SAND2015-7653C

Rapid PL Imaging For Defect Identification in Semiconductor Sensors

September 2015

Nathaniel K Grady, Grant Soehnel, Daniel A. Bender

Sandia National Laboratories
1515 Eubank SE
Albuquerque, NM 87123

Rapid PL Imaging For Defect Identification in Semiconductor Sensors

Material defects are a common problem limiting the fabrication of high performance, large area semiconductor detectors. These defects create high rates of carrier recombination which in turn produces pixels in focal plane arrays (FPAs) with high dark currents and other undesirable characteristics. Defects can range from individual pixels to clusters hundreds of pixels or more in size. In this presentation we introduce a new measurement technique based on spatially mapping photoluminescence intensity variations, which we show are strongly correlated with carrier lifetime variations in infrared detectors, which will identify material quality deficiencies and defects. Our primary focus is infrared detectors and materials for creating infrared detectors, but the approach can be used for any direct bandgap semiconductor material. This technique can be applied to sensor layers before patterning or hybridization, allowing bad sensors to be discarded early in the manufacturing process. Near-infrared light is directed onto a sample housed inside a Dewar, which will excite carriers in the sample. For samples containing a direct bandgap semiconductor, radiative recombination in the sample will then produce a photoluminescent return leaving the sample which is imaged with a mid-infrared camera. The contrast in this image provides a map closely related to the carrier lifetime variations in the sample when the incident intensity is in an appropriate range. The optics and camera are mounted on motorized stages while the sample under test is in a fixed position inside the Dewar. This allows us to obtain multiple image tiles that can be stitched together into an image of the entire sample, thereby enabling imaging of samples larger than the system's field of view. In comparison to our previous single point-by-point scanning photoluminescence lifetime system, which was well received at last year's symposium, the present system

allows images with more than 10x high resolution to be obtained in less than 1/100 of the time.

Key Words: Minority Carrier Lifetime, Photoluminescence, Mercury Cadmium Telluride (MCT), Dark Current, Defects, Infrared FPAs

1.0 Introduction

Material defects are a common problem limiting the fabrication yield and performance of large area semiconductor detectors. These defects cause high rates of carrier recombination which in turn produces pixels in focal plane arrays (FPAs) with high dark currents and other undesirable characteristics. Defects can range from individual pixels to clusters hundreds of pixels or more in size. Here we introduce a new measurement technique based on spatially mapping photoluminescence intensity variations, which we show are strongly correlated with carrier lifetime variations in infrared detectors, to identify material quality deficiencies and defects. Our primary focus is infrared detectors and materials for creating infrared detectors, but the approach can be used for any direct bandgap semiconductor material. This technique can be applied to sensor layers regardless of patterning or hybridization, allowing bad sensors to be discarded early in the manufacturing process. Further, our technique could aid in determining if a bad part of the FPA is due to the detector or electronics (*e.g.* ROIC, bonding). In comparison to our previous single point-by-point scanning photoluminescence lifetime system, the present system allows images with more than 30x high resolution to be obtained in $1/100^{th}$ the time.

When a semiconductor absorbs a photon with an energy exceeding the semiconductor bandgap, an electron from the valence band (E_v) is promoted to the conduction band (E_c), leaving a hole in the valance band (*i.e.* forms an electron-hole pair exciton) as shown in Fig. 1 [1, 2]. These excited carriers are responsible for the photocurrent in a semiconductor photodiode. However, some excitons will spontaneously recombine before they can contribute to the photo current through three processes: (1) radiative recombination, (2) Shockley-Read-Hall (SRH) recombination, and (3) Auger recombination. This first process is responsible for the photoluminescence emission we exploit. Minimizing the rate of these processes is critical to obtaining high photodiode performance as (1) excited carriers that decay through these processes do not contribute to the photocurrent, (2) the random nature of these processes contributes the noise, and (3) the dark current increases with increasing SRH and Auger recombination rates.

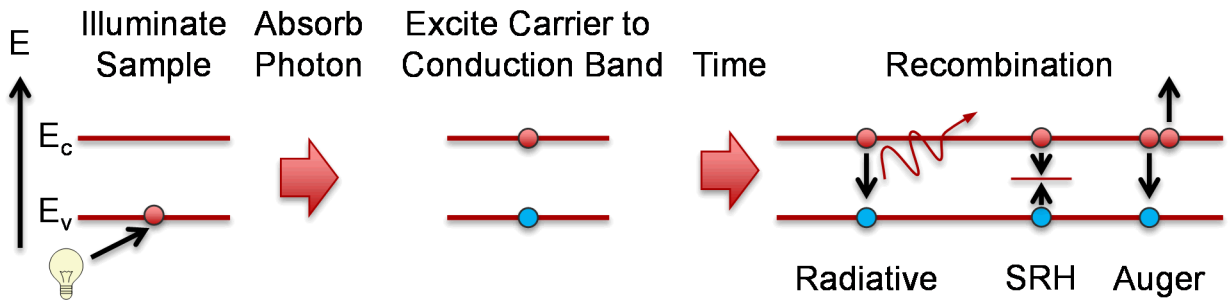

Figure 1: Diagram illustrating excitation of carriers in a semiconductor through absorption of a photon and subsequent recombination.

