

Information-Theoretic Criterion for the Performance of Single-Photon Avalanche Photodiodes

David A. Ramirez, Majeed M. Hayat, *Senior Member, IEEE*, Sergio N. Torres, Bahaa E. A. Saleh, *Fellow, IEEE*, and Malvin C. Teich, *Life Fellow, IEEE*

Abstract—A channel-capacity metric is introduced for assessing the performance of single-photon avalanche photodiodes (SPADs) when used as detectors in laser communication systems. This metric is employed to theoretically optimize, with respect to the device structure and operating voltage, the performance of SPADs with simple InP or $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -InP heterojunction multiplication regions. As the multiplication-region width increases, an increase is predicted in both the peak and the full-width at half-maximum of the channel capacity curve versus the normalized excess voltage. Calculations also show the existence of an optimal $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -InP heterojunction multiplication region that maximizes the peak channel capacity beyond that of InP.

Index Terms—Avalanche photodiodes (APDs), channel capacity, dead space, detection efficiency, Geiger mode, heterostructures, impact ionization, InAIAs, InP, single-photon detection.

I. INTRODUCTION

A COMMONLY used performance metric for the performance of Geiger-mode avalanche photodiodes (APDs), also known as single-photon avalanche photodiodes (SPADs), is the single-photon quantum efficiency (SPQE), which is also referred to as the detection efficiency [1]. Another important metric is the dark-count probability P_d , which is one of the key factors affecting the SPQE. These metrics are well accepted and established for sensing applications, where the objective is to assess the capability of the SPAD in detecting and counting individual photons by examining the detection and false-alarm probabilities.

Today, however, there are emerging applications in laser communications that require the use of SPADs as detectors [2], as they are used to determine the presence or absence of a photon in individual bits of optically transmitted binary data. For these applications, we can view the SPAD as a binary channel, in which case the channel capacity would be the natural metric for assessing the performance of the SPAD. In this context, the relevance of the channel capacity as a performance metric is clear:

Manuscript received June 7, 2005; revised July 7, 2005. This work was supported by the National Science Foundation under Award ECS-0334813 and Award ECS-0428756. The work of D. A. Ramirez and S. N. Torres was supported by Fundación Andes Grant C-13855 and by Milenio ICM Grant P02-049.

D. A. Ramirez and S. N. Torres are with the Department of Electrical Engineering, University of Concepción, Casilla 160-C, Concepción, Chile (e-mail: davramir@gmail.com; storres@die.udec.cl).

M. M. Hayat is with the Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87131-1356 USA (e-mail: hayat@ece.unm.edu).

B. E. A. Saleh and M. C. Teich are with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215-2421 USA (e-mail: besaleh@bu.edu; teich@bu.edu).

Digital Object Identifier 10.1109/LPT.2005.856406

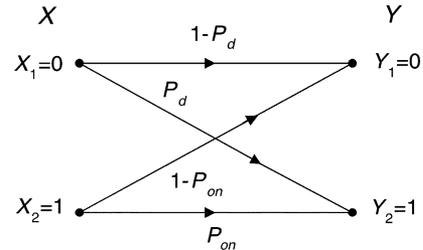


Fig. 1. SPAD communication channel.

It represents the maximum rate of information we are allowed to transmit while maintaining the ability to reduce the bit-error probability to an arbitrarily small value through proper modulation and encoding for error correction [3]. Indeed, for certain deep-space communication systems, the current state of the art in single-photon detection is inadequate to achieve acceptable bit-error probabilities [2]. Reliable operation of such systems would require effective error correction and modulation of the data to be transmitted. Thus, knowledge of the channel capacity of an SPAD is crucial since it provides the fundamental limit of the rate of information-transmission above which reliable operation is impossible.

In this letter, we introduce the notion of channel capacity for single-photon detectors and utilize it to optimize the performance of SPADs with simple (InP) or heterojunction ($\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -InP) multiplication regions. The optimization is performed over the operating voltage and the width of the multiplication region. It is shown theoretically that if the operational reverse-bias voltage of the SPAD is optimized with respect to the SPQE, the channel capacity achieved at that voltage is significantly below the maximum achievable value, which can be realized at a different voltage.

II. CHANNEL CAPACITY

Consider the binary channel shown in Fig. 1 representing the input-output relationship of an SPAD. We call this channel the *SPAD channel*. The random variables $X \in \{0, 1\}$ and $Y \in \{0, 1\}$ represent, respectively, the possible input and output alphabets. In particular, $X = 0$ represents the case when no optical power is present in a bit (zero signal) while $X = 1$ corresponds to the case when an optical power is present. Similarly, $Y = 0$ means that no avalanche breakdown has occurred during the detection period (bit) of the SPAD while $Y = 1$ represents the occurrence of an avalanche breakdown. With these definitions, the transition probabilities of the SPAD channel can be identified with the total detection probability and the dark-count probability, which are described as follows. The total detection

probability P_{on} is the probability that either a dark carrier or a photon triggers an avalanche breakdown given that there is an optical pulse during the detection period. It is given by $P_{\text{on}} = 1 - e^{-(\eta P_a N_o + Q_a N_d)}$, where P_a is the probability of an avalanche breakdown caused by a photocarrier that is injected into the multiplication region, Q_a is the probability that a dark carrier (generated at a random location in the multiplication region) triggers an avalanche breakdown, and N_d is the average number of dark carriers generated in the multiplication region during the detection time [4]. Here, the average number of photons per pulse is denoted by N_o , and η is the quantum efficiency of the SPAD. The dark-count probability P_d is given by $P_d = 1 - e^{-Q_a N_d}$ [1]. We now observe that $\text{P}\{Y = 1 | X = 0\} = P_d$, $\text{P}\{Y = 1 | X = 1\} = P_{\text{on}}$, etc., as shown in Fig. 1. In this letter, the breakdown probabilities P_a and Q_a are calculated according to the dead-space multiplication theory [4], [5].

According to principles of digital communication, before defining the channel capacity we must define the mutual information of the SPAD channel, given by [3]

$$\mathcal{I} = \sum_{j=0}^1 \sum_{i=0}^1 \text{P}\{X = j, Y = i\} \frac{\text{P}\{X = j | Y = i\}}{\text{P}\{Y = i\}} \quad (1)$$

where the output-alphabet probabilities are given by $\text{P}\{Y = 0\} = (1 - P_d)\pi_0 + (1 - P_{\text{on}})(1 - \pi_0)$ and $\text{P}\{Y = 1\} = P_d\pi_0 + P_{\text{on}}(1 - \pi_0)$, and $\pi_0 \triangleq \text{P}\{X = 0\}$ is the *a priori* input-symbol probability. Upon substitution, the mutual information for the channel shown in Fig. 1 reduces to

$$\begin{aligned} \mathcal{I} = & \pi_0(1 - P_d) \log_2 \left(\frac{1 - P_d}{(1 - P_d)\pi_0 + (1 - P_{\text{on}})(1 - \pi_0)} \right) \\ & + \pi_0 P_d \log_2 \left(\frac{P_d}{P_d\pi_0 + P_{\text{on}}(1 - \pi_0)} \right) \\ & + (1 - \pi_0)(1 - P_{\text{on}}) \log_2 \left(\frac{1 - P_{\text{on}}}{P_d\pi_0 + P_{\text{on}}(1 - \pi_0)} \right) \\ & + (1 - \pi_0) P_{\text{on}} \log_2 \left(\frac{P_{\text{on}}}{P_d\pi_0 + P_{\text{on}}(1 - \pi_0)} \right). \quad (2) \end{aligned}$$

The channel capacity C with units of bits (or more generally, bits per symbol), can now be defined as the maximum of the mutual information \mathcal{I} over all $0 \leq \pi_0 \leq 1$, i.e., $C \triangleq \max_{0 \leq \pi_0 \leq 1} \mathcal{I}$. It also follows that $0 \leq C \leq 1$ [3].

III. RESULTS

In our calculations, we focus on the dark-carrier generation and breakdown only in the high electric field multiplication region of the SPAD. Band-to-band tunneling of carriers is considered as the dominant source of dark current in the multiplication layer [4], [6]. Throughout, we assume that the SPAD is operated in a gated mode, whereby it is dc-biased just below its breakdown voltage V_{BR} during the off-mode, and then pulse-biased above V_{BR} during the on-period. We assume a 2-ns on-pulse with repetition rate of 500 kHz [6], $N_o = 0.1$ photons, and $\eta = 0.5$ [7].

The results of the calculations for an SPAD with a simple hole-injected InP multiplication region are shown in Fig. 2, which depicts the channel capacity as a function of the normalized excess voltage. (The normalized excess voltage is defined

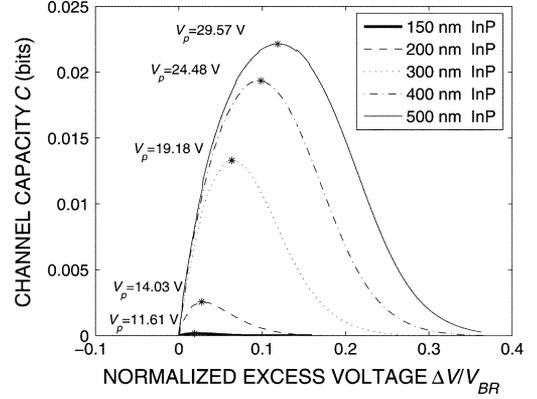


Fig. 2. Channel capacity corresponding to InP multiplication region as function of the normalized excess voltage for each width. The reverse-bias voltage corresponding to the peak channel capacity V_p is also shown for each width.

as $\Delta V/V_{\text{BR}}$, where $\Delta V = V - V_{\text{BR}}$ and V is the voltage across the multiplication region.) The peak channel capacity (shown by the symbol “*”) and the corresponding reverse-bias voltage V_p are shown for each width. Clearly, the peak channel capacity increases as the multiplication-region width increases. The key factor causing this performance enhancement is the reduction in the breakdown electric field (as the width is increased), which, in turn, leads to a reduction in tunneling dark carriers [6]. Another advantage of increasing the width of the multiplication region is reduced sensitivity to variation in the applied voltage. This is measured by the full-width at half-maximum (FWHM) of the channel capacity curve versus the normalized excess voltage; the FWHM increases as the multiplication-region width increases.

The trend in the channel capacity, as a function of the multiplication-region width, is similar to that of the SPQE observed in our earlier studies [4]. However, the voltage at which the channel capacity reaches its optimal value (peak channel capacity) is different from that corresponding to the optimal SPQE. For example, for a 500-nm multiplication region, the peak SPQE occurs at 32 V (according to the model reported in [4]) while the peak channel capacity occurs at 29.6 V. Notably, if we reverse bias the SPAD at 32 V, the channel capacity would be 40% below its peak value. Thus, it can be inferred that optimizations with respect to the SPQE and the channel capacity are not equivalent; SPAD optimization is, therefore, application (sensing versus communication) dependent.

We next consider an SPAD with an $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -InP heterojunction multiplication region, which has the same general structure of the heterostructure APDs considered for minimizing the excess noise factor [8]. For this type of SPAD, photogenerated holes are injected into the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer, which is termed the energy-buildup layer. Our calculations indicate that for a fixed total width of the heterojunction multiplication region (consisting of both layers), the peak channel capacity depends on the fraction of the width occupied by the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer. The peak channel capacity (at the optimal applied voltage) starts at a minimum when the multiplication region is entirely made of $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$. As the fraction of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -layer-width decreases, the peak channel capacity increases, reaching a maximum value, after which it decreases monotonically. This behavior is shown in Fig. 3 and

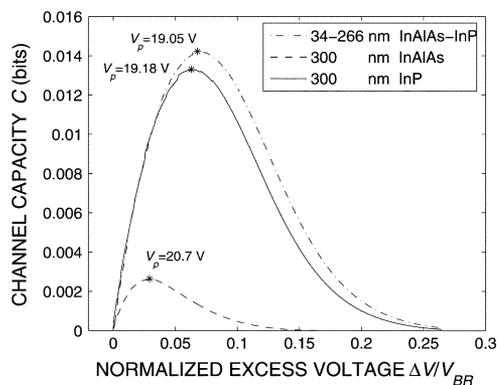


Fig. 3. Channel capacity corresponding to InP, $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$, and optimal $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -InP multiplication regions of width 300 nm. The reverse-bias voltage corresponding to the peak channel capacity V_p is also shown for each multiplication region. The 34-266-nm heterostructure is the optimal structure among all other width combinations that have a total width of 300 nm.

it is attributable to the dead-space effect [4]. Thus, for a fixed total multiplication-region width, there is an optimal width of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer that maximizes the peak channel capacity. For example, for total multiplication-region widths of 300 and 200 nm, the optimal widths of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer are 34 and 33 nm, respectively. However, the channel capacity in this type of heterojunction is less sensitive to variations in the width of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer compared to variations in the applied voltage. Thus, a small change in the width of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer (i.e., within 4 nm) from its optimal value will not result in a significant change in the channel capacity.

The peak channel capacities calculated are indeed very small (i.e., $\ll 1$ bits), which is a consequence of the ultraweak nature of the received signal (recall that $N_o = 0.1$, corresponding to a subphoton level), which results in a high “miss” probability (e.g., $P\{Y = 0 | X = 1\} = 0.96$ in the case of a 300-nm InP SPAD operated at 19.2 V). If N_o is risen to 10, for example, the peak-channel capacity increases to approximately 0.9 bits.

For a given total width of the multiplication region, the highest peak channel capacity occurs when the width of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer is between the hole’s dead spaces in InP and $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$, at the breakdown field. (A similar behavior was shown earlier in the case of the SPQE [4].) This behavior can be explained in the context of the initial-energy effect [4], [9]: The high bandgap $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer serves to energize (or heat) the injected holes before they enter the low-bandgap InP layer. If the width of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer is just below the hole dead space in $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$, then the holes will energize without ionizing in the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer. However, once they enter the InP layer, they require no dead space and may impact ionize immediately. This mechanism results in an enhancement in the breakdown probability, which in turn, reduces the electric field required for avalanche breakdown, thereby reducing tunneling dark carriers. Our analysis neglects phonon scattering, which may result in loss of energy of carriers as they are energized, thereby reducing the significance of the initial-energy effect.

Indeed, the calculated channel capacity of an optimized $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -InP heterojunction is higher than that for a simple InP multiplication region of the same width, as shown

in Fig. 3. The improvement in the peak channel capacity offered by the optimal $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ -InP heterostructure, relative to the simple InP multiplication region, for the total multiplication-region widths of 100, 200, 300, 400, and 500 nm is 129%, 47%, 7%, 2%, and 1%, respectively. On the other hand, for the same range of widths, the improvement in the FWHM is 16%, 7%, 6%, 3%, and 2%, respectively. Note that the performance advantage is significant only in thin structures, which is due to the diminishing role played by the dead space in thick multiplication regions.

A similar trend exits when considering the SPQE [4]; however, the optimal heterostructure that maximizes the channel capacity is different from that maximizing the SPQE. For example, when the total width is 300 nm, the optimal SPQE-maximizing width of the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ layer is 32 nm, compared to 34 nm in the case of the channel-capacity optimization. Additionally, the applied voltage at which the SPQE peaks ($V = 20.5$ V) is different from that at which channel capacity peaks ($V = 19$ V). If the channel-capacity-optimized heterostructure SPAD is operated at $V = 20.5$ V, the channel capacity would be 58% below its maximum value (at $V = 19$ V).

IV. CONCLUSION

The channel capacity is the appropriate metric for the performance assessment and optimization of SPADs when used as detectors in a digital communication system. While both the channel capacity and the SPQE generally follow similar trends, device and operational-voltage optimizations based on these metrics are not equivalent; optimization with respect the wrong metric may result in a performance level that is well below the optimal performance.

REFERENCES

- [1] Y. Kang, H. X. Lu, Y. H. Lo, D. S. Bethune, and W. P. Risk, “Dark count probability and quantum efficiency of avalanche photodiodes for single-photon detection,” *Appl. Phys. Lett.*, vol. 83, pp. 2955–2957, 2003.
- [2] A. Biswas and W. H. Farr, “Detectors for ground-based reception of laser communications from Mars,” in *17th Annu. Meeting IEEE Lasers and Electro-Optics Society (LEOS)*, vol. 1, 2004, pp. 74–75.
- [3] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.
- [4] D. A. Ramirez, M. M. Hayat, G. Karve, J. C. Campbell, S. N. Torres, B. E. A. Saleh, and M. C. Teich, “Generalized breakdown probabilities and detection efficiencies for single-photon avalanche photodiodes,” *IEEE J. Quantum Electron.*, submitted for publication.
- [5] M. M. Hayat, Ü. Sakoğlu, O.-H. Kwon, S. Wang, J. C. Campbell, B. E. A. Saleh, and M. C. Teich, “Breakdown probabilities for thin heterostructure avalanche photodiodes,” *IEEE J. Quantum Electron.*, vol. 39, no. 1, pp. 179–185, Jan. 2003.
- [6] G. Karve, S. Wang, F. Ma, X. Li, and J. C. Campbell, “Origin of dark counts in $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ avalanche photodiodes operated in Geiger mode,” *Appl. Phys. Lett.*, vol. 86, pp. 63 505–1–63 505–3, 2005.
- [7] G. Karve, X. Zheng, X. Zhang, X. Li, S. Wang, F. Ma, A. Holmes, J. C. Campbell, G. S. Kinsey, J. C. Boisvert, T. D. Isshiki, R. Sudharsanan, D. S. Bethune, and W. P. Risk, “Geiger mode operation of an $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ avalanche photodiodes,” *IEEE J. Quantum Electron.*, vol. 39, no. 10, pp. 1281–1286, Oct. 2003.
- [8] O.-H. Kwon, M. M. Hayat, S. Wang, J. C. Campbell, B. E. A. Saleh, and M. C. Teich, “Optimal excess-noise reduction in thin heterojunction $\text{Al}_{0.6}\text{Ga}_{0.4}\text{As}/\text{GaAs}$ avalanche photodiodes,” *IEEE J. Quantum Electron.*, vol. 39, no. 10, pp. 1287–1296, Oct. 2003.
- [9] M. M. Hayat, O.-H. Kwon, S. Wang, J. C. Campbell, B. E. A. Saleh, and M. C. Teich, “Boundary effects on multiplication noise in thin heterostructure avalanche photodiodes: Theory and experiment,” *IEEE Trans. Electron Devices*, vol. 49, no. 12, pp. 2114–2123, Dec. 2002.