 1/f noise, thermal and shot noise, as well as those of the f and* $f^2$ *contributions*

$$\begin{cases} C_d + C_p = 0.2pF, C_{gs} = 0.2pF \\ I_g = 150\ nA, I_{Diode} = 20\ nA \\ R_p = 30\ k\Omega, g_m = 12\ mS; P = 1.9, R = 0.4 \\ C_{re} = 0.2, C_{im} = 0.4 \\ A_c = 10^{11}\ pA^2, A_g = 10^3\ pA^2 \end{cases}$$

At high frequencies, in the case where $Cg_d\omega$ remains smaller than $g_m$, equation [7.88] can be written:

$$\begin{aligned} S_{i_\Sigma}(f) = 4kT & \left\{ \frac{1}{R_T} + \left(R_{series} + Rg\right) Cd^2 \omega^2 + \frac{C_{gs}^2 \omega^2}{g_m} R \right. \\ & \left. + \frac{P}{g_m} \left( \frac{1}{R_T^2} + C_T^2 \omega^2 \right) + 2 C_{gs} \omega \sqrt{RP} \left( Cre \frac{1}{R_T g_m} + C_{im} \frac{C_T \omega}{g_m} \right) \right\} \end{aligned} \qquad [7.90]$$

Starting from equation [7.90], it can be seen how the width *W* of the transistor gate can be calculated to match it to the photodetector, setting:

$$\begin{aligned} & g_m = g_{mo} W \\ & C_{gs} = C_{gso} W \\ & C_{gd} = C_{gdo} W \text{ and } Co = C_{gso} + C_{gdo} \end{aligned}$$

We then have:

$$\begin{aligned} S_{i_\Sigma}(f) = 4kT & \left\{ \frac{1}{R_T} + \left(R_{series} + Rg\right) Cd^2 \omega^2 + \frac{Cgs_o^2}{g_{m_0}} W \omega^2 R + \frac{P}{g_{m_0} W} \frac{1}{R_T^2} \right. \\ & + P \frac{C_o^2 W \omega^2}{g_{m_0}} + P \frac{Cd^2 \omega^2}{g_{m_0} W} + 2P \frac{CdCo}{g_{m_0}} \omega^2 \\ & + \frac{2Cgso}{g_{m_0}} \sqrt{RP} C_{im} Co W \omega^2 + \frac{2Cgso}{g_{m_0}} \sqrt{RP} C_{im} Cd \omega^2 \\ & \left. + 2 \frac{Cgso}{g_{m_0}} \omega \sqrt{RP} \frac{Cre}{R_T} \right\} \end{aligned} \qquad [7.91]$$

In terms of the variable *W*, this last equation can be rearranged to obtain:

$$\begin{aligned}S_{i_\Sigma}(f) &= W\frac{4kT}{g_{m_0}}\left\{Cgso^2R + PCo^2 + 2Cgso\sqrt{RP}C_{im}Co\right\}\omega^2\\ &+4kT\left\{\frac{1}{R_T} + 2\frac{P}{g_{m_0}}CdCo\omega^2 + 2\frac{Cgso}{g_{m_0}}\sqrt{RP}\left(C_{im}Cd\omega^2 + \frac{Cre}{R_T}\omega\right)\right\}\\ &+\frac{1}{W}4kT\frac{1}{g_{m_0}}\left\{P\left(\frac{1}{R_T^2} + Cd^2\omega^2\right) + \left(R_{s`rie} + Rg\right)Cd^2\omega^2Wg_{m_0}\right\}\end{aligned} \qquad [7.92]$$

Since the gate access resistance is inversely proportional to:

$$W\left(Rg = \frac{Rgo}{W}\right)$$

it is reasonable to make the following approximation when integrating the photoreceiver:

$$\left(R_{s`rie} + Rg\right)Cd^2\omega^2Wg_{m_0} = \left(R_{s`rie_0} + Rg_o\right)Cd^2\omega^2g_{m_0} \qquad [7.93]$$

Additionally, the term representing the noise induced by the channel in the gate, $4kT\frac{Cgs^2\omega^2}{g_m}R$, can be written as:

$$4kT\frac{Co^2\omega^2}{g_m}R' \text{ with } R' = \frac{Cgs^2}{Co^2}R \qquad [7.94]$$