2.0 Carrier Recombination Model

The observed photoluminescence is due to radiative recombination across the semiconductor bandgap. We developed a numerical model for the time-dependent carrier concentration upon excitation by light as part of our previous investigation into using PL lifetime mapping to identify in semiconductor sensors [3]. Essentially, the carrier concentration of electrons n and holes p after a time step Δt is determined by the

relative rates of carrier generation and recombination through radiative, SRH, and Auger processes as described in Eqs. (1) and (2):

$$n(t+\Delta t) = n(t) + \Delta t G_{optical}(t) - \Delta t(U_{aug} + U_{rad} + U_{SRH}) \quad (1)$$

$$p(t+\Delta t) = p(t) + \Delta t G_{optical}(t) - \Delta t(U_{aug} + U_{rad} + U_{SRH}) \quad (2)$$

where $G_{optical}(t)$ is the optical carrier generation rate, and U_{aug}, $Urad$, U_{SRH} are the net recombination rates from auger, radiative, and SRH recombination. Both pulsed laser excitation and CW excitation can be modeled through $G_{optical}(t)$. The recombination rates are calculated using Eqs. (3)-(5) [4], where n_0 and p_0 are the equilibrium concentrations of electrons and holes, n_i is the intrinsic carrier concentration, and n_1 and p_1 are the SRH densities (set equal to n_0 and p_0 because the exact nature of the defect states present is neither known nor critical for this application). τ_{SRH-n} and τ_{SRH-p} are the SRH lifetimes for electrons and holes, which are set equal to each other and regarded as the primary variable of interest. B, C_n, and C_p are recombination rate coefficients given by Eqs. (7)-(9) [5, 6, 7], m_0 is the mass of an electron, m_c^* and m_v^* are the effective masses of electrons and holes (in the conduction and valence bands), and $|F_1F_2|$ is the Bloch Function overlap integral which we set to 0.2 as its value is not precisely known.

$$U_{aug} = C_n(n^2p - n_0^2p_0) + C_p(np^2 - n_0p_0^2) \quad (3)$$

$$U_{rad} = B(np - n_i^2) \quad (4)$$

$$U_{SRH} = \frac{(np - n_i^2)}{\tau_{SRH-n}(p + p_1) + \tau_{SRH-p}(n + n_1)} \quad (5)$$

$$n_0p_0 = n_i^2 \quad (6)$$

$$B = 5.8\times10^{-13}\sqrt{\varepsilon_\infty}\left(\frac{m_0}{m_c^* + m_v^*}\right)^{\frac{3}{2}}\left(1 + \frac{m_0}{m_c^*} + \frac{m_0}{m_v^*}\right) \times\left(\frac{300}{T}\right)^{\frac{3}{2}}\left(E_g^2 + 3kTE_g + 2.75k^2T^2\right) \quad (7)$$

$$C_n = \frac{\left(\frac{m_c^*}{m_0}\right)|F_1F_2|^2}{2n_i^2(3.8\times10^{-18})\varepsilon_\infty^2\left(1 + \frac{m_c^*}{m_v^*}\right)^{\frac{1}{2}}\left(1 + \frac{2m_c^*}{m_v^*}\right) \times\left(\frac{E_g}{kT}\right)^{\frac{-3}{2}}\exp\left(-\frac{1 + \frac{2m_c^*}{m_v^*}}{1 + \frac{m_c^*}{m_v^*}}\frac{E_g}{kT}\right)} \quad (8)$$

$$C_p = C_n \left[\frac{1 - \frac{3E_g}{kT}}{6\left(1 - \frac{5E_g}{4kT}\right)} \right] \tag{9}$$

The experiments here focus on devices where the active layer is $Hg_{(1-x)}Cd_xTe$ (MCT), for which the intrinsic carrier concentration n_i and band gap E_g depend on temperature T and Cd mole fraction x following Eqs. (10) and (11) [8, 9]. Finally, the equilibrium carrier concentrations are given by Eqs. (12) and (13), where N_D and N_A are the donor and acceptor doping concentrations [2].

$$E_g = -0.302 + 1.93x + 5.35\times10^{-4}T(1-2x) - 0.81x^2 + 0.832x^3 \tag{10}$$

$$n_i = (5.585 - 3.82x + 1.753\times10^{-3}T - 1.364\times10^{-3}xT) \times 10^{14} E_g^{\frac{3}{4}} T^{\frac{3}{2}} \exp\left(\frac{-E_g}{2kT}\right) \tag{11}$$

$$n_0 = \frac{N_D - N_A}{2} + \sqrt{\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2} \tag{12}$$

$$p_0 = \frac{N_A - N_D}{2} + \sqrt{\left(\frac{N_A - N_D}{2}\right)^2 + n_i^2} \tag{13}$$

The simulation is run by setting the initial carrier concentrations to the equilibrium concentrations, then applying light to excite the material producing carriers ($G_{optical(t)}$). It is assumed that the illumination is circular and of uniform intensity and that charge diffusion rapidly spreads this excitation evenly through the depth of the material such that carrier concentration is always treated as uniform in the illuminated volume of the material. The devices measured were p-n photodiodes, but since they are left unbiased during the experiment and the p-type implant takes up a very small volume in the n-type absorber, only the n-type material was simulated. Eqs. (1) and (2) are used to simulate the time-dependent change in carrier concentration which will either reach and equilibrium with the excitation rate in the case of constant illumination or rapidly rise concentration during the laser pulse and subsequent decay back to the equilibrium concentration. Figure 2 shows the result of the simulation using a CW illumination (λ = 2.5 μm) of MCT at 100 K with a composition x = 0.287 for several input intensities. Note that the y-axis is arbitrary as several scale factors such as the collection efficiency of the optics is not precisely known. This graph indicates that the incident intensity must be selected such that the variation in the non-radiative recombination rate results in a significant change in the carrier concentration. At high CW intensities, such that the generation rate $G_{optical}$ is very high compared to the spontaneous decay rates, carriers will almost immediately be re-excited and the decay rates therefore have only a small influence on the observed PL signal. At low CW intensities, the PL signal will be weak and therefore difficult to measure. This did not prove to be a problem for this experiment as the PL intensity is easily observable in the regime where the carrier lifetime produces a strong variation in PL signal.

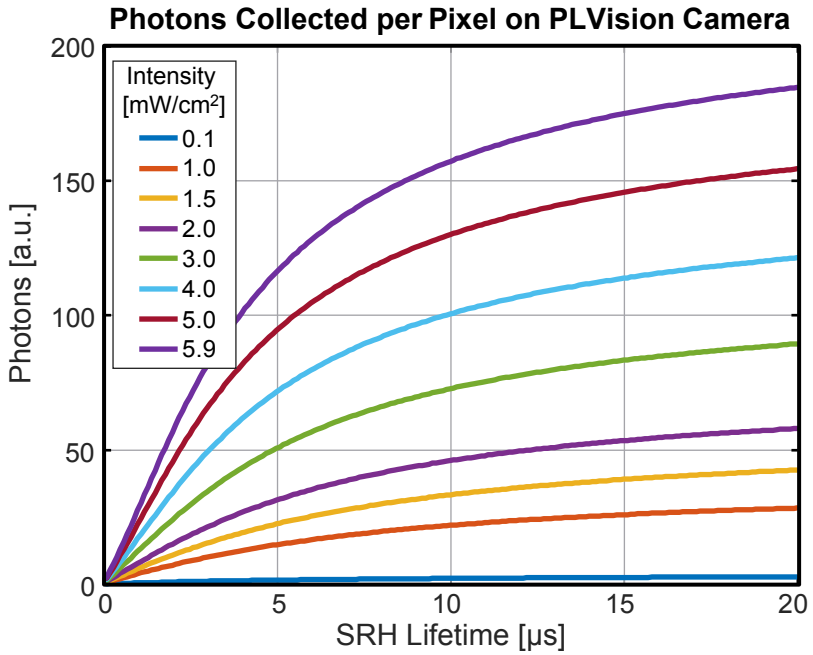

Figure 2: Carrier recombination simulation with CW excitation for MCT at 100K.

3.0 Experiment

3.1. Overview

As shown in Fig. 3, near-infrared light from a 150 W QTH lamp is directed onto a sample housed inside a Dewar through a fiber ring illuminator, exciting carriers in the sample. The standard microscope fiber ring illuminator surrounds the collection lens, and produces a hollow cone of light that, due to symmetry, results in an even illumination of the sample over the system's field-of-view (FOV). For samples containing a direct bandgap semiconductor like MCT, radiative recombination will then produce a photoluminance return leaving the sample which is imaged with a mid-infrared InSb camera. The contrast in this image provides a map closely related to the carrier lifetime variations in the sample. The optics and camera are mounted on motorized stages while the sample under test is in a fixed position inside the Dewar. Moving the optics and camera allows us to obtain multiple image tiles that can be stitched together into an image of the entire sample, thereby enabling imaging of samples larger than the system's field of view. We found that using a fiber ring illuminator to direct light onto the sample provides a nearly dark-field measurement: light reflected from a flat sample will emerge at an equal and opposite angle to the sample normal, hitting the opposite side of the ring illuminator rather than being captured by the imaging optics. Only scattered light from features protruding from the sample and PL generated by the illumination are collected by the imaging optics. Using a long-pass filter in the imaging optics to further reject the excitation light did not significantly change the PL image, but did reduce the intensity and increase the background due to thermal self-emission, so one was not employed for the measurements presented here.

