Statistical Correlation of Gain and Buildup Time in APDs And Its Effects on Receiver Performance

Peng Sun

Majeed M. Hayat

Bahaa E.A. Saleh

Malvin Carl Teich

Electrical and Computer Engineering Faculty Research and Publications/College of Engineering

Statistical Correlation of Gain and Buildup Time in APDs And Its Effects on Receiver Performance

Peng Sun
Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM

Majeed M. Hayat
Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM

Bahaa E.A. Saleh
Department of Electrical and Computer Engineering, Boston University, Boston, MA

Malvin Carl Teich
Department of Electrical and Computer Engineering, Physics, and Biomedical Engineering, Boston University, Boston, MA

Abstract:

This paper reports a novel recurrence theory that enables us to calculate the exact joint probability density function (pdf) of the random gain and the random avalanche buildup time in avalanche photodiodes (APDs)

including the effect of dead space. Such calculations reveal a strong statistical correlation between the gain and the buildup time for all widths of the multiplication region. To facilitate the calculation of the photocurrent statistics in the presence of this correlation, the impulse-response function of the APD is approximately modeled by a function of time whose prespecified shape is appropriately parameterized by two random variables: the gain and the buildup time. The evaluation of the variance of the photocurrent under this model leads to the definition of the shot-noise-equivalent bandwidth of the APD, which captures the statistical correlation between the gain and the buildup time. It is shown that the shot-noise-equivalent bandwidth in GaAs APDs is greater, by approximately 30%, than the traditional buildup-time-limited 3-dB bandwidth, which is calculated from the mean of the impulse-response function. A thorough analysis of the performance of APD-based integrate-and-dump digital receivers reveals that the strong correlation between the gain and the buildup time accentuates intersymbol interference (ISI) noise, and thus, adversely affects receiver sensitivity at high transmission rates beyond previously known limits.

Keywords

Performance gain, Bandwidth, Photoconductivity, Intersymbol interference, Probability density function, Avalanche photodiodes, Statistics, Shape, Random variables, Gallium arsenide

SECTION I. Introduction

Avalanche photodiodes (APDs) are widely used photodetectors in many high-speed optical receivers, including those deployed in 10-Gb/s/channel lightwave systems. The popularity of APDs is due to their ability to provide high internal optoelectronic gains. In high-speed systems, the gain that an APD provides translates into improved receiver sensitivity, as the gain combats the Johnson noise in the preamplifier stage of an optical receiver. This benefit, however, comes at the often tolerable expense of an increase in shot noise by a factor called the APD's excess noise factor F, which is a measure of the gain uncertainty [1]–[2][3][4]. In addition, the APD's avalanche buildup time, which is the duration of the APD's single-photo-excited impulse-response function, limits the APD's bandwidth and causes intersymbol interference (ISI) in digital communication. The buildup time has heretofore been the factor that limits the use of APDs in 40-Gb/s systems to this date.

Advances in APD technology, such as the development of APDs that may feature evanescent light coupling into very thin multiplication regions (where both carrier transit time and avalanche buildup time are significantly reduced without much sacrifice in coupling efficiency), may be a promising path for bringing APDs to the 40-Gb/s arena [5]–[6][7][8][9][10][11][12][13][14]. On the other hand, to know the exact device requirements for specified system performance requires (yet to be developed) exact theoretical models for assessing the receiver performance at high speeds, where factors such as ISI and the stochastic nature of the APD's impulse-response function play critical roles. In this paper, we exactly compute the joint statistics of the random gain and random buildup time in APDs for the first time and incorporate these statistics in the analysis of APD-based receivers at high transmission speeds. This theory is expected to impact the way we translate system requirements (e.g., those for next-generation lightwave systems) to device requirements. Moreover, this paper may also impact the design and performance of emerging strategies (such as equalization and forward-error correction) used to mitigate ISI and polarization mode dispersion in high-speed lightwave systems.

The APD's impulse-response function is a stochastic process, with a random duration (RD) (viz., avalanche buildup time) and a random area representing the multiplication factor, or gain. Moreover, the random gain and random buildup time are statistically correlated. One manifestation of this correlation is the APD's gain-bandwidth product (GBP), traditionally calculated as the product between the mean gain and the 3-dB bandwidth (calculated from the Fourier transform of the average impulse-response function). However, the GBP conveys the correlation in the gain and the buildup time in a very limited way—being the product of mean

quantities. In particular, it does not capture the inherent correlation in the fluctuations in the gain and the buildup time. For example, even if an APD is operated to yield a certain fixed mean gain, specific realizations of the avalanche process that exhibit high gains are accompanied with long buildup times (those realizations result in greater ISI and adversely affect receiver sensitivity), and vice versa. Consequently, at transmission rates near the APD's bandwidth, where buildup-time uncertainty is expected to affect ISI, it is plausible to suspect that the statistical correlation between the random gain and the random buildup time may play an intricate role in receiver performance.