Equation [7.92] then takes the final form:

$$\begin{aligned} S_{i_\Sigma}(f) = 4kT \Bigg\{ & W\frac{Co^2}{g_{m_0}}\left(R' + P + 2\sqrt{R'P}C_{im}\right) \\ & + \frac{1}{R_T} + 2\frac{P}{g_{m_0}}CdCo\omega^2 + 2\frac{Co}{g_{m_0}}\sqrt{R'P}\left(C_{im}Cd\,\omega + \frac{Cre\,\omega}{R_T}\right) \\ & + \frac{1}{W}\frac{1}{g_{m_0}}\left[P\left(\frac{1}{R_T^2} + Cd^2\omega^2\right) + Cd^2\omega^2\left(R_{series_0} + Rg_o\right)g_{m_0}\right]\Bigg]\Bigg\} \end{aligned} \quad [7.95]$$

The derivative of this final result, with respect to *W*, leads us to an optimal value of *W* which minimizes the noise, such that:

$$W_{opt} = \frac{Cd}{Co}\sqrt{\frac{P\left(1 + \frac{1}{R_T^2Cd^2\omega^2}\right) + g_{m_0}\left(R_{series} + Rg_o\right)}{R' + P + 2\sqrt{RP}C_{im}}} \quad [7.96]$$

We then have:

$$\begin{aligned} Si_\Sigma(f) = 4kT & \left\{\frac{1}{R_T} + 2\frac{CoCd\omega^2}{g_{m_0}}\left[P + \sqrt{R'P}\left(C_{im} + \frac{Cre}{R_TCd\omega}\right)\right]\right. \\ & + \sqrt{\left\{P\left(1 + \frac{1}{R_T^2Cd^2\omega^2}\right) + g_{m_0}\left(R_{series} + Rgo\right)\right\}\left(R' + P + 2\sqrt{R'P}C_{im}\right)} \end{aligned} \quad [7.97]$$

In the case where the value of $R_T$ dominates, we have:

$$\begin{aligned} & Si_\Sigma(f) = 8kT\frac{Co}{g_{m_0}}Cd\omega^2 \\ & \left\{P + \sqrt{PR'}C_{im} + \sqrt{P + g_{m_0}\left(R_{series} + Rg_o\right)}\sqrt{R' + P + 2\sqrt{R'P}C_{im}}\right\} \end{aligned} \quad [7.98]$$

which can be written as:

$$Si_{\Sigma}\left(f\right)_{\min}=16kT\frac{Co}{g_{m_0}}Cd\omega^2\Gamma \qquad [7.99]$$

with:

$$\Gamma=\frac{1}{2}\left\{P+\sqrt{PR'}C_{im}+\sqrt{\left\{P+\left(R_{series}+Rg_o\right)g_{m_0}\right\}\left\{R'+P+2\sqrt{R'P}C_{im}\right\}}\right\}$$

We can consider $\Gamma$ as an equivalent noise parameter for the channel which allows us to treat the system as if the gate noise is zero ($R = C_{\mathrm{r}} = 0$). Equation [7.99] accounts for the high-frequency noise of the receiver. Now returning to the sources of noise that we have ignored up to now, the total input noise after the gate width has been optimized is:

$$\begin{aligned}Si_{\Sigma}\left(f\right)_{min}&=2q\left(I_{dark}+I_{signal}+I_g\right)+\frac{A_g}{f}+A_c4\pi^2\frac{C_T2}{g_m^2}f\\&+64\pi^2kT\frac{Co}{g_{m_0}}Cd\Gamma f^2\end{aligned} \qquad [7.100]$$

This equation gives the frequency dependence of the optimized noise.

## 7.8. Comments and conclusions

Several comments can be made based on the example that we have worked though. There are four sources of noise that are a function of frequency:

– a $1/f$ contribution which dominates at low frequencies and is connected with the manufacturing technology of the devices;

– a white noise component which represents the minimum noise level of the photoreceiver;

– a component proportional to frequency, associated with the high frequency performance of the transistor and the capacitative coupling between the channel and the gate, which brings into play the low frequency noise of the channel;

– a component proportional to the frequency squared, associated with the Γ coefficient defined above.

The noise contribution from each element in the photoreceiver circuit is clear:

– the intrinsic characteristics of the PIN photodiode ($I_{obs}$, $C_d$ and $R_{series}$) contribute to the white noise and the high frequency noise as a result of its matching to the amplifier;

– the bias resistance of the photodiode has an effect on the white noise and low frequency noise of the system;

– the transistor characteristics ($I_g$, $g_m$, $C_{gs}$, $C_{gd}$, $A_c$, $A_g$ and Γ) affect the noise performance of the photoreceiver across the whole frequency range of its operation.

Finally, at the level of the whole photoreceiver, and to summarize all the concepts presented in this chapter, we calculate the signal-to-noise ratio for an incident optical power $P_{opt}$ and a photodiode response coefficient $R$ through the equation:

$$\frac{S}{B} = \frac{\left(RP_{opt}\right)^2}{\int_{B_{en}} Si_{\Sigma}\left(f\right)df} \qquad [7.101]$$

where $B_{en}$ represents the equivalent noise bandwidth.

This type of calculation can be applied to different photoreceivers involving bipolar or field effect transistors connected to PIN or avalanche photodiodes.

## 7.9. Bibliography

[BER 93] BERTHIER P., Transistors à effet de champ AlInAs/(Al)GaInAs(P) pour photodétection intégrée à 1,3-1,5 µm, PhD thesis, Montpellier University, December 1993.

[BER 94] BERTHIER P., GIRAUDET L., SCAVENEC A., RIGAUD.D, VALENZA M., DAVIES J.I., BLAND S.W., "InGaAsP channel HFET's on InP for OEIC applications", *Journal of Light Wave Technology*, vol. 12, n° 12, December 1994.

[BIQ 92] BIQUARD M.F., *Signaux, systèmes linéaires et bruit en électronique*, Ellipses, 1992.

[DAS 95] DAS M.B., CHEN J.W., JOHN E., "Designing optoelectronic integrated circuit (OEIC) receivers for high sensitivity and maximally flat frequency response", *Journal of Light Wave Technology*, vol. 13, n° 9, September 1995.

[FIS 93] FISH P.J., *Electronics noise and low noise design*, Macmillan New Electronics series, 1993.

[HEN 89] HENTSCHEL C., *Fiber Optics Handbook*, Hewlett-Packard, 1989.

[HOO 76] HOOGE F.N., "1/f noise", *Physica*, vol. B83, 1976.

[JOI 96] JOINDOT I.M., *Les télécommunications par fibres optiques*, Dunod, 1996.

[KIM 97] KIM M.J., KIM D.K., KIM S.J., DAS M.B., "Determination of bit-rate and sensitivity limits of an optimized p-i-n/HBT OEIC receiver using SPICE simulations", *IEEE Transactions on Electron Devices*, vol. 40, n° 4, April 1997.

[KRE 93] KRESSEL H., *Semiconductors Devices for Optical Communication*, Springer-Verlag, Topics in Applied Physics, vol. 39, 1987.

[LEG 93] LEGROS E., Photorécepteur intégré pour transmission cohérente sur fibre optique, University Thesis, Orsay, December 1993.

[PAR 93] PARK M.S., MINASIAN R.A., "Ultralow noise 10Gb/s p-i-n-HEMT optical receiver", *IEEE Photonics Technology Letters*, vol. 5, n° 2, February 1993.

[MCW 57] MC WHORTER A.L., *Semiconductor Surface Physics*, University of Pennsylvania Press, Philadelphia, 1957.

# List of Authors

Robert ALABEDRA
CEM2
University of Montpellier II
France

Baudoin DE CREMOUX
Central research laboratory
Thomson-CST
France

Didier DECOSTER
IEMN
Lille University of Science and Technology
France

Carmen GONZALEZ
Alcatel Thales III-V Lab
Marcoussis
France

Joseph HARARI
IEMN
Lille University of Science and Technology
France

Vincent MAGNIN
IEMN
Lille University of Science and Technology
France

Antoine MARTY
LAAS-CNRS
Toulouse
France

Eva MONROY
INAC
Nanophysics and Semiconductors Laboratory
Grenoble
France

Franck OMNES
Neel Institute
Grenoble
France

Dominique RIGAUD
CEM2
University of Montpellier II
France

Gérard RIPOCHE
Alcatel Research Centre
Marcoussis
France

# Index