8-1-2009

Optimization of InP APDs for High-Speed Lightwave Systems

Daniel S.G. Ong

Jo Shien Ng

Majeed M. Hayat

John P.R. David

Journal of Lightwave Technology, Vol. 27, No. 15 (August 1, 2009) : 3294-3302. DOI.

Optimization of InP APDs for High-Speed Lightwave Systems

Daniel S. G. Ong
Department of Electronic & Electrical Engineering, University of Sheffield, Sheffield, UK
Jo Shien Ng
Department of Electronic & Electrical Engineering, University of Sheffield, Sheffield, UK
Majeed M. Hayat
Department of Electrical and Computer Engineering, the University of New Mexico, Albuquerque, NM
Peng Sun
Western Digital Corp., Lake Forest, CA
John P.R. David
Department of Electronic & Electrical Engineering, University of Sheffield, Sheffield, UK

Abstract:

Calculations based on a rigorous analytical model are carried out to optimize the width of the indium phosphide avalanche region in high-speed direct-detection avalanche photodiode-based optical receivers. The model includes the effects of intersymbol interference (ISI), tunneling current, avalanche noise, and its correlation with

the stochastic avalanche duration, as well as dead space. A minimum receiver sensitivity of -28 dBm is predicted at an optimal width of 0.18 mum and an optimal gain of approximately 13, for a 10 Gb/s communication system, assuming a Johnson noise level of 629 noise electrons per bit. The interplay among the factors controlling the optimum sensitivity is confirmed. Results show that for a given transmission speed, as the device width decreases below an optimum value, increased tunneling current outweighs avalanche noise reduction due to dead space, resulting in an increase in receiver sensitivity. As the device width increases above its optimum value, the receiver sensitivity increases as device bandwidth decreases, causing ISI to dominate avalanche noise and tunneling current shot noise.

Keywords

Indium phosphide, Intersymbol interference, Tunneling, Optical receivers, Noise level, Analytical models, Optical noise, Stochastic resonance, Electrons, Communication system control

SECTION I. Introduction

Indium phosphide (InP) avalanche photodiodes (APDs) have become the photodetectors of choice in present-day high-speed direct-detection lightwave communication systems [1]. Compared to p-i-n photodetectors, these APDs offer bias-controllable optoelectronic gain, G, generated by carrier impact ionizations in the InP avalanche region, which amplifies the photocurrent. This amplification suppresses Johnson noise and ultimately improves the receiver sensitivity [2]. However, determining the optimum avalanche-region width and the associated gain (or bias voltage) is quite complex. There are three main competing factors that govern the sensitivity of APD-based optical receivers at high speeds. They are (i) the avalanche noise of the APD, represented by the excess noise factor, which governs the penalty brought about by the stochastic nature of the impact-ionization process; (ii) the stochastic avalanche duration (or buildup time), which is strongly correlated with gain and governs the APD's speed and ultimately the level of intersymbol interference (ISI); and (iii) the APD's dark current, which is typically dominated by tunneling in the avalanche region.

Generally, as the gain increases, so do the excess noise factor and the avalanche buildup time. Thus, for a fixed avalanche-region width, there is an optimal sensitivity minimizing gain that offers a balance between suppressing Johnson noise while keeping the degrading contributions from the excess noise factor and ISI at a minimum. More importantly, changing the width of the avalanche region strongly affects the receiver sensitivity, as all of the aforementioned three factors change. On the one hand, reducing the thickness of the avalanche region serves to reduce the excess noise factor (due to the dead-space effect) [3]–[4][5] and minimize ISI via reducing carrier transit times across the avalanche region. On the other hand, the increase in the field in narrow regions accentuates tunneling current at exponential rates [6].

The most rigorous analytical approach for optimizing the width of the avalanche region must be based upon considering the performance metric of receiver sensitivity, which is the minimum average optical power in each bit required to produce a receiver bit error rate (BER) of 10^{-12}. As such, it is critical to have an accurate model for the BER, a model that takes into account the various device and system related factors that govern the performance of the receiver. There have been a few reports in the literature on the analytical formulation of the BER for APD-based receivers with the inclusion of ISI. The work reported in [7], for example, includes an analysis of BER for noninstantaneous APDs with the inclusion of the dead space using a Gaussian approximation for the probability density function (PDF) of the receiver output, albeit with the exact first-order and second-order statistics. Groves and David [8] have recently performed a Monte Carlo analysis of InP receivers including the effects of carrier velocity, the dead space, and the width of the APD's avalanche region. Also recently, Sun et al. [9]developed a rigorous model for the performance of high-speed direct-detection APD-based integrate-and-dump receivers. Sun's model allows us to calculate the BER and therefore determine the receiver sensitivity. The

model in [9] includes the effects of ISI, nonlocalized ionization to account for the dead-space effect, as well as the stochastic correlation between the gain and the avalanche duration. Similar to the model in [7], Sun et al. [9] also adopted a Gaussian approximation in their model for the PDF of the receiver output. More recently, Sun et al. [10] extended their earlier model [9] to include the exact analysis of the BER (using the exact statistics of the avalanche multiplication) as well as an asymptotic analysis based on large deviation theory; the latter is a generalization of the earlier asymptotic analyses by Letaief and Sadowsky [11]and Choi and Hayat [12]. However, no form of dark current, which can significantly contribute to APD noise and sensitivity, was modeled in [9]–[10][11][12].

In this paper, we extend the work in [9] by including InP tunneling current based on [13] and rigorously solve the optimization problem associated with the InP avalanche-region width. Despite the generality and sophistication of the model reported in [9] and its generalization in the present paper, it is fairly easy to use. The model offers for the first time compact expressions for the mean and variance of the receiver's output, with well-defined parameters that capture ISI, detector speed relative to the transmission speed, and the complex correlation between the APD's gain and buildup time. Moreover, to the best of our knowledge no rigorous model has been set forth heretofore for optimizing the sensitivity of APD-based receivers over the width of the avalanche region while taking into consideration all of the performance determining factors. This work may therefore enable device engineers to identify the optimal InP-based APD and the associated operation bias voltage for use at a prescribed digital transmission speed.