For comparison, we previously characterized the PL lifetime used a time-consuming point-by-point time-resolved mapping of the photoluminescence and the dark current of each pixel. Due to the slow acquisition rate, the lifetime maps are limited to much lower resolution. Nonetheless, it shows that the

time-averaged CW PL provides similar information in the case of a typical FPA. Note that for some exotic devices such as nanoantenna-enhanced detector, the strong modification of the absorption and emission properties by the antennas will strongly influence the observed CW PL intensity and the obtained image contrast may not correlate to lifetime variations.

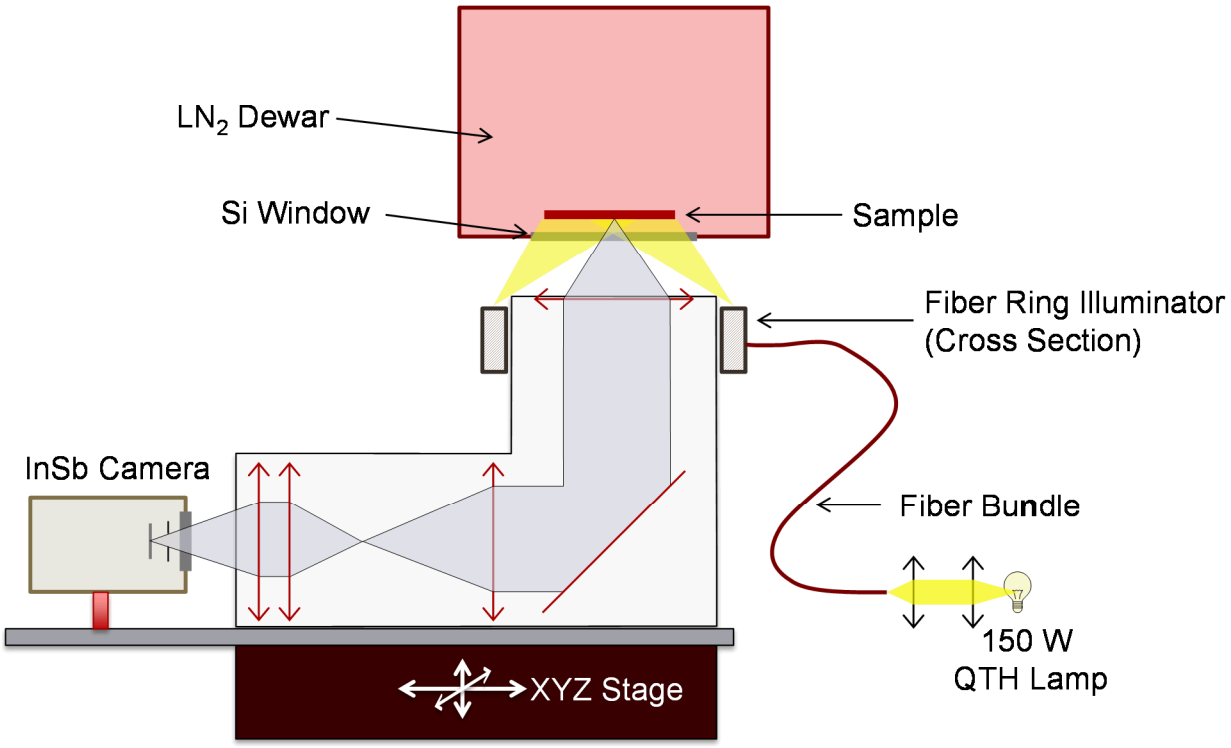

Figure. 3: Schematic diagram of the PLVision apparatus.

3.2. IR camera

We used a commercially-available 640x512 pixel InSb FPA with a compact closed-cycle cryocooler for operation at ~ 80K. The camera has a custom AR coated sapphire window to transmit ~1 to 5 µm. The spectral range of this camera limits us to studying arrays with a cut-off wavelength < 5 µm as the observed band-edge PL is emitted at approximately the same wavelength as the cutoff of the detector being tested. In the current experiment, we looked at SWIR detectors that emitted around 3.6 µm.

3.3. Lens design

Designing a suitable lens for PLVision is quite challenging. There are several constraints on the design of the lens: (1) the fixed cold stop in the InSb camera, (2) the modest maximum mass that could be mounted on the translation stages used to select the area of the sample imaged, (3) a strong preference for using available off-the-shelf lenses, and (4) keeping distortion and vignetting low enough to stitch images together without obvious errors in the overlapping regions. The 1x magnification f/4.1 lens we developed is shown in Fig. 4. A re-imaging approach was chosen because it enabled a high cold shield efficiency (i.e. low vignetting) and object-side telecentricity (no magnification change with focus). The latter was desirable because the active layer was buried a significant distance into the sample and the surrounding electronics were also at a different height. With a non-telecentric design, these features would appear at a different size making evaluation of the images more difficult. The effective f/number was limited by the

FPA cold stop. For 1 in diameter off-the-shelf optics and a 1x magnification, the loss of effective objective lens size that is the primary downside of a telecentric design was not an issue. For room-temperature measurements, a much simpler lens would suffice as there would be no need to correct for the aberration due to imaging through a thick Si window. While the optimized design of the imaging optics, shown in Fig. 5, provided a 1x magnification, mechanical limitations in mounting the fiber ring illuminator around the objective lens forced us to operate with the objective further from the cryostat window than we planned and the actual system magnification was ~ 0.8x. The performance, as can be seen in the images of a resolution test target presented in Fig. 5, is nonetheless quite good: group 4, elements 4 and 5, visible in Fig. 5(c), correspond to 23 and 25 lp/mm, respectively, or just over 2 px/lp.

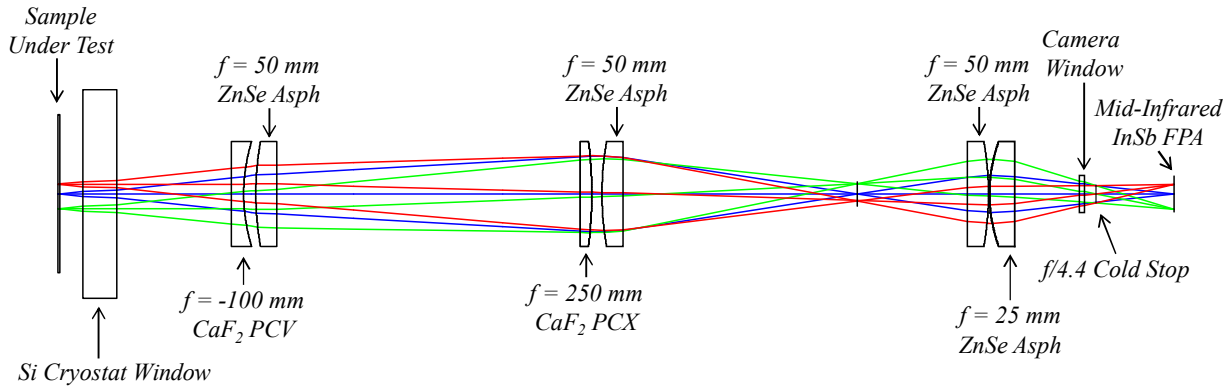

Figure 4. Imaging optics.

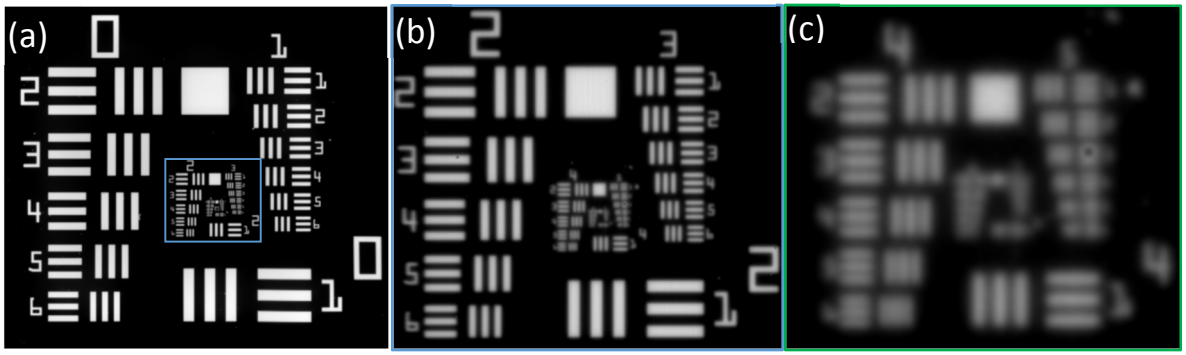

Figure 5. Images of USAF-1951 test target with PLVision system. No additional light source beyond ambient and thermal self-emission was used.

4.0 Image Processing

4.1. Flat-Fielding / Background Subtraction

IR cameras, including the InSb camera used here, exhibit significant fixed pattern noise. In addition, the room temperature optics, cryostat window, and reflection from the sample contribute a significant background. Non-uniformity correction to deal with these issues is a significant area of research. Most such research focuses on more traditional applications such as remote sensing from aerospace platforms and handheld thermal imagers and may not perform as well on the "scenes" imaged in our PL system. Simply imaging a black-body at several temperatures is insufficient in this case as the environment of the

sample both includes coatings (AR coating, CdZnTe), the surrounding electronics and mounting structure, and the cryostat itself. We assume that the background is the constant part over all images of different positions on the sample. Initially, we tried just averaging or taking the median of all frames, throwing out some of the frames with large amounts of structure manually. For the more interesting samples, many bright features are arranged non-uniformly around the sample making it difficult or impossible to find frames to average together which did not result in strong features significantly influencing the average. Instead, we used a robust principle components analysis (RPCA) scheme, specifically Low Rank Recovery (LRR) by Robust Orthonormal Subspace Learning (ROSL) [10] as implemented in the open source Matlab LRSLibrary [11, 12]. This is a somewhat arbitrary choice; other algorithms we tried, such as ssGoDec [13], worked equally well. Approaches more commonly used in the IR imaging literature, such as Kalman filters, are another possible solution but most implementations are tuned assuming a large sequence of images with a much smaller translation step between them as would be obtained from, for example, a hand-held video-rate thermal imaging camera being panned across a scene.

4.2. Montage Formation

Even low levels of distortion and residual illumination inhomogeneities can become obvious in the stitched images: for a sharp-edged feature, say a sharp line, split across two frames, even a 1 or 2 px error is obvious. Given a 512x512 px camera, stitching images with a 20% overlap (~ 102 px), a lens with 1% distortion over the full field corresponds to a ~2 px error between the appearance of such a structure in the two images! Registration is also critical as small errors are obvious when montaging; the stage we used was quite accurate, but the close-cycle cooler on the camera resulted in many frames being shifted 1-2 px. For the images presented here, we found the ImageJ MIST [14, 15] plugin generally produced reasonable results. We then processed the resulting images in Matlab to produce the final false-color images.

5.0 Results

We studied two similar SWIR MCT p-n diode FPAs. FPA #1 has a small useful area with the rest of the array suffering from many hot pixels and therefore serves as a good demonstration of our techniques ability to highlight poor areas of an array.

5.1. PL Intensity Compared to Lifetime, Dark Current, and Relative Response

In Fig. 6 we present images comparing 38x38 mm images of the PL lifetime, intensity, and dark current maps for FPA #1. In all three cases the color scale has been chosen such that bright areas are relatively good parts of the array; note that the dark current is presented as -1 x $\log_{10}(I_{dark})$ in order to aid comparison by making bright colors in all three images correspond to better areas of the array. The 68 x 68 px, 0.5 μm spot size lifetime map in Fig. 6(a) took ~ 3 hours to collect, while the 2.2k *x* 2.2k px intensity map in Fig. 6(b) took < 5 mins to collect. In the latter case, less than a second of this time was actual image integration time, with most of the time actually taken moving the stage between frames and saving each frame with less. We also show the lifetime and intensity maps for a 2nd array in Fig. 7.

Excellent correspondence between the major features can be seen in all three images, proving that this approach provides a viable method for locating problematic areas of the array. Only features in the dark current due to the detector itself are mirrored in the PL images; a few features that are electronic in origin, such as the vertical bar in the upper half of the image and the overall difference between the top and bottom half of the FPA, only appear in the dark current map. This indicates another useful aspect of the PL maps – they can help identify the origin of problems with the finished detector.

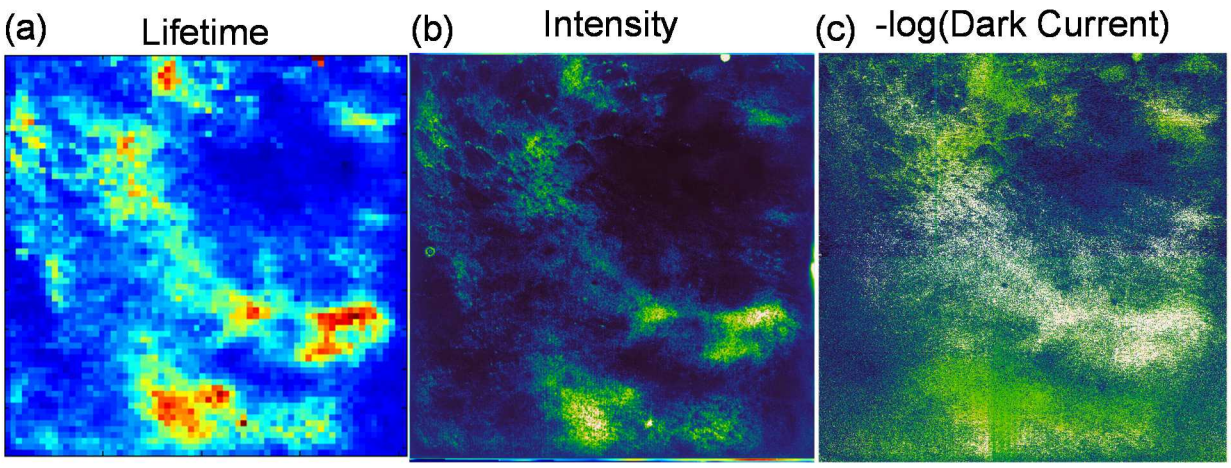

Figure 6: (a) SRH lifetime from point-by-point time-resolved PL mapping (bright = long lifetime), (b) CW PL image (bright = high intensity), and (c) -1*log$_{10}$ of the dark current (bright = *low* dark current) of SWIR MCW FPA #1. The sample was cooled to 110 K for (a) and (b), and 100 K for (c).

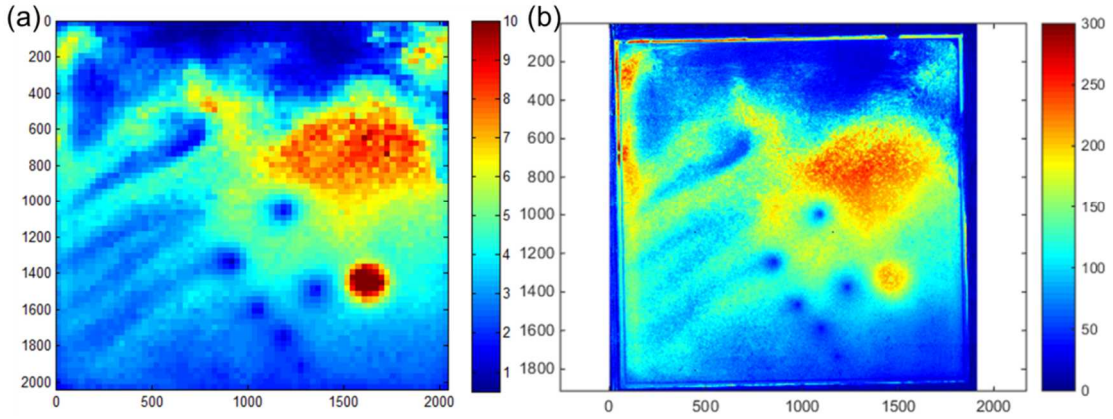

Figure 7: (a) lifetime and (b) PL Intensity for FPA #2.

5.2. Temperature Dependence

So far, all the measurements described were performed with the samples cooled to 110 K with LN_2. However, this presents a difficulty for practical application during the manufacture of FPAs both due to the need to mount the sample in a cryostat and the several hours it takes to evacuate the cryostat to high vacuum levels and safely cool to cryogenic temperatures. We therefore measured the PL images with the samples at temperatures ranging from 110 K to room temperature (295 K), shown in Fig. 8. While the PL intensity is reduced and the thermal background increases at high temperatures, the results indicate that this measurement can be performed at room temperature. The thermal population of carriers is higher, which increases the relative relaxation rates (both SRH and Auger recombination depend on the carrier density), and therefore reduces the PL intensity. Nonetheless, the areas with significantly higher defect densities are still clearly identifiable and the increased background would be the same for all devices in a screening application. At 200 K (-73 C), the result is nearly as good as at 110 K. This is a much easier temperature to work at: lab freezers that hold large volumes at -80 C are common, a dry N_2 or air environment can be used instead of high vacuum (saving pump-down time), and the entire optical assembly could easily be placed in the -80 C environment which would significantly reduce the thermal

background. Traditional electrical characterization of the dark current, noise, and relative response have to be carried out in at the target operating temperature for the FPA (e.g. LN_2 temperatures for the FPAs presented here), making this another significant advantage of this technique.

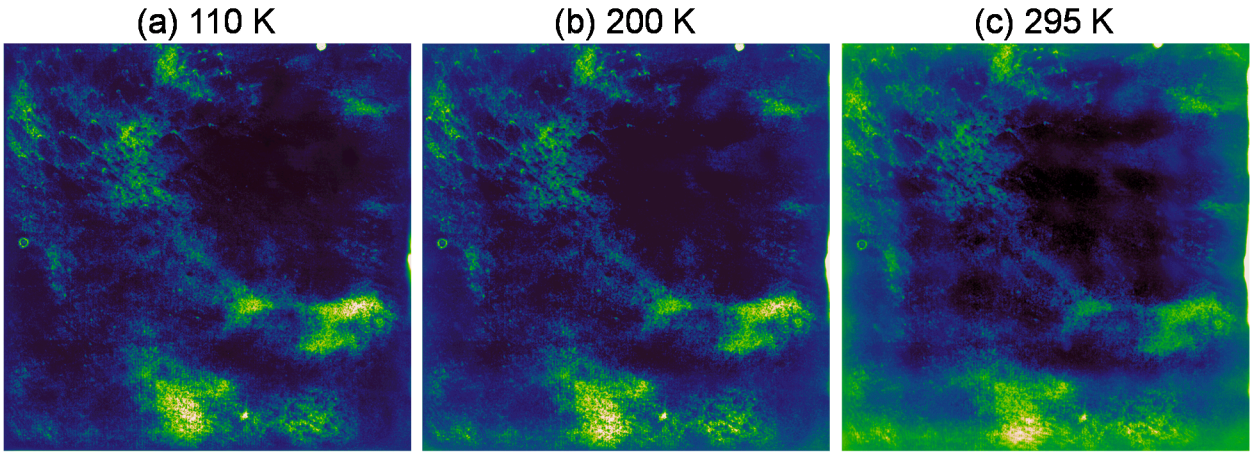

Figure 8: PL intensity images taken with the sample at three different temperatures.

6.0 Summary

We have introduced a new technique for defect mapping in infrared FPAs based on rapidly imaging the photoluminescence intensity. The results presented here were for finished FPAs, but the PL maps do not depend on any electrical properties of the array and could be applied at other steps in the manufacturing process; for example, to identify good wafers, or decide which detectors to hybridize to a ROIC. This technique can be applied under ambient conditions, allowing wafers, detector dies, and completed FPAs to be rapidly screened in a few minutes instead of hours or days.

7.0 References

[1] S. Banerjee and B. Streetman, Solid State Electronic Devices, Upper Saddle River, NJ: Prentice Hall, 2000.

[2] S. Sze and K. Ng, Physics of Semiconductor Devices, Hoboken, NJ: John Wiley & Sons, 2007.

[3] G. Soehnel, "Time Resolved Photo-luminescent Decay Measurements of Infrared Materials," in *2014 Meeting of the Millitary Sensing Symposia (MSS) Specialty Groups on Passive Sensors, Battlefield Survivability & Discrimination, Detectors, and Materials*, Springfield, VA, 2014.

[4] S. Rein, Lifetime Spectroscopy, Berlin, Germany: Springer, 2005.

[5] G. L. Hansen, J. L. Schmit and T. N. Casselman, "Energy gap versus alloy composition and temperature in Hg1−x Cd x Te," *J. Appl. Phys.*, vol. 53, no. 10, pp. 7099-7101, 1982.

[6] P. Saxena, "Modeling and simulation of HgCdTe based p+-n-n+ LWIR photodetector," *Infrared*

Physics and Technology, vol. 54, pp. 25-33, 2011.

[7] G. Hansen and J. Schmidt, "Calculation of intrinsic carrier concentration in $Hg_{(1-x)}Cd_{(x)}Te$," *J. Appl. Phys.,* vol. 54, pp. 1639-1640, 1983.

[8] A. Itsuno, J. Phillips and S. Velicu, "Predicted performance improvement of auger-suppressed HgCdTe photodiodes and p-n heterojunction detectors," *IEEE Transactions on Electron Devices,* vol. 58, pp. 501-507, 2011.

[9] P. Chakrabarti, P. Saxena and R. Lal, "Analytical simulation of an InAsSb photovoltaic detector for mid-infrared applications," *Int. J. Infrared. Milli. Waves.,* vol. 27, pp. 1119-1132, 2006.

[10] X. Shu, F. Porikli and N. Ahuja, "Robust Orthonormal Subspace Learning: Efficient Recovery of Corrupted Low-rank Matrices," in *Proc. of International Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, Ohio, 2014.

[11] A. Sobral, T. Bouwmans and E.-h. Zahzah, "LRSLibrary: Low-Rank and Sparse tools for Background Modeling and Subtraction in Videos," in *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, CRC Press, Taylor and Francis Group, 2015.

[12] A. Sobral, "Low-Rank and Sparse Tools for Background Modeling and Subtraction in Videos," 2015. [Online]. Available: https://github.com/andrewssobral/lrslibrary. [Accessed Oct. 2015].

[13] T. Zhou and D. Tao, "GoDec: Randomized Low-rank & Sparse Matrix Decomposition in Noisy Case," in *28th International Conference on Machine Learning*, Bellevue, Washington, 2011.

[14] T. Blattner, "Image Stitching," The National Institute of Standards and Technology (NIST), 2014. [Online]. Available: https://isg.nist.gov/deepzoomweb/resources/csmet/pages/image_stitching/image_stitching.html. [Accessed Sept. 2015].

[15] T. Blattner, J. Chalfoun, B. Stivalet, M. Brady and S. Zhou, "A Hybrid CPU-GPU System for Stitching Large Scale Optical Microscopy Images," in *43rd International Conference on Parallel Processing (ICPP)*, Minneapolis, MN, 2014.