To understand the complex interplay between the buildup time and the gain, one must appeal to the statistical analysis of the APD's impulse-response function. For example, to calculate the variance of the photocurrent, knowledge of the second moment of the impulse-response function is required at each time. Further, in order to calculate the variance of the integrated photocurrent, as required when assessing the bit error rate (BER) of an integrate-and-dump receiver, the autocorrelation function of the impulse response is also necessary [15], [16]. Although recurrence relations for these quantities are available, their solutions require intensive computing, especially for the autocorrelation function [16]. Moreover, calculation of higher order statistics (e.g., the probability distribution of the photocurrent, which is required in the calculation of the probability distribution of the output of an integrate-and-dump receiver) is practically impossible with existing models.

In this paper, we also provide a novel approximate method for calculating the statistics of the impulse-response function using our exact knowledge of the joint statistics of the random gain and buildup time. The rationale is to first approximate the random impulse-response function by a function of time whose prespecified shape is parameterized by the random gain and the random buildup time. A simple example of such a function is a rectangular RD (RD-R) function whose duration is the buildup time and whose area is proportional to the gain (more complex functions may also be considered). The exact calculation of the joint probability distribution function (PDF) of the gain and buildup time is accomplished by developing a novel recursive theory that generalizes the existing recursive techniques for computing the marginal PDFs of the gain [17], [18] and the buildup time [19]. The computations are carried out for homojunction APDs under the simplifying assumptions of a constant electric field (in the multiplication region) and a constant drift velocity for carriers. We emphasize that the theory developed here incorporates the dead-space effect, which is inherent in the process of cascaded impact ionizations. Dead space is the minimum distance a newly generated carrier must travel before it becomes capable of effecting an impact ionization. It is well known that this effect becomes increasingly more important in thin multiplication layers, where the fast buildup time is accompanied by reduced excess noise factor.

We also utilize this avalanche multiplication theory to analyze the receiver performance in an on–off keying (OOK) optical receiver. We particularly investigate the effect of the statistical correlation in the gain and the buildup time on the receiver performance in high-speed transmission settings.

SECTION II. Joint Probability Distribution of Gain and Buildup Time

Consider an APD with a multiplication region extending from $x = 0$ to $x = w$. Assume that an electron is injected into the multiplication region from the left at $x = 0$, thereby starting the avalanche process. Upon entering the multiplication region, the injected electron travels to the right a random distance $X_{e,0}$, after which it impact ionizes, resulting in two offspring electrons and a hole. The probability density function (pdf) of the free path $X_{e,0}$ of the injected electron is denoted by $h_{e,0}(\xi)$. For example, according to the dead-space multiplication theory (DSMT) [20], $h_{e,0}(\xi) = \alpha(\xi)\exp\left[-\int_{d_{e,0}}^{\xi} \alpha(y)dy\right]u(\xi - d_{e,0})$, where $\alpha(\xi)$ is the position-dependent (field-dependent) nonlocalized ionization coefficient, $d_{e,0}$ is the dead space associated with the injected electron, and u(x) is the unit step function [i.e., $u(x) = 1$ if $x \geq 0$, and zero otherwise]. After each

impact ionization, a newly created electron (or adversely, hole) born at position x travels a free distance $X_{e,x}$ to the right (or adversely, a distance $X_{h,x}$ to the left) before it impacts ionizes. Let $h_e(\xi|x)$ and $h_h(\xi|x)$ denote the pdfs of $X_{e,x}$ and $X_{h,x}$, respectively [in the DSMT model [20], $h_e(\xi|x) = \alpha(x + \xi)\exp\left[-\int_{d_e(x)}^{\xi}\alpha(x + y)dy\right]u(\xi - d_e(x))$ and $h_h(\xi|x) = \beta(x - \xi)\exp\left[-\int_{d_h(x)}^{\xi}\alpha(x - y)dy\right]u(\xi - d_h(x))$, where $d_e(x)$ and $d_h(x)$ are, respectively, the position-dependent (field-dependent) dead spaces associated with the electron and holes that are created at x].

For a linear-mode operation of the APD (i.e., when the APD is operated below breakdown, that is, the probability of avalanche breakdown is zero, or equivalently, the mean gain is finite), the cascade of impact ionizations terminate at some finite random time T, called the avalanche buildup time, yielding a net random gain G. Note that if we consider the electrical current, or the APD's impulse-response function due to the injected electron, resulting from the carriers generated by the cascade of impact ionizations, then T is precisely the RD of the impulse-response function. Our goal is to determine $f_{G,T}(m, t) = P\{G = m, T \le t\}$, the joint PDF of the random variables G and T, where m is the number of electron-hole pairs involved in the avalanche buildup, and t is the time by which the avalanche buildup is completed.

We first define certain intermediate quantities that are essential in the formulation of the recursive (renewal) equations that would ultimately yield $f_{G,T}(m, t)$. For each $x \in [0, w]$, let $Z(x)$ be the total number of carriers (electrons and holes, collectively) generated by a parent electron positioned at x, and let $T_e(x)$ be the random time until all these carriers exit the multiplication region. Similarly, we define $Y(x)$ as the total number of carriers generated as a result of a parent hole positioned at x and let $T_h(x)$ denote the time until all these carriers exit the multiplication region. Note that $G = 0.5(Z(0) + 1)$ and $T_e(0) = T$. Let us now define the joint PDF of $Z(x)$ and $T_e(x)$ by $f_e(m, t; x) = P\{Z(x) = m, T_e(x) \le t\}$. Similarly, define $f_h(m, t; x) = P\{Y(x) = m, T_h(x) \le t\}$. Note that, as a special case, $f_{G,T}(m, t) = f_e(2m - 1, t; 0)$. We now invoke a renewal argument that will allow us to recursively characterize $f_e(m, t; x)$ and $f_h(m, t; x)$

A. Recursive Equations

In avalanching, once a parent carrier impact ionizes, the regenerated parent carrier and the offspring carriers independently repeat a similar process as their parent. Suppose that the parent electron at x first impacts the ionization after traveling a distance $X_e = \xi \in [0, w - x]$ from x. In such an event, there would be two newly created carriers (an electron and a hole). Now, the key observation is that, conditional on the occurrence of this initial ionization, the event that $T_e(x)$ is less than or equal to t and $Z(x) = m$ is precisely the event that the responses due to the newly created carriers all terminate in the remaining time (i.e., t less the electron transport time from x to $x + \xi$, or simply $t - \frac{\xi}{v_e}$) and that the three carriers collectively generate m total offsprings within the remaining time of $t - \frac{\xi}{v_e}$. By invoking the well-known convolution principle for the probability mass function of sums of independent random variables, we obtain the convolution in (1) and (2). Here, v_e is the electron saturation velocity in the multiplication region. By means of this renewal argument and by using standard properties of conditional expectations and sums of independent random variables, we can see that by conditioning on the occurrence of the initial ionization at ξ, the conditional PDF $f_e(m, t; x|\xi)$ can be written as

$$f_e(m, t; x|\xi) = f_e\left(m, t - \frac{\xi}{v_e}; x + \xi\right) * f_e\left(m, t - \frac{\xi}{v_e}; +\xi\right)$$
$$* f_h\left(m, t - \frac{\xi}{v_e}; x + \xi\right)$$

(1)

where $*$ denotes discrete convolution (in the variable m). Similarly, the conditional PDF $f_h(m, t; x|\xi)$ can be written as

$$f_h(m, t; x|\xi) = f_h\left(m, t - \frac{\xi}{v_h}; x - \xi\right) * f_h\left(m, t - \frac{\xi}{v_h}; x - \xi\right)$$
$$* f_e\left(m, t - \frac{\xi}{v_h}; x - \xi\right) \tag{2}$$

where v_h is the hole saturation velocity in the multiplication layer.

The condition on the location ξ of the first impact ionization can now be removed by averaging over all possible locations ξ in the interval $[0, w - x]$. However, it is possible that the parent electron (or hole) born at x does not impact ionize at all, corresponding to the event $\{X_e > w - x\}$ (or $\{X_h > x\}$). In particular, if $X_e > w - x$ (or $X_h > x$), then $T_e(x) = \frac{w-x}{v_e}$ [or $T_h(x) = \frac{x}{v_h}$] and $Z(x) = 1$ [or $Y(x) = 1$].

With the above considerations and after performing an average over all possible locations of the first ionization, we obtain the following coupled recursive equations that together characterize $f_e(m, t; x)$ and $f_h(m, t; x)$:

$$f_e(m, t; x)$$
$$= g_e(x, t)\delta_{m-1}$$
$$+ \int_0^{w-x} \{f_e(m, t - \frac{\xi}{v_e}; x + \xi) * f_e\left(m, t - \frac{\xi}{v_e}; x + \xi\right)$$
$$* f_h(m, t - \frac{\xi}{v_e}; x + \xi)\}h_e(\xi|x)d\xi$$

$$f_h(m, t; x)$$
$$= g_h(x, t)\delta_{m-1}$$
$$+ \int_0^{x} \{f_h(m, t - \frac{\xi}{v_h}; x - \xi) * f_h\left(m, t - \frac{\xi}{v_h}; x - \xi\right)$$
$$* f_e(m, t - \frac{\xi}{v_h}; x - \xi)\}h_h(\xi|x)d\xi$$

where $\delta(k)$ is equal to one if $k = 0$, and zero otherwise

$$g_e(x, t) = (1 - \int_x^w h_e(\xi|x)d\xi) \, u\left(t - \frac{w-x}{v_e}\right) \tag{5}$$

and

$$g_h(x, t) = (1 - \int_0^x h_h(\xi|x)d\xi)u\left(t - \frac{x}{v_h}\right). \tag{6}$$

The convolutions in the above equations can be converted into products by taking the z transform ($0 \leq |z| \leq 1$) of $f_e(m, t; x)$ and $f_h(m, t; x)$ with respect to the variable m. The transformed quantities, denoted by $F_e(z, t; x)$ and $F_h(z, t; x)$, respectively, will satisfy

$$F_e(z, t; x) = g_e(x, t)z + \int_0^{w-x} F_e^2 \left(z, t - \frac{\xi}{v_e}; x + \xi \right)$$
$$\times F_h \left(z, t - \frac{\xi}{v_e}; x + \xi \right) h_e(\xi | x) d\xi \tag{7}$$

and

$$F_h(z, t; x) = g_h(x, t)z + \int_0^x F_h^2 \left(z, t - \frac{\xi}{v_h}; x - \xi \right)$$
$$\times F_e \left(z, t - \frac{\xi}{v_h}; x - \xi \right) h_h(\xi | x) d\xi. \tag{8}$$

In actuality, the energy states of the carriers after ionizations are not identical. The carriers with excess energy above the ionization threshold do not necessarily end up with zero energy after ionization. Therefore, they may travel a relatively shorter dead-space length even in a constant field. Das and Deen [21] accommodated this effect in their work by assuming that the excess energy after ionization is equally distributed. Alternatively, the effect of uncertainty in energy states of the carriers can be considered, for example, through adopting a stochastic model for the carrier's dead space, as done in [22]. For simplicity and brevity, in this paper, we assume zero energy state of the carriers after each ionization as well as a deterministic dead space. We next introduce a numerical recipe for computing F_e, F_h, f_e, and f_h.

B. Numerical Solutions

Calculation of F_e and F_h: For any fixed z, we can solve (7) and (8) numerically using the simple iterative method described below. We first select a maximum limit t_{max} for the range of the buildup time to be considered. We then select a mesh size for the time t and the space x allowing the discretization of the function F_e and F_h, and hence, converting the integrals into summations. The flowchart is presented in Fig. 1. We set the zeroth iteration of the functions $F_e(z, t; x)$ and $F_h(z, t; x)$, denoted by $F_e^{(0)}(z, t; x)$ and $F_h^{(0)}(z, t; x)$, respectively, to be $g_e(x, t)z$ and $g_h(x, t)z$, for $0 \leq x \leq w$ and $0 \leq t \leq t_{max}$. Next, the first iterates $F_e^{(1)}(z, t; x)$ and $F_h^{(1)}(z, t; x)$ are computed by substituting the zeroth iterates into the right-hand side of (7) and (8). Subsequent iterates are generated in the same way. This procedure is continued until the iterates converge uniformly in t and x. More precisely, we stop the iteration process when the maximum relative change from the $(n - 1)$th to the nth iterate in the functions $F_e^{(n)}(Z, t; x)$ and $F_h^{(n)}(Z, t; x)$ is below a predefined tolerance level. At such point, the procedure for computing F_e and F_h is complete and we set them to the respective values at the last round of iterations.
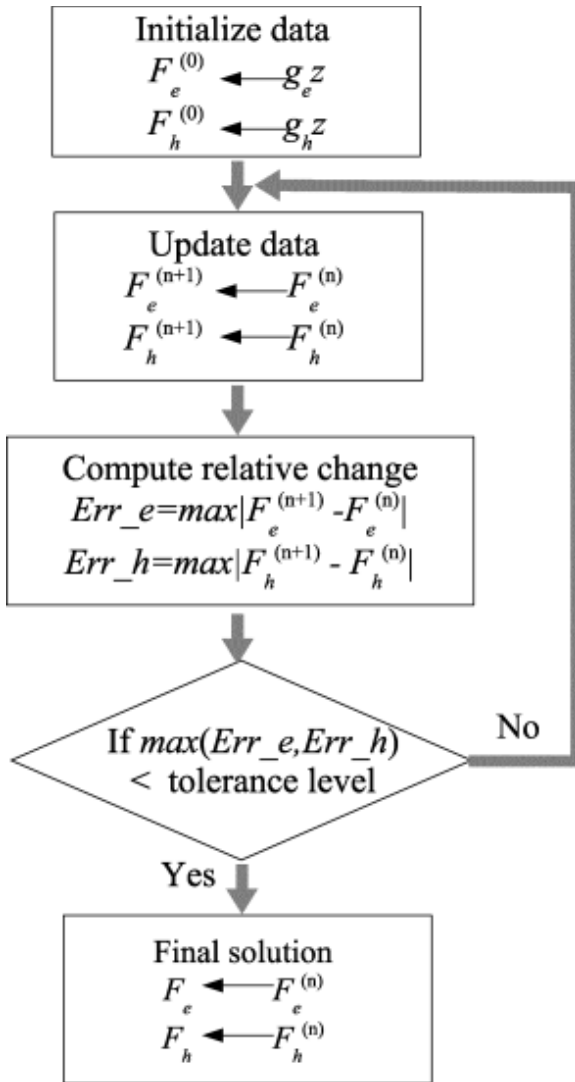
Fig. 1. Flowchart describing the iterative procedure (Picard iterations) used to solve the renewal equations.

Calculation of $f_{G,T}$: Let $F_{G,T}(z,t)$ be the z transform of $f_{G,T}(m,t)$ with respect to m. Then, since $f_{G,T}(m,t) = f_e(2m-1,t;0)$, we obtain

$$F_{G,T}(z,t) = \sqrt{z}F_1(\sqrt{z},t;0)$$

which we need to convert back to the m domain. Since G is an integer-valued random variable, we can apply the efficient technique proposed by Gubner and Hayat [23] to recover the $f_{G,T}(m,t)$ from $F_{G,T}(z,t)$, with z restricted to the unit circle.

In this paper, the electric field required for achieving a certain average gain is obtained by using the direct approach reported in [17], with the ionization coefficients, dead spaces, and ionization-threshold energies taken from [24]. Subsequently, these electric fields and their corresponding ionization coefficients and dead spaces are utilized in the proposed renewal equations to obtain the solution of the joint PDF.

Fig. 2 shows an example of the joint PDF $f_{G,T}(m,t)$ calculated for a homojunction GaAs APD with a 160-nm multiplication layer exhibiting an average gain of 10.46 (from the direct approach). The electron and hole saturation velocities are assumed to be 0.67×10^7 cm/s (from [25]) and the nonlocalized ionization coefficients for electrons and holes and their respective ionization-threshold energies are taken from [24]. The joint pdf,

shown in Fig. 3, is obtained by taking the derivative of $f_{G,T}(m,t)$ with respect to t. Recall that if we restrict the random gain realization to unity (the cross section corresponding to $m = 1$ in Fig. 2), then the PDF of T becomes a step function with a jump at the electron transit time $\frac{w}{v_e}$. Consequently, the joint pdf of T is expected to exhibit a Dirac impulse at the electron transit time. These two features are observed in Figs. 2 and 3, respectively. Fig. 3 also shows that as the realized value m of the random gain increases, the pdf of the T (i.e., cross section of the joint pdf along the t-axis) shifts toward a higher buildup-time mean and a larger spread about the mean. These are clear indications of the statistical correlation between the gain and the buildