Discussion about photodiode architectures for space applications

O. Gravrand
G. Destefanis
C. Cervera
J.-P. Zanatta
et al.
DISCUSSION ABOUT PHOTODIODE ARCHITECTURES
FOR SPACE APPLICATIONS

O. Gravrand1, G. Destefanis1, C. Cervera1, J.P. Zanatta1, N. Baier1, A. Ferron1, O. Boulade2
1 CEA-LETI-Minatec Campus, 17 rue des martyrs, 38054 Grenoble France
2 CEA-IRFU-SAp, CEA Saclay, 91191 Gif sur Yvette France

I. INTRODUCTION: CLASSICAL SPACE NEEDS

Detection for space application is very demanding on the IR detector: all wavelengths, from visible-NIR (2-3µm cutoff) to LWIR (10-12.5µm cutoff), even sometimes VLWIR (15µm cutoff) may be of interest. Moreover, various scenarios are usually considered. Some are imaging applications where the focal plane array (FPA) is used as an optical element to sense an image. However, the FPA may also be used in spectrometric applications where light is triggered on the different pixels depending on its wavelength. In some cases, star pointing is another use of FPAs where the retina is used to sense the position of the satellite.

In all those configurations, we might distinguish several categories of applications:
- low flux applications where the FPA is staring at space and the detection occurs with only a few number of photons.
- high flux applications where the FPA is usually staring at the earth. In this case, the black body emission of the earth and its atmosphere ensures usually a large number of photons to perform the detection.

Those two different categories are highly dimensioning for the detector as it usually determines the level of dark current and quantum efficiency (QE) requirements. Indeed, high detection performance usually requires a large number of integrated photons such that high QE is needed for low flux applications, in order to limit the integration time as much as possible. Moreover, dark current requirement is also directly linked to the expected incoming flux, in order to limit as much as possible the SNR degradation due to dark charges vs photocharges. Note that in most cases, this dark current is highly depending on operating temperature which dominates detector consumption. A classical way to mitigate dark current is to cool down the detector to very low temperatures.

This paper won't discuss the need for wavefront sensing where the number of detected photons is low because of a very narrow integration window. Rigorously, this kind of configuration is a low flux application but the need for speed distinguishes it from other low flux applications as it usually requires a different ROIC architecture and a photodiode optimized for high response speed.

II. NARROW GAP PHOTODIODE DARK CURRENTS FOR DUMMIES

Before getting into discussion about photodiode structures, it is interesting to go back to basics, examining classical expressions of dark currents in narrow gap photodiodes. We distinguish diffusion current from depletion current.

First the famous Shockley ideal diode equation describes the diffusion current, ie the effect of thermal generation of minority carriers aside from the junction, diffusing up to the space charge region during its lifetime $\tau_{\text{diff}}$. If the thermal generation is $G_{\text{th}}$ in the volume $V_{\text{diff}}$, the diffusion current is written as:

$$I_{\text{diff}} = q \ G_{\text{th}} \ V_{\text{diff}}$$

The thermal generation being the ratio between the minority carrier concentration $N_{\text{mino}}$ and minority carrier lifetime, it might be written as:

$$G_{\text{th}} = \frac{N_{\text{mino}}}{\tau_{\text{diff}}} = \frac{n_i^2}{N_{\text{Dop}} \ \tau_{\text{diff}}}$$

where $n_i$ is the intrinsic carrier density and $N_{\text{Dop}}$ is the majority carrier density in the extrinsic regime (at low temperature). We therefore see an important figure of merit for a diffusion limited photodiode: the product $N_{\text{Dop}} \times \tau_{\text{diff}}$.

Concerning depletion currents, thermal generation into the space charge region (SCR) of volume $V_{\text{SCR}}$ might induce a so called generation-recombination (GR) current usually expressed as:

$$I_{\text{GR}} = q \ G_{\text{th}} \ V_{\text{SCR}}$$

This time the thermal generation in the space charge region (depleted of carriers) is given by a simplification of the well-known Shockley-Read-Hall formula:

$$G_{\text{th}} = \frac{n_i^2}{n \ \tau_e + p \ \tau_h} \approx \frac{n_i}{2 \ \tau_{\text{SRH}}}$$

where we made the assumption that $n = p = n_i$ in the SCR region and $\tau_e = \tau_p = \tau_{\text{SRH}}$ far a mid-gap neutral recombination SRH center.
We therefore face here the classical statement: diffusion current is proportional to $n_i^2$ whereas GR current is proportional to $n_i$. In other words intrinsic carrier concentration being proportional to $\exp\left(-\frac{E_g}{2kT}\right)$, the activation energy of diffusion current is the energy gap $E_g$ whereas it is half the gap energy for GR currents. Hence, diffusion dark current decreases twice faster than GR current when decreasing the temperature so that at low temperature GR current always end up dominating the dark current.

Another picture would be that, in the case of a homojunction, the ratio between diffusion and GR given by:

$$\frac{I_{diff}}{I_{GR}} = \frac{\tau_o n_i}{V_{diff} \tau_{diff} n_{\text{dop}} \frac{V_{GR}}{V_{diff}}} \propto n_i \frac{\tau_o}{\tau_{diff}} V_{diff} V_{GR}$$

Therefore, apart from the SCR extension diffusion tends to dominate GR for high intrinsic carrier concentration (ie narrow gap), whereas GR will easily dominate diffusion for larger gap (short cutoff wavelength). The effect doping level will be discussed later on (section III) as it has an effect only on diffusion current. Naturally this ratio depends also on the ratios of the respective volumes and lifetime.

Note also that the evolution of diffusion dark current is weakly depending on bias $V$, for biases higher than a few $kT/q$. However, because of the space charge width variation, GR dark current exhibits a stronger variation with bias (typically proportional to $\sqrt{V}$ for an abrupt junction and closer to $\sqrt[2]{V}$ for a linearily graded junction).

Another important difference between diffusion and GR dark current lies in the fact that diffusion might be limited by fundamental recombination mechanisms such as Auger whereas GR is due SRH processes which are highly depending on material quality.

Other sources of depletion dark currents are tunneling currents, namely direct band to band tunneling (BBT) and trap assisted tunneling (TAT) currents. Those tunnel currents arise from the fact that in narrow gap semiconductors, the SCR may be pinched so that the top of the valence band on the P side overcome the bottom of the conduction band on the N side and minority carrier may cross the SCR by tunnel. Those currents are highly depending on operating bias and doping level. Hence, they are usually not limiting the photodiode for low dopings in NIR to LWIR bands, but limiting the polarization plateau (see Fig. 1). Usually, tunnel currents are not correlated with the junction area but correlated with junction perimeter, demonstrating the fact that it is barely a fundamental limitation but rather a technological induced limitation.
Surface currents are often mentioned as a source of current limitation for narrow gap photodiodes. Those surface currents might be classified in two main categories. Shunt leakage from the surface passivation occurs when the passivation is not fully isolating both sides of the junction (regarding the dark current requirement that might be extremely low). Surface recombination arises when the narrow gap material is damaged underneath the surface by all the different processing steps. Consequently, a localised high trap density might be the source of an excess GR current degrading the total performance. The same kind of effect might also be obtained by simply breaking the crystal symmetry with an abrupt interface resulting in a high level of interface traps, potentially sources of GR currents in the SCR.

III. NARROW GAP MATERIALS AND MINORITY CARRIER LIFETIME

As we have seen recombination mechanisms are of primarily importance for narrow gap photodiode performances (see Fig. 2). At the end, the resulting inverse lifetime is given by the sum of all different recombination inverse lifetimes. Apart from the radiative recombination occurring only at high temperature, we mainly face with Shockley-Read-Hall (SRH) recombination and Auger recombination (Auger1 dominating in N type material and Auger7 usually dominating P type material). Auger recombination is a multicharrier process that might be understood as the reverse process of an impact ionisation process. As a consequence, it highly depends on carrier concentration ie on doping $N_{dop}$ for low operating temperatures:

$$\tau_{Auger} \propto \frac{1}{N_{dop}} \quad (6)$$

On the other side, SRH recombination is a single carrier process therefore exhibiting a very mild dependence on doping. This recombination relies on an intermediate trap level in the bandgap. The energy difference between band edge and trap state being lower than the gap, the recombining minority carrier uses this trap to thermally jump easily from its initial band to this intermediate state and then jump again up to the other band to recombine. This recombination process depends on trap density and cross section, but also on trap location in the gap. The most optimum neutral recombination center is located at the intrinsic energy level $E_i$, usually close to midgap. For HgCdTe, due to the large disparity in the effective masses of the electron and heavy hole, it might be closer to a band edge but this matter remains under active discussion in the community $[1]$ $[2]$.

As a consequence, Auger is a fundamental recombination process, ie a perfect semiconductor crystal (free of any defects) will undergo this kind of recombination limiting the minority carrier lifetime. On the contrary, SRH depends on trap density and will therefore highly depend on diode process or material growth and passivation. Typical values for HgCdTe are from 2 to 20μs for very high quality crystals ($[1]$), confirmed by measurement carried out at LETI for In doped material, see Fig. 3), whereas other 3-5 materials usually exhibit more active SRH centers resulting in lower lifetime. Typical values of the literature are generally taken below 300ns even for very high quality InSb, resulting with very strong GR current compared to HgCdTe, even at 80K (typically 4 orders of magnitude, $[1]$). We end up with the conclusion that high lifetime for low doping requires ultra-high quality material. Any concession made on this point will lead to degraded lifetime and thus lower performances.

The case of type 2 super lattice (SL2) is more complex and need further discussions. The idea of this material is to use the ability of molecular beam epitaxy (MBE) to modulate the crystal potential in one direction using a very thin alternations (typically 8 to 10 monolayers) between two wide gap semiconductors. By symmetry considerations, it allows to fold the 1st Brillouin zone creating minibands. The resulting synthetic material behaves like an absorbing bulk semiconductor without any selection rules as it is the case in QWIPs. Therefore, using widegap materials MBE it is possible to synthetize an narrow band semiconductor with direct absorption as a bulk material. The thicknesses of the alternating layers determine the miniband and thus the cutoff wavelength. Using for example InAs and InGaSb allows the growth of MWIR to LWIR, up to VLWIR materials.

It is often argued that the cutoff control is better in this case because it relies on the management of layers thicknesses rather than an alloy composition, which is sometimes claimed as easier, especially for LWIR and VLWIR materials. However this ability might still be discussed, given the fact that the thickness control has to be done typically on thousands of layers.

Moreover, playing with the band alignment properties between the two materials, it seems possible to optimize the minority carrier lifetime. Grein $[1]$ computed Auger lifetime exceeding HgCdTe values. Later, those computations where compared to lifetime measurement by Youngdale $[2]$. These experimental values were very close to expected theoretical values for very high dopings. However, lower doping values seemed limited below 10ns by a strong SRH process. Other references have been trying to estimate this SRH lifetime ending up with sometimes slightly higher values, 35ns for LWIR$[6]$ 80ns for MWIR $[7]$. Ultimately it appears that InAs/InGaSb material system seems limited by a strong SRH process, wherever the material is grown (this assessment has been done by different laboratory, using different MBE machines).
Ga free SL (InAs/InAsSb) material has therefore been introduced to overcome this low lifetime limitation. In this system, the electrons and holes are spatially separated in the two layers (InAs or InAsSb) so that the recombination is supposed to be drastically reduced. Experimental values around 400ns has been reported for undoped material \[8\], decreasing with doping level, which is significantly higher than the preceding values, but it appears that this carrier separation drastically reduce the optical absorption of the material, driving to a priori weak QEs \[9\].

Fig. 3: Summary of minority carrier lifetime values for main IR absorbing materials \[1\] \[10\] \[5\] \[6\] \[17\] \[8\]

Back to HgCdTe, this material may be grown with different growth methods: liquid phase epitaxy (LPE), molecular beam epitaxy (MBE) and metal organic vapour phase epitaxy (MOVPE). LPE is the oldest technic and allows the growth of single layers on lattice matched CdZnTe substrates. It uses the melt of a large mother ingot containing the right amount of elements (Cd, Hg, Te, and dopants) and all layers grown from the same mother ingot exhibit very close characteristics (in terms of alloy composition and doping). The resulting crystalline quality is very high and this growth process remains the standard growth process for mass production at least at Sofradir, but also some other HgCdTe manufacturing companies in the US (DRS, Raytheon).

As opposed to LPE, MBE and MOVPE are two vapour phase epitaxies that allow the precise control of the alloy composition during the epitaxial growth. As a consequence, it is therefore possible to adapt the composition and doping profiles for optimised diode performances using heterostructures. Going further, trying to compare MBE and MOVPE is a hard task as both technologies have their pros and cons \[11\]. Notice however that MBE allows the growth on lattice matched CdZnTe substrate to minimize the resulting dislocation densities, whereas MOVPE is usually performed on highly mismatched GaAs substrates (14\%). The crystalline orientation being different in this case (100 instead of 211), dislocation density assessment is difficult to quantify by the usual etch pit density (EPD) techin.

Fig. 4: Example of photodiode optimization for diffusion current in MW

When dealing with the optimization of diffusion limited photodiode, the doping level appears as the main technological parameter. As see in equation (2), the figure of merit is the \(N_{dop} \times \tau\) product. Therefore, the
game is to get the highest lifetime with the highest doping. In this frame, the use of Hg vacancies in HgCdTe is damaging, and it should be preferred to switch to extrinsic doping, p on n for instance. Then, the transition between SRH limited lifetime (for low doping levels) to an Auger limited lifetime (for higher doping) introduces a bell curve with an optimum doping level regarding diffusion current. For p on n HgCdTe photodiodes, this optimum doping corresponds to the very popular Rule07 [12] describing empirically the dark current of the Auger1 to SRH limitation. This situation is also illustrated in Fig. 4. This figure compares the situation of a MW p/n HgCdTe photodiode to an equivalent n/p SL2 photodiode. As expected from the lifetime discussion, the optimum is higher for the HgCdTe material system. However, the difference between the two is lower than the SRH lifetime differential, because of the fact that SL2 optimum involves a higher doping. This consideration supposes that such a high doping level is technologically accessible and won’t degrade other photodiode parameters such as QE or tunnel currents.

IV. NARROW GAP STRUCTURES (DIODES AND BARRIODES)

The previous considerations are considering diffusion dark current (usually dominating at high temperature), where the photodiode is limited by an Auger recombination mechanism, which stands for a fundamental limitation of the narrow gap material. Decreasing the temperature to decrease this dark current for low flux detection, the photodiode might get into a GR limited regime, where the system becomes limited by depletion current depending on SRH trap density. The photodiode structure might therefore be optimized to mitigate the sensitivity to defects and therefore stay longer limited by diffusion current. A narrow gap photodiode structure might be of different types, namely planar, mesa and vertically integrated (Fig. 5). The vertically integrated one won’t be long discussed here because it is barely used for space applications. In a planar diode, the passivation is performed on a planar surface whereas for the mesa structure, the passivation has to be deposited on the sidewalls, which is less comfortable for a good conformal deposition. However, mesa delineation allows taking full benefits of MBE/MOVPE heterostructures, whereas in the planar case, the junction has to be formed usually using ion implantation. 

In this latter case, the doping profile is driven by the implantation conditions. However, in both cases, the game is almost identical: one wants to position most of the SCR in a region where the bandgap is the largest. For example, the Teledyne planar structure (double layer photodiode heterostructure DLPH) consists in growing a N type wide bandgap layer on top of the narrow gap N type absorbing layer. The P region is formed using As ion implantation. The implantation conditions are such that the SCR lies as much as possible in the wideband region, thus minimizing the GR current. However, the bottom of the SCR has to enter the narrow gap layer otherwise a barrier will form on the valence band impeding the flow of photocarriers coming from the absorbing layer (The resulting QE degradation will be discussed later in next paragraph). GR current depends exponentially on the gap. This means that for a DLPH structure with an abrupt interface between the two layers, the GR current will still be dominated by the narrow gap part of the SCR (exhibiting the corresponding activation energy). In this case the gain in equation (5) is due to the ratio of the diffusion volume to this GR volume. In this case the GR decrease is linear, not exponential. Going further into the analysis, the interfaces are never perfectly abrupt in HgCdTe (because of diffusion during technological bakes) so that the p region may end up in a composition grading. In the case of such a pn junction lying in a graded interface, the valence band will always exhibit a small barrier [21]. At high temperature, the minority carriers might have high enough energy to overcome this barrier (thermionic effect) whereas at lower temperature the thermal energy might become insufficient and the barrier starts filtering out minority carriers resulting in a degraded QE. As a consequence playing with heterostructures to optimize GR current might represent a risk concerning QE at low temperatures.
The same kind of heterostructure may also be obtained with LPE layers. As an example, the LETI-Sofradir p on n process makes use of a smart 3D centered interdiffusion process \(^{[14]}\), where the Cd interdiffusion is accelerated in the presence of As. As a consequence, the gap widens in the SCR resulting in a 3D auto centered heterostructure. This effect might even be enhance with a strongly interdiffused passivation \(^{[15]}\) to decrease GR current. This effect has been successfully used at LETI under CNES contract in the LWIR range. It resulted in a significant decrease in the depletion-GR knee temperature \(^{[15]}\).

At last, there exists another detection structure becoming very popular in the last few years: the barriode or nBn structure \(^{[18]}\). In this case, the band gap engineering versatility of HgCdTe \(^{[19]}\) or SL2 \(^{[20]}\) (or even InAs \(^{[21]}\) for MWIR blue band) is used to build a unipolar N structure that mimics the p on n valence band for the transport of the minority holes while the majority electrons are blocked by a strong barrier on the conduction band. The structure consists in a low doped N type absorbing layer. On top of it, a N type wide gap barrier layer is grown followed by a thin smaller gap contact layer (see Fig. 6). At last, only this contact layer is etched to reticulate the pixels. In order to maintain an optimum crystalline quality, the material of the barrier layer is usually chosen very close to the lattice matched conditions. The structure being unipolar, it might be understood as unity gain photoconductor limited by minority carriers. Hence, it doesn’t collect at zero bias and for a good hole transport across the barrier (ie a good QE), the operating bias is slightly higher than for a photodiode to overcome the turn-on voltage. The choice of the material for the barrier layer has to take into account the band offsets between the absorbing layer and the barrier layer in order to minimize the formation of a potential barrier onto the valence band, increasing this turn-on voltage. In case of a barrier formation, the QE is strongly degraded just like in the case of a standard heterojunction. Moreover, the barrier has to be high enough onto the conduction band in order to filter out efficiently majority electrons and mitigate as much as possible the thermionic current (majority carriers from the contact layer able to “jump” over the conduction band barrier).

Being unipolar, there is no pn junction strictly speaking. However, under operating bias, the barrier is depleted, just like in the case of conventional pn junction. If the structure is correctly designed, the depletion region stays in the widegap layer thus drastically minimizing GR current in comparison with a simple pn homojunction (High gains have been simulated at LETI for HgCdTe, see Fig. 8). However, for large biases, the depletion might start to extend in the narrow gap region and GR current starts to limit the photodiode, just like for classical pn homojunctions. The gain of a nBn bariode structure is expected on the GR current. As a consequence, a “bad” material system (in the SRH sense) will exhibit a larger gain. However, there will be no gain in terms of performances compared to a diffusion limited photodiode. Remind also that Fig. 8 is the comparison between nBn bariode and p/n diode. In both cases, the absorption is done in the N type diffusion layer. The difference only lies in the collection mechanism. The comparison is done with a pn homojunction with the same amount of SRH traps. At 78K, the computed diode is strongly dominated by GR current. However, we have seen that real p/n diodes in HgCdTe are heterojunctions, which means that the volume of the SCR in the narrow gap material has been reduced, leading to a reduced GR current. Hence, the resulting dark current of such heterostructures should be lower, probably between the nBn bariode and the pn homojunction.

**Fig. 7: Dark current optimization of HgCdTe LWIR photodiodes for low flux detection**\(^{[15]}\) \(^{[16]}\)
V. QE ISSUES

Usually, photodiode structures give access to very high QE, typically larger than 80% for carefully designed photodiodes \(^{[22]}\). A photodiode QE is usually driven by several features: (a) the absorption properties of the narrow gap material (the higher the better), (b) the ability of the structure to collect the photogenerated charges (either by diffusion or by drift in the depletion region itself (the optical fill factor (FF) of the pixel will be considered as a part of this collecting feature) and last but not least (c) the incident interface transmission. The last item is generally solved by means of an AR coating and interesting transmissions are usually obtained with a simple quarter wave layer (typically 5% loss).

Due to Beer-Lambert law, the first item (a) is highly depending on material properties (linear absorption coefficient \(\alpha\)), but also on absorbing layer thickness. Direct band gap bulk semiconductors such InGaAs, InSb or HgCdTe give usually a high absorption within a few micron layer thickness such that the thickness is barely limiting. The question of the SL material might be discussed in details to ensure both a high \(\alpha\) and a large enough thickness. In some SL structure, the anisotropy of the material is such that diffusion length perpendicular to the SL layer is weaker than the in-plane diffusion resulting in a weaker QE \(^{[23]}\).

However, the question of the collection efficiency (item b) is more complex as it is strongly depends on the diode technology and structure. Deep mesa structure may result in a loss of fill factor because the lateral etch removes some of the absorbing material. For a given trench QE loss will be more pronounced for small pixel pitches. As an example, a 2 \(\mu\)m wide mesa etch will result in more than 20% loss in FF for a 18\(\mu\)m pixel pitch \(^{[24]}\). This loss might be compensated by smart sidewall reflections to recover a good optical fill factor. Yet, this might induce an extra source of response spatial fluctuations at FPA level. Indeed, the management of uniform mesa profile geometry still remains a hard task. This effect has been put in evidence for QWIPs, comparing structures manufactured by ARL using beveled mesas for optical coupling instead of the classical diffraction gratings \(^{[25]}\). The use of mesa trenches are also critical concerning crosstalk features. Indeed, it might allow to mitigate the diffusion MTF for small pixel pitches, but it has been demonstrated that diffraction onto the trench sharp grounds followed by a guided propagation into contact-window layers might in some cases induce large distance crosstalk issues \(^{[26]}\). At last, the question of the formation of a potential barrier in photodiode heterostructure and its effect on QE has been discussed in the preceding paragraph. Remind that its effect might be more important at low temperatures than at higher temperature due thermionic effect over the barrier. However, in a diffusion extracted photodiode it is difficult to distinguish this effect from the simple contraction of the diffusion length at low temperature; resulting in a loss of fill factor (the collection might be less efficient in the corner of the pixel in case of short diffusion length \(^{[27]}\)). In VLWIR, the materials being less absorbing, thicker layers are usually necessary to maintain high QEs (typical thicknesses are of the order of the cutoff wavelength for HgCdTe for example). In this case, the diffusion length must be large enough to ensure a full collection in the full layer thickness.

VI. CONCLUSION

We have widely discussed the different materials and structures available for space IR detection, in terms of QE and dark current. Two material systems might offer the versatility needed to cover the whole space need in terms of spectral range. HgCdTe has already demonstrated its potentiality in terms of performance and reliability. High lifetimes are obtained resulting in low dark current FPAs for a large range of cutoff wavelengths. However, photodetection structures may still be optimized further to improve the performance or
operating temperature ranges. There was no room in this paper for discussion about the potentiality of HgCdTe APDs for low flux imaging applications. This option might be of first interest and should be discussed elsewhere in this conference [28] [29].

However, the level of maturity is not ready today to fulfill space needs. Apart from the radiation hardness that degrading the lifetime therefore limiting the accessible performances at low temperatures. Moreover, the passivation remains often an issue, particularly for LWIR diodes. Hence, there remains a lot of work to do, both at growth and process levels, to reach the level of performances requested for low flux applications.

**VII. REFERENCES**

[1] M. A. Kinch (2005), *JEM*, vol. 34, no. 6 pp.,
[22] O. Boulade, “Development and characterization of MCT detectors for space astrophysics at CEA”, *this conference*.
[26] O. Gravrand (2008), *JEM*, vol. 37, no. 9, pp. 1205–1211
[27] C. Cervera (2014), “Diffusion length estimator from HgCdTe planar photodiode quantum efficiency measurements”, *presented at QSIP14, to be published in IR physics and technology*
[28] J. Rothman, “HgCdTe APDs for space applications”, *this conference*
[29] E. Deborniol, “SWIR HgCdTe avalanche photodiode FPA performances evaluation”, *This conference*