SECTION II. Model

A. Review of the Model Reported in [9]

For its relevance to the present paper, we begin by reviewing germane aspects of the probabilistic model for receiver-sensitivity analysis developed in [9], which can be summarized in the following four main points. First, a recursive method was developed characterizing the joint probability distribution function associated with the random variables comprising the APD's stochastic gain, G, and its stochastic avalanche duration time, T, resulting from a single avalanche trigger. Second, a stochastic parametric model, in terms of G and T, was developed to approximate the APD's stochastic impulse-response function (as a result of a single avalanche trigger); accurate approximations were derived for the first- and second-order statistics of the APD's impulse-response function. Third, the model was further exploited in a rigorous probabilistic framework to yield the moment-generating function (which is akin to the z-transform of the probability distribution function) of the random variable comprising the total charge accumulated within the integration time of an integrate-and-dump receiver. The latter formalism enabled the calculation of the statistics of the output of the integrate-and-dump receiver while capturing ISI (arising from an infinite random stream of past bits), as well as the dead-space effect. Fourth, such receiver characterization enabled the calculation of the receiver BER and sensitivity.

To better understand the stochastic nature of the APD's buildup-time-limited bandwidth, which is a result of the stochastic nature of the avalanche duration and its statistical correlation with the gain, Sun et al. [9] also introduced the so-called *shot-noise-equivalent bandwidth*, defined as $B_{\mathrm{sneq}} = \langle G^2/T\rangle/2\langle G\rangle^2 F$, where F is the APD's excess noise factor, defined as $F = \langle G^2\rangle/\langle G\rangle^2$, where brackets represent ensemble averaging. The quantity B_{sneq} can be computed using the joint probability distribution of G and T developed in [9]; it is also the precise bandwidth that when used in the usual formula for APD-amplified shot noise, i.e., $\sigma^2 = 2e\langle\frac{G)^2 F B_{\mathrm{sneq}}\eta P}{h\nu}$, the correct value of the shot-noise variance is obtained [9], where η is the APD quantum efficiency, P is the optical power, h is Planck's constant, and ν is the photon's frequency. It was shown in [9] that due to the stochastic coupling between T and G, B_{sneq} is generally greater than the conventional 3 dB bandwidth of the APD, $B_{3\mathrm{dB}}$, which is often taken as the 3 dB drop point in the Fourier transform of the APD's mean impulse-

response function [9]. This discrepancy can be as high as 30% [9], leading to a similar error in the prediction of the APD-amplified shot-noise variance if $B_{3\text{dB}}$ is incorrectly used in place of B_{sneq}.

We now describe the Gaussian approximation method used in [9] to calculate the BER. The output of the integrate-and-dump receiver (in units of electrons) was approximated by a Gaussian random variable, albeit with the exact mean and variance, and the BER was computed using the usual formula [14]

$$\text{BER} \approx \frac{1}{4}\left[\text{erfc}\left(\frac{\theta-\mu_0}{\sqrt{2}\sigma_0}\right) + \text{erfc}\left(\frac{\mu_1-\theta}{\sqrt{2}\sigma_1}\right)\right] \quad (1)$$

where μ_0 and σ_0^2 denote the mean and variance for the receiver's output conditional on the present bit (i.e., the information bit corresponding to the receiver's present integration period) being "0," and μ_1 and σ_1^2 are similar quantities conditional on the present bit being "1." The decision threshold θ is taken as

$$\theta = \frac{\mu_0\sigma_1+\mu_1\sigma_0}{\sigma_0+\sigma_1} \quad (2)$$

which is a convenient accurate approximation to the optimal decision threshold that minimizes the BER [14]. The expressions for the parameters μ_0, σ_0^2, μ_1, and σ_1^2 are derived as [9]

$$\begin{aligned}
\mu_0 &= \frac{1}{2}\frac{n_0\langle G\rangle}{\kappa\lambda}\left(1-e^{-\kappa\lambda}\right) \\
\sigma_0^2 &= \frac{1}{4}\frac{n_0^2\langle G\rangle^2}{\kappa\lambda^2}\frac{(1-e^{-\kappa\lambda})^4}{(1-e^{-2\kappa\lambda})} \\
&\quad + \frac{n_0\langle G\rangle^2 F}{2\kappa\lambda}\left(1-e^{-\kappa\lambda}-\kappa\lambda e^{-\kappa\lambda}\right) + \sigma_J^2 \qquad (3)(4)(5)(6) \\
\mu_1 &= \mu_0 + \frac{n_0\langle G\rangle}{\kappa\lambda}\left(\kappa\lambda-1+e^{-\kappa\lambda}\right) \\
\sigma_1^2 &= \sigma_0^2 + \frac{n_0\langle G\rangle^2 F}{\kappa\lambda}\left(\kappa\lambda-2+2e^{-\kappa\lambda}+\kappa\lambda e^{-\kappa\lambda}\right)
\end{aligned}$$

where n_0 is the average number of absorbed photons per "1" bit, $\kappa = \frac{4B_{\text{sneq}}}{2\pi B_{3\text{dB}}}$is the so-called *bandwidth correction factor* (it accounts for the discrepancy between B_{sneq} and $B_{3\text{dB}}$) and λ, termed here as the *detector speed factor*, is a measure of the detector's relative speed, defined as $\lambda = \frac{2\pi B_{3\text{dB}}}{R_\text{b}}$, where $R_\text{b} = \frac{1}{T_\text{b}}$is the bit transmission speed, and T_b is the bit duration. Finally, the term σ2J represents the variance of Johnson noise accumulated in the integration time. We also emphasize that μ_0, σ_0^2, μ_1, and σ_1^2 are quantities that are averaged over all possible past bit patterns.

We note that the expressions in (3)–(6) are generalizations of the traditional expressions for the output statistics of APD-based receivers found in optical communication literature [14]. Whilst these expressions capture the usual effects of shot noise and the excess noise due to avalanche multiplication, they additionally capture the effects of ISI, relative speed of the detector, as well as the stochastic coupling between the APD's gain and buildup time through the effective use of the novel parameters κ and λ. It is interesting to observe that for an instantaneous detector for which $B_{3\text{dB}} = \infty$, the detector speed factor λ is infinite, and the expressions shown in (3)–(6) collapse to the traditional expressions for the receiver mean and variance in the absence of ISI [14]: $\mu_0 = 0, \sigma_0^2 = \sigma_J^2, \mu_1 = n_0\langle G\rangle \; and \; \sigma_0^2 = \sigma_J^2 + n_0\langle G\rangle^2 F$. Moreover, in detectors for which the gain is unity (e.g., a simple pin diode), the bandwidth correction factor κ is unity, resulting in simplified versions of (3)–(6) that continue to capture the effect of ISI. However, the parameters μ_0, σ_0^2, μ_1, and σ_1^2 shown above do not

include the contribution from the APD's tunneling current generated in the multiplication region of the APD, which acts to trigger avalanches just as the detected photons do.

We end this section by making important observations when tunneling current is included in the analysis. First, we point out that the term in μ_0 and the first two terms in σ_0^2, as shown in (3) and (4), respectively, are due entirely to contributions from ISI resulting from the random stream of bits (preceding the present bit). In particular, they arise from contributions from photo-generated carriers generated within the bits that precede the present bit. Second, the second term in (5) and second term in (6) are due to contributions from carriers generated during the present bit.

B. Variant of the Gaussian-Approximation Approach for BER Calculation

In this paper, we replace the assumption in [9], which dictates that the receiver output, conditional on the state of the present bit, is a Gaussian random variable, with the more realistic assumption that the receiver output, conditional on the state of the present bit *and the entire past bit stream*, is a Gaussian random variable. One can then compute the BER conditional on the entire past bit stream, and then average the resulting pattern-specific BERs over all possible past bit patterns and obtain the overall average BER. The advantage of this approach compared to the one in [9] is that it relaxes the often unrealistic assumption of a unimodal PDF for the receiver output conditional on the state of the present bit. While this variant approximation will yield an improved approximation of the average BER, it does come, however, with a slight increase in computational cost. The details are described next.

Consider the scenario for which the *n*th past bit ($n = 1,2, \ldots$) is a "1" bit and all other past bits are "0" bits. Assume further for the moment that the present bit (corresponding to $n = 0$) is a "0" bit. In [9], the authors provided a formula for the mean and variance of the output of an integrate-and-dump receiver excited by a random bit pattern ((13) and (46) in [11]). If we specialize these expressions to the above deterministic bit pattern whose only nonzero bit is the *n*th bit, then after some algebra, we can obtain expressions for the mean and the variance of the receiver output when excited by this particular bit pattern. As shown in the Appendix, for $n = 1,2, \ldots$, the mean and variance of the ISI contributions in the receiver output from the *n*th past bit alone are respectively given by

$$\mu_{\mathrm{ISI},n} = \frac{2n_0\langle G\rangle e^{-\kappa\lambda n}}{\kappa\lambda}\left(\cosh(\kappa\lambda) - 1\right) \quad (7)$$

and

$$\sigma^2_{\mathrm{ISI},n} = \frac{n_0\langle G\rangle^2 F}{\kappa\lambda} e^{-\kappa\lambda n}\left(e^{-\kappa\lambda} - 1\right)\left(1 - \kappa\lambda e^{-\kappa\lambda} - e^{-\kappa\lambda}\right). \quad (8)$$

Now if we consider an arbitrary past bit pattern, I_j, of length L bits, then one can calculate the mean of the receiver output when the current bit is zero by adding up the contributions from each of the ISI terms from the past bits in the pattern I_j; this yields the expression

$$\mu_0(I_j) = \sum_{k=1}^{L} a_k(I_j)\mu_{\mathrm{ISI},k} \quad (9)$$

where $a_k(I_j) = 0$ unless the kth bit in the pattern I_j is a "1" bit, in which case $a_k(I_j)$ assumes the value 1. The receiver mean output, $\mu_1(I_j)$, when the present bit is a "1" bit is obtained by adding to $\mu_0(I_j)$ the contributions from the photons in the current bit. From the discussion near the end of Section II-A, we know that the contribution to the mean of the receiver output from the photons available in the present bit

is $\left(\frac{n_0\langle G\rangle}{\kappa\lambda}\right)\left(\kappa\lambda - 1 + e^{-\kappa\lambda}\right)$ (which is merely the second term in (5)). When we combine this component with the contributions from the ISI terms, we obtain

$$\mu_1(I_j) = \mu_0(I_j) + \frac{n_0\langle G\rangle}{\kappa\lambda}\left(\kappa\lambda - 1 + e^{-\kappa\lambda}\right). \quad (10)$$

Similarly, one can calculate the variance of the receiver output associated with the pattern I_j when the current bit is a "0" bit by adding up the ISI contributions from past bits in the specific pattern as well as contribution from Johnson noise and obtain

$$\sigma_0^2(I_j) = \sum_{l=1}^{L} a_l(I_j)\sigma_{\mathrm{ISI},l}^2 + \sigma_J^2 . \quad (11)$$

The corresponding expression when the current bit is "1" is obtained by simply adding to $\sigma_0^2(I_j)$the contributions from the photons in the current bit; this yields

$$\sigma_1^2(I_j) = \sigma_0^2(I_j) + \frac{n_0\langle G\rangle^2 F}{\kappa\lambda}\left(\kappa\lambda - 2 + 2e^{-\kappa\lambda} + \kappa\lambda e^{-\kappa\lambda}\right). \quad (12)$$

Note that the second term in (12) corresponds to the contributions to the variance of the receiver output from photons available to the present bit, which is identical to the second term in (6).

Next, for every pattern I_j, $j = 1, \ldots 2^L$, we can calculate the *pattern-specific* BER as follows

$$\mathrm{BER}(I_j) \approx \frac{1}{4}\left[\mathrm{erfc}\left(\frac{\theta - \mu_0(I_j)}{\sqrt{2}\sigma_0(I_j)}\right) + \mathrm{erfc}\left(\frac{\mu_1(I_j) - \theta}{\sqrt{2}\sigma_1(I_j)}\right)\right] \quad (13)$$

where θ is calculated as before from (2), in conjunction with (3)–(6). To calculate the overall BER, we simply compute the ensemble average of the pattern-specific BERs over all possible past bit patterns. More precisely, we have

$$\mathrm{BER} = \frac{1}{2^L}\sum_{j=1}^{2^L} \mathrm{BER}\,(I_j).$$

C. Generalization of the Model to Include Multiplied Tunneling Current

The un-multiplied band-to-band tunneling current, I_{tunn}, is modeled by [15]

$$I_{\mathrm{tunn}} = \frac{(2m^*)^{0.5} q^3 \Im V A}{h^2 E_g^{0.5}} \exp\left(-\frac{2\pi\sigma_T (m^*)^{0.5} E_g^{1.5}}{q h \Im}\right) \quad (15)$$

where m^* is the effective electron mass, q is the electron charge, $\Im$ is the electric field, V is the applied reverse bias voltage, A is the device area, E_g is the direct energy band gap, and σ_T is the tunneling fitting parameter; $\sigma_T = 1.16$ was used for InP in [13]. The average number of dark carriers generated per bit time interval T_b is given by $n_d = {I_{\mathrm{tunn}} T_b}/{q}$. Since the dark-carrier generation has Poisson statistics, it is plausible to attempt to include the effect of dark carriers on the parameters μ_0, σ_0^2, μ_1 and σ_1^2 by treating dark carriers as photo-generated carriers. However, there is a caveat; this procedure must be performed with care since dark-carrier generation exists independently of the status of the optical signal, whereas the photo-carrier generation is modulated by a random stream of binary random variables. One of the contributions of this paper is to

generalize the parameters given by (9)–(12), and hence the expression for the BER in (14), to include tunneling current.

We begin by deriving the new version of $\mu_0(I_j)$ that includes tunneling current. To this end, we initially consider the expression in (3) but assume that the photon arrival rate is held constant (i.e., no binary modulation is present). In this scenario, which resembles the case when only dark carriers are allowed to trigger the APD, the expression for the mean $\mu_0(I_j)$should include the term $\left(\frac{n_d\langle G\rangle}{\kappa\lambda}\right)(1-e^{-\kappa\lambda})$, where n_o in (3) is replaced with $2n_d$ to compensate for the absence of random modulation in dark-carrier generation. Note, however, that the resultant will yet exclude the contributions from dark carriers generated during the present bit. To include the latter contributions, we must include the second term of (5) with n_o replaced with n_d. Putting all this together, we obtain the new expression for $\mu_0(I_j)$

$$\begin{aligned}\mu_0(I_j) &= \sum_{l=1}^{L} a_l(I_j)\mu_{\mathrm{ISI},l} + \frac{n_d\langle G\rangle}{\kappa\lambda}(1-e^{-\kappa\lambda}) \\ &\quad + \frac{n_d\langle G\rangle}{\kappa\lambda}(\kappa\lambda - 1 + e^{-\kappa\lambda}) \qquad \text{.(16)} \\ &= \sum_{l=1}^{L} a_l(I_j)\mu_{\mathrm{ISI},l} + n_d\langle G\rangle.\end{aligned}$$

The expression for $\mu_1(I_j)$is identical in form to that shown in (10) with the proviso that $\mu_0(I_j)$is now represented by (16) and not by (9).

The derivation of a new expression for $\sigma_0^2(I_j)$ is more complicated as we must appeal to (24) in [9], which represents the variance of the receiver output assuming a constant optical power that extends from the infinite past to the end of the present bit [9]

$$\sigma^2 = \phi\frac{2c}{b_{se}^2}\left[T_b - \frac{1}{b_{se}}(1-e^{-b_{se}T_b})\right] \text{(17)}$$

View Source where φ is the constant photon flux, $b_{se} = 4B_{\mathrm{sneq}}$ and $^{c}/_{b_{se}} = 2B_{\mathrm{sneq}}\langle G^2\rangle$. Note that by using the definition of B_{sneq}, we obtain the simplification $^{c}/_{b_{se}^2} = {}^{\langle G^2\rangle}/_{2}$. All that is required now is to utilize the expression in (17) while replacing the photon flux with the dark-carrier generation rate. After substituting the expressions for κ and λ in (17) while using $\phi = {}^{n_d}/_{T_b}$ (representing the dark-carrier generation rate), we obtain the new expression for $\sigma_0^2(I_j)$

$$\sigma_0^2(I_j) = \sum_{l=1}^{L} a_l(I_j)\sigma_{\mathrm{ISI},l}^2 + \frac{n_d\langle G\rangle^2 F}{\kappa\lambda}(\kappa\lambda - 1 + e^{-\kappa\lambda}) + \sigma_J^2 \text{..(18)}$$

The expression for $\sigma_1^2(I_j)$ is identical in form to that in (12) with the proviso that $\sigma_0^2(I_j)$is now given by (18) and not by (11). Note that when the detector is approximated as instantaneous, we can substitute $\lambda = \infty$ in (18) and obtain $\sigma_0^2 = (n_0\langle G\rangle^2 F/2) + n_d\langle G\rangle^2 F + \sigma_J^2$, which is once again the traditional expression for the variance of the output of the integrate-and-dump receiver in the absence of ISI [14].

Finally, we calculate the pattern-specific BER using (13), where θ is calculated from (2)but with μ_0 and σ_0^2 given by (16) and (18), respectively, and μ_1 and σ_1^2 given by (10) and (12), respectively. The bit-length parameter, L, can be chosen to be sufficiently large to capture all significant ISI terms; in our calculation (detailed

in Section III), we found $L = 10$ to be an adequate choice beyond which no tangible change in the BER was observed.

All the results to follow are generated using the model described in this subsection.

SECTION III. Results

In this paper, the joint probability distribution function of the gain and avalanche duration that the model in [9] requires is obtained from the random path length (RPL) model [16]. For the scope of this paper, this approach offers a computationally simple Monte Carlo-based alternative to the recursive analytical method reported in [9]. With the joint probability distribution at hand, we can calculate all the parameters of the model, which are κ, λ, $\langle G \rangle$, and F. The field-dependent nonlocalized ionization coefficients and the ionization threshold energies for InP are obtained from [13]. The accuracy of the RPL method was first cross-checked against the recursive technique using test runs. In our calculations, we considered transmission speeds in the range of 2.5–20 Gb/s.

In order to obtain the Johnson noise level, we investigated transimpedance amplifiers (TIAs) for 2.5–40 Gb/s operation. The input noise current density, i_n, and bandwidth, BTIA, of each TIA was obtained from [17]–[18][19][20][21]. These parameters were also obtained from data sheets for commercial TIA modules manufactured by Applied Micro Circuits Corporation (S3160, S3170), Maxim Integrated Products (MAX3271), Sumitomo Electric (F0100504B, F0100505B, F0100604B, F0100612B, F0100613B), Analog Devices (ADN2882, ADN2821), TriQuint Semiconductor (TGA4805-EPU, TGA4815-EPU, TGA4816, TGA4817-EPU, TGA4812), and Texas Instruments (ONET9901TA, ONET8501T, ONE8511T). The average of i_n for each transmission speed was fitted linearly to obtain the average i_n as a function of transmission speed. The fit yielded the equation $i_n = 4.81 \times 10^{-10} R_b + {5.87\text{pA}}/{\sqrt{\text{Hz}}}$.Similarly, by fitting the average of BTIA against transmission speed, we obtained the average B_{TIA} as a function of transmission speed, given by B_{TIA}=0.91Rb GHz. Using these averaged i_n and B_{TIA} values, we were then able to obtain the average Johnson noise levels, σ_J, as a function of transmission speed using the formula $\sigma_J = \left({\sqrt{B_{\text{TIA}} i_n^2}}/{q}\right)\left({1}/{R_b}\right)$.

A. Optimum Avalanche Width for a Given Transmission Speed

The investigation was conducted on a series of 25-μm-radius InP p-i-n diodes, with avalanche-region widths, w, ranging from 0.13 to 0.5 μm. Ideal electric-field profile (with negligible depletion into the p and n claddings and a constant electric field across the i-region) was assumed. The RPL calculations used InP ionization coefficients and threshold energies reported in [13], which are valid for electric fields in the range 180–850 kV/cm. The 0.13 μm diodes required extrapolations of ionization coefficients when the electric field exceeded 850 kV/cm. As holes impact ionize more readily than electrons in InP, the calculations used avalanche statistics due to pure hole injection into the InP avalanche region, as realized in practice by separate absorption multiplication InGaAs/InP APDs.

Sensitivity versus gain curves were calculated for the diodes and the results are compared in Fig. 1 at a transmission speed of 10 Gb/s. The key observation is that for each diode, there exists an optimum mean gain that achieves the lowest sensitivity. While this result is expected, it is the first time that it is reported with the inclusion of both ISI and tunneling current, which further enables us to make a correct prediction of the optimal operation gain. In Fig. 2, we plot the lowest sensitivity for each device and corresponding optimal mean gain both as functions of the avalanche-region width; this plot allows us to identify the optimum avalanche width for a given transmission speed, thereby yielding the optimized sensitivity for a given transmission speed. This is a

key contribution of this paper. Indeed, our calculations predict an optimum avalanche width of 0.19 μm for InP APDs, yielding a lowest sensitivity of −28 dBm at an optimal gain of approximately 13 for a 10 Gb/s system.

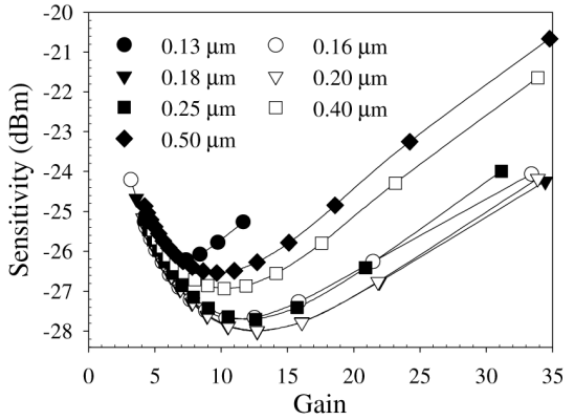

Fig. 1. Receiver sensitivity versus gain for the InP p-i-n APDs investigated for a 10 Gb/s transmission system.

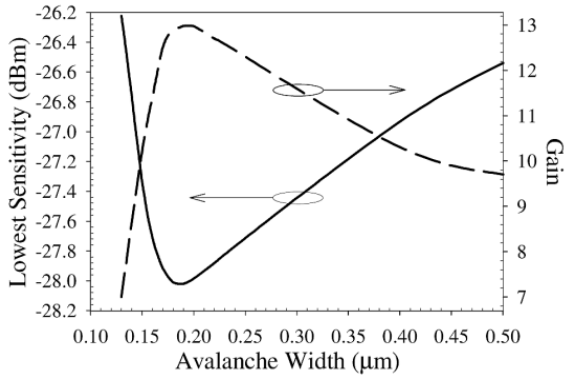

Fig. 2. Lowest sensitivity (solid line, left axis) and its corresponding optimal mean gain (dashed line, right axis) versus InP APD avalanche width for a 10 Gb/s transmission system.

For clarity, we show how we obtain the lowest sensitivity and corresponding optimal mean gain for each avalanche width in Fig. 3, which shows G and the sensitivity as functions of electric field for a 0.19 μm InP APD. To give the reader an idea of the scale of B_{sneq} and $B_{3\mathrm{dB}}$ as functions of electric field, these quantities are also shown in Fig. 3, along with the values for κ and λ, which are derived from B_{sneq} and $B_{3\mathrm{dB}}$.

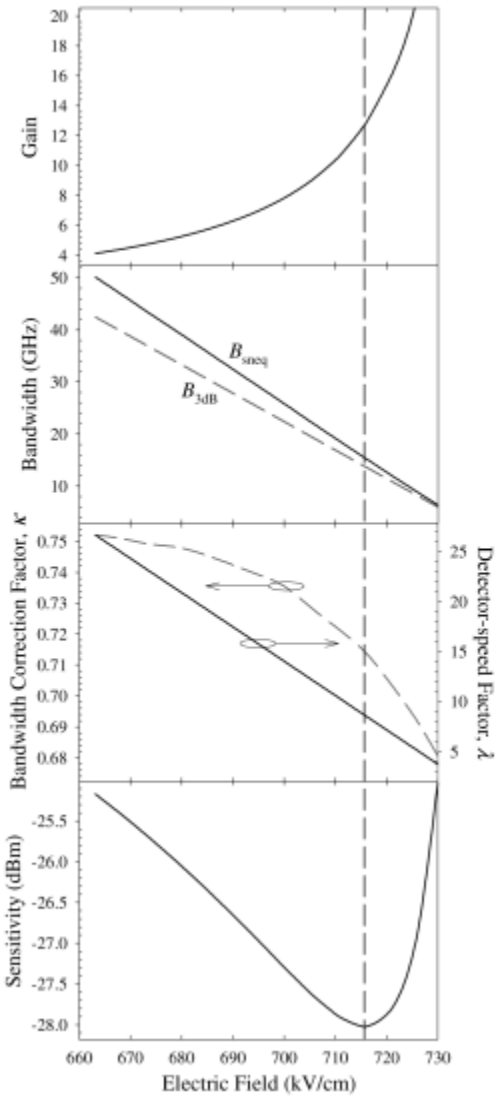

Fig. 3. Mean gain, shot-noise-equivalent bandwidth, 3 dB bandwidth, bandwidth correction factor, detector speed factor, and sensitivity, all functions of the electric field in the avalanche region for a 10 Gb/s transmission system and a 0.19 µm InP APD. The dashed line that runs through the graphs corresponds to the lowest sensitivity.

B. Competing Effects of Avalanche Excess Noise, Bandwidth, and Tunneling Current

The calculations conducted so far included effects of ISI, tunneling current, and presence of dead space, which we refer to hereafter as the *complete* calculations. In order to independently assess the significance of (i) ISI, (ii) device bandwidth, (iii) tunneling current, and (iv) the dead space in the ionization process, we carried out four additional sets of calculations, which we refer to as the *incomplete* calculations (all at 10 Gb/s). Each set in the incomplete calculations ignores exactly one of the above four effects. ISI was excluded from the calculations by setting $L = 0$ in (16) and (18). The device bandwidth constraint was removed by setting $\lambda = \infty$, which corresponds to an instantaneous APD ($B_{3\text{dB}} = \infty$). It is worth noting that the effect of ISI is automatically ignored in an instantaneous APD. It is also important to note that even when ISI is excluded from the model by means of setting $L = 0$, the receiver output is still affected by the bandwidth through the parameter λ in the second terms of (10) and (12), which, in turn, represent the attenuation in the receiver output resulting from the APD's bandwidth constraint. This shows the capability of the model to exclude ISI effects alone without the need for assuming an infinite APD bandwidth. Tunneling current can be excluded by setting $n_d = 0$. Lastly, presence of dead space can be ignored by assuming zero dead space for both electrons and holes and by employing ionization coefficients available for bulk InP, in which the dead-space effects are inherently neglected [22].

Results from each of these three sets of incomplete calculations are compared to those from the complete calculation in Fig. 4. The calculations excluding ISI predicted lower sensitivity values than those of the complete calculation for $w > 0.13\mu$m. When ISI is ignored, a slightly thicker optimum avalanche width of about 0.22 μm was predicted compared to the complete calculations, yielding a lower optimum sensitivity of about −29 dBm. However, the sensitivity values for diodes with w<0.16 μm are indistinguishable for the two sets of calculations, confirming that the significance of ISI decreases for narrower diodes with larger bandwidths.

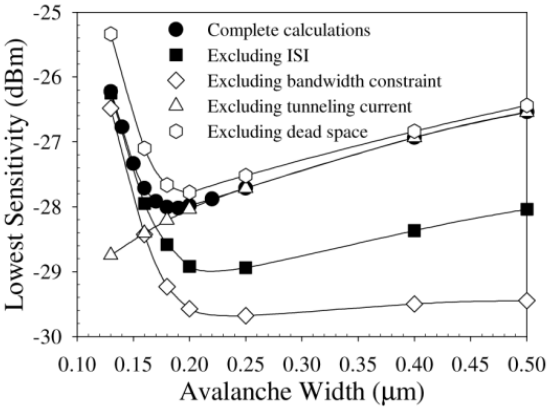

Fig. 4. Sensitivity versus avalanche width for the complete and various incomplete calculation conditions for a 10 Gb/s system. Different curves identify the distinct roles of ISI, device bandwidth, avalanche excess noise, and tunneling current.

When the device bandwidth constraint was removed from the calculations, even lower sensitivity values were predicted than those from the complete calculation and those excluding ISI. By removing the bandwidth constraint, a thicker optimum avalanche width, of about 0.25 μm, was predicted compared to the complete calculations, yielding a lower optimum sensitivity of about −29.7 dBm. Above this optimum width, the rate of increase in sensitivity is slower when the bandwidth constraint is removed compared to the case when only the ISI is removed.

Next, we consider the third set of incomplete calculations by excluding the effect of tunneling current. The calculations resulted in sensitivity versus avalanche-region width characteristics indistinguishable from the complete calculation except when $w < 0.20$ μm, for which reduced sensitivity values are observed. This behavior suggests that at high fields (i.e., in narrow diodes) the degrading effect of tunneling current on the sensitivity strongly outweighs the benefits of reduced ISI and excess noise.

In the fourth set of incomplete calculations for which the dead space is neglected, the results predicted an optimum width of 0.20 μm, yielding an optimum sensitivity of about −27.8 dBm. In particular, the sensitivity values are higher than those obtained from the complete calculation because the local model overestimates the excess noise.

C. Effect of Transmission Speed on Sensitivity Optimization

The APD bandwidth becomes more crucial as the transmission speed increases. For a given transmission speed, the significance of ISI is expected to increase with w since the bandwidth decreases. As a result, this affects the receiver sensitivity and hence, it is important to investigate the dependence of optimum sensitivity on transmission speed. To do so, we repeated the complete sensitivity calculations for a range of transmission speeds from 2.5 to 20 Gb/s. A similar trend in the optimized sensitivity versus width characteristic, as that in Fig. 4, is also observed at different speeds. For a clearer analysis, the gain-and-width optimized sensitivity is plotted against transmission speed, as shown in Fig. 5, as well as the optimal avalanche width (yielding the lowest sensitivity). As expected, the optimized sensitivity increases with the transmission speed, as the

significance of ISI increases; narrower diodes are required as the transmission speed increases. Our calculations also show that the optimum gain decreases with the transmission speed (not shown here) in order to avoid the more significant tunneling current in narrower diodes.

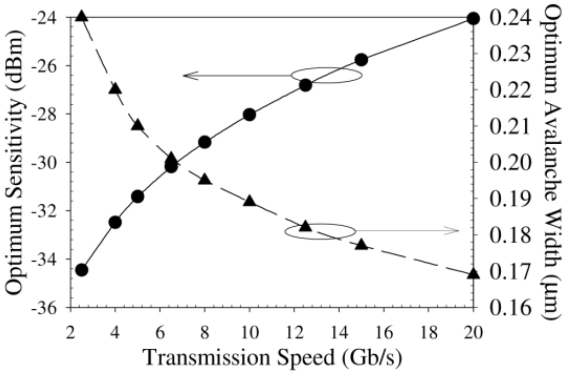

Fig. 5. Optimum sensitivity (•, left axis) and avalanche width (▲, right axis) versus transmission speed for InP diodes.

SECTION IV. Discussion

For a given transmission speed, the optimized sensitivity versus width characteristics are controlled in a very complex fashion by three device-related factors, namely the tunneling current, excess noise characteristics, and the device bandwidth. As the device width decreases, the operating field increases, resulting in increased tunneling current, as shown in Fig. 6. It can also be observed from Fig. 6 that k_{eff} (a parameter commonly used to describe excess noise versus multiplication factor characteristics) decreases with thinner devices confirming a lower excess noise factor, as the dead-space effect becomes more significant [6], [13]. At the same time, the APD's bandwidth decreases with *w*, which causes weaker receiver output as well as an increase in the significance of ISI, thereby causing an elevation in the sensitivity.

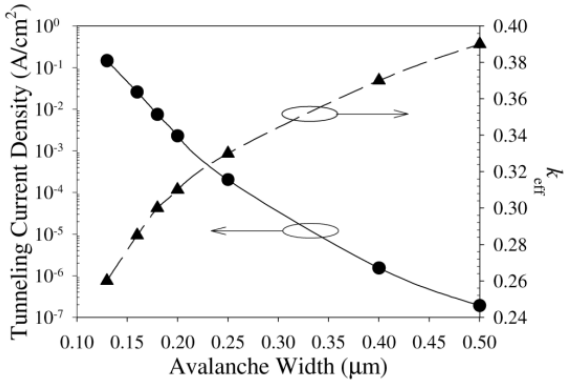

Fig. 6. Tunneling current density (•, left axis) and keff (▲, right axis) versus device avalanche width for a 10 Gb/s system.

In the results of complete calculations, high sensitivity values for diodes narrower than the optimum avalanche width are due to high tunneling current. For diodes wider than the optimum avalanche width, sensitivity increases with *w*, as described above. However, the relative dominance of increasing k_{eff} (resulting in an increase in the excess noise) on the one hand and decreasing diode bandwidth on the other hand becomes clear only through the rigorous modeling and calculations performed in this paper. Sensitivity results from the calculations that exclude the bandwidth constraint are only affected by changes in the excess noise when *w* is increased beyond the optimum width. Consequently, we observe that the sensitivity increases more slowly with avalanche width compared to that obtained from the complete calculation, suggesting that a decreasing device bandwidth plays a more dominant role than increasing excess noise on sensitivity as *w* increases. As such,

calculations that ignore bandwidth effects erroneously predict higher optimal device gains compared to those predicted by the complete calculation.

SECTION V. Conclusion

We have generalized the APD-based receiver model in [9] to include tunneling current and used it for the purpose of optimization of the avalanche-region width for best receiver sensitivity for an arbitrarily prescribed transmission speed. The model offers compact analytical expressions for the mean and the variance of the output of the integrate-and-dump APD-based receiver that capture, in the presence of dark current, the complex effects of ISI and the stochastic correlation between the APD's gain and bandwidth. These expressions, which can also be used to calculate the signal-to-noise ratio of the receiver output, are generalizations of the traditional counterparts that neglect the effects of ISI and the stochastic coupling between the APD's gain and bandwidth.

Application of the theory to InP receivers showed that for a 10 Gb/s system, an optimal width of 0.19 μm is predicted, yielding a minimum sensitivity of −28 dBm at an optimal gain of approximately 13. The factors that control the optimized sensitivity versus width characteristics were also confirmed theoretically. As device width decreases below its optimum value, increased tunneling current results in increasing receiver sensitivity. Conversely, as device width increases above its optimum, a decreasing device bandwidth causes the receiver sensitivity to increase, and is the dominating factor compared to increasing excess noise.

ACKNOWLEDGMENT

The authors thank Mr. Caoxie Zhang from Shanghai Jiao Tong University for providing many helpful comments.

Appendix Derivation of $\mu_{\mathrm{ISI},n}$

Consider the first term in (13) in [9], which is given by

$$\int_{-\infty}^{0} \langle I_p(t-\tau)\rangle\, \phi(\tau) d\tau. \textbf{(A1)}$$

This quantity represents the photocurrent input to the integrate-and-dump receiver arising from photons that are present in the interval $(-\infty, 0)$. Note that the photon flux, ϕ, is arbitrary in this equation. For $n = 1,2,3$ …,let us select $\phi(u) = \phi_n(u)$as follows:

$$\phi_n(u) = \begin{cases} \phi_0, & -nT < u \le -nT + T \\ 0, & \text{otherwise} \end{cases} \textbf{(A2)}$$

where ϕ_0 is a constant. With $\phi(u)$ in (A1) replaced by $\phi_n(u)$, the integral of the first term in (A1) over the interval $[0, T_b]$ yields $\mu_{\mathrm{ISI},n}$, which is the output of the integrate-and-dump receiver arising from photons that are present in the interval $(-nT, (n+1)T]$. Note that $\mu_{\mathrm{ISI},n}$is precisely the ISI contribution to the mean of the receiver output arising from the nth past bit, which we have denoted as $\mu_{\mathrm{ISI},n}$. To reiterate

$$\mu_{\mathrm{ISI},n} = \int_0^{T_b} \int_{-nT}^{-nT+T} \langle I_p(t-\tau)\rangle\, \phi_n(\tau) d\tau dt. \textbf{(A3)}$$

We now substitute from (21) in [9] the approximation $\langle I_p(t)\rangle \approx a e^{-b_{se} t} u(t)$, and after some algebra, we obtain

$$\mu_{\mathrm{ISI},n} = \phi \frac{a}{b^2} e^{-bnT}(-2 + e^{-bT} + e^{bT}). \text{ (A4)}$$

Recall that $\kappa = {4B_{\mathrm{sneq}}}/{2\pi B_{3\mathrm{dB}}}$ and $\lambda = 2\pi B_{3\mathrm{dB}} T$, so $4B_{\mathrm{sneq}} T = \kappa\lambda$. Also, from [9] we have ${a}/{b_{se}} = \langle G \rangle$. If we substitute these in **(A4)** and use $n_0 = \phi_0 T$, we obtain

$$\mu_{\mathrm{ISI},n} = \frac{n_0 \langle G \rangle e^{-\kappa\lambda n}}{\kappa\lambda}\left(-2 + e^{-\kappa\lambda} + e^{\kappa\lambda}\right) \text{ (A5)}$$

which simplifies to

$$\mu_{\mathrm{ISI},n} = \frac{2n_0 \langle G \rangle e^{-\kappa\lambda n}}{\kappa\lambda}(\cosh(\kappa\lambda) - 1). \text{(A6)}$$

Remark

If we sum up **(A5)** over all $n = 1,2,3 \ldots$ we obtain $\left(\frac{n_0\langle G \rangle}{\kappa\lambda}\right)\left(1 - e^{-\kappa\lambda}\right)$, which correctly agrees with twice the quantity in **(3)** since there is binary modulation in the derivation of **(A6)**.

Appendix Derivation of $\sigma^2_{\mathrm{ISI},n}$

The equivalent of (46) in [9] for the case of a deterministic photon flux $\phi_n(.)$is

$$\sigma^2_{\mathrm{ISI},n} = \int_0^T \int_0^T \int_{-nT}^{-nT+T} \langle I_p(\mu - \xi) I_p(\nu - \xi) \rangle \, \phi_n(\xi) d\xi d\mu d\nu. \text{(A7)}$$

We now substitute in the approximation $\langle I_p(\mu) I_p(\nu) \rangle \approx c e^{-b_{se} \max(\mu,\nu)}$from (21) in [9], carry out the triple integral and obtain

$$\sigma^2_{\mathrm{ISI},n} = \phi \frac{c}{b_{se}^2} e^{-b_{se} T n}(e^{b_{se} T} - 1) \times \left\{-2Te^{-b_{se}T} - 2\frac{e^{-b_{se}T}}{b_{se}} + \frac{2}{b_{se}}\right\}. \text{(A8)}$$

Upon using $4B_{\mathrm{sneq}} T = \kappa\lambda$ and $c/b_{se} = \langle G^2 \rangle / T = 2B_{\mathrm{sneq}} \langle G \rangle^2 F$ from [9], we obtain

$$\sigma^2_{\mathrm{ISI},n} = \frac{n_0 \langle G \rangle^2 \mathrm{F}}{\kappa\lambda} e^{-\kappa\lambda n}\left(e^{-\kappa\lambda} - 1\right)\left(1 - \kappa\lambda e^{-\kappa\lambda} - e^{-\kappa\lambda}\right). \text{(A9)}$$

References

1. J. C. Campbell, S. Demiguel, F. Ma, A. Beck, X. Y. Guo, S. L. Wang, X. G. Zheng, X. W. Li, J. D. Beck, M. A. Kinch, A. Huntington, L. A. Coldren, J. Decobert, N. Tscherptner, "Recent advances in avalanche photodiodes", *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, no. 4, pp. 777-787, Jul./Aug. 2004.

2. S. D. Personick, "Receiver design for digital fiber-optic communication systems Parts I and II", *Bell Syst. Tech. J.*, vol. 52, no. 6, pp. 843-886, Jul./Aug. 1973.

3. K. F. Li, D. S. Ong, J. P. R. David, G. J. Rees, R. C. Tozer, P. N. Robson, R. Grey, "Avalanche multiplication noise characteristics in thin GaAs \${rm p}^{+}-{rm i-n}^{+}\$ diodes", *IEEE Trans. Electron Devices*, vol. 45, no. 10, pp. 2102-2107, Oct. 1998.

4. M. A. Saleh, M. M. Hayat, P. Sotirelis, A. L. Holmes, J. C. Campbell, B. E. A. Saleh, M. C. Teich, "Impact-ionization and noise characteristics of thin III-V avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 48, pp. 2722-2731, Dec. 2001.

5. C. H. Tan, J. P. R. David, S. A. Plimmer, G. J. Rees, R. C. Tozer, R. Grey, "Low multiplication noise thin \${hbox {Al}}_{0.6}{hbox {Ga}}_{0.4}{hbox {As}}\$ avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 48, no. 7, pp. 1310-1317, Jul. 2001.

6. S. R. Forrest, M. DiDomenico, Jr., R. G. Smith, H. J. Stocker, "Evidence of tunneling in reverse-bias III-V photodetector diodes", *Appl. Phys. Lett.*, vol. 36, no. 7, pp. 580-582, Apr. 1980.

7. M. M. Hayat, B. E. A. Saleh, J. A. Gubner, "Bit-error rates for optical receivers using avalanche photodiodes with dead space", *IEEE Trans. Commun.*, vol. 43, pp. 99-107, Jan. 1995.

8. C. Groves, J. David, "Effects of ionization velocity and dead space on avalanche photodiode bit error rate", *IEEE Trans. Commun.*, vol. 55, pp. 2152-2158, 2007.

9. P. Sun, M. M. Hayat, B. E. A. Saleh, M. C. Teich, "Statistical correlation of gain and buildup time in APD and its effects on receiver performance", *J. Lightw. Technol.*, vol. 24, no. 2, pp. 755-768, Feb. 2006.

10. P. Sun, M. M. Hayat, A. K. Das, "Bit error rates for ultrafast APD based optical receivers: Exact and large deviation based asymptotic approaches", *IEEE Trans. Commun.*.

11. K. B. Letaief, J. S. Sadowsky, "Computing bit-error probabilities for avalanche photodiode receivers by large deviations theory", *IEEE Trans. Inf. Theory*, vol. 38, pp. 1162-1169, May 1992.

12. B. Choi, M. M. Hayat, "Computation of bit-error probabilities for optical receivers using thin avalanche photodiodes", *IEEE Comm. Lett.*, vol. 10, pp. 56-58, Jan. 2006.

13. L. J. J. Tan, J. S. Ng, C. H. Tan, J. P. R. David, "Avalanche noise characteristics in sub-micron InP diodes", *IEEE J. Quantum Electron.*, vol. 44, no. 4, pp. 378-382, Apr. 2008.

14. G. P. Agrawal, "Optical Receivers" in Fiber-Optic Communication Systems, New York:Wiley, pp. 172, 1997.

15. S. R. Forrest, R. F. Leheny, R. E. Nahory, M. A. Pollack, "\${rm In}_{0.53}{hbox {Ga}}_{0.47}{hbox {As}}\$ photodiodes with dark current limited by generation-recombination and tunnelling", *Appl. Phys. Lett.*, vol. 37, no. 3, pp. 322-325, 1980.

16. J. S. Ng, C. H. Tan, B. K. Ng, P. J. Hambleton, J. P. R. David, G. J. Rees, A. H. You, D. S. Ong, "Effect of dead space on avalanche speed", *IEEE Trans. Electron Devices*, vol. 49, no. 4, pp. 544-549, Apr. 2002.

17. A. Maxim, "A \$54 {hbox {dB}}Omega+42 {hbox {dB}}\$ 10 Gb/s SiGe transimpedance-limiting amplifier using bootstrap photodiode capacitance neutralization and vertical threshold adjustment", *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 1851-1864, Sep. 2007.

18. C.-F. Liao, S.-I. Liu, "40 Gb/s transimpedance-AGC amplifier and CDR circuit for broadband data receivers in 90 nm CMOS", *IEEE J. Solid-State Circuits*, vol. 43, no. 3, pp. 642-655, Mar. 2008.

19. H. Fukuyama, K. Sano, K. Murata, H. Kitabayashi, Y. Yamane, T. Enoki, H. Sugahara, "Photoreceiver module using an InP HEMT transimpedance amplifier for over 40 Gb/s", *IEEE J. Solid-State Circuits*, vol. 39, no. 10, pp. 1690-1696, Oct. 2004.

20. J.-D. Jin, S. S. H. Hsu, "A 40-Gb/s transimpedance amplifier in 0.18- \$mu{hbox {m}}\$ CMOS technology", *IEEE J. Solid-State Circuits*, vol. 43, no. 6, pp. 1449-1457, Jun. 2008.

21. J. S. Weiner, J. S. Lee, A. Leven, Y. Baeyens, V. Houtsma, G. Georgiou, Y. Yang, J. Frackoviak, A. Tate, R. Reyes, R. F. Kopf, W.-J. Sung, N. G. Weimann, Y.-K. Chen, "An InGaAs-InP HBT differential transimpedance amplifier with 47-GHz bandwidth", *IEEE J. Solid-State Circuits*, vol. 39, no. 10, pp. 1720-1723, Oct. 2004.

22. L. W. Cook, G. E. Bulman, G. E. Stillman, "Electron and hole impact ionization coefficients in InP determined by photomultiplication measurements", *Appl. Phys. Lett.*, vol. 40, no. 7, pp. 589-591, Apr. 1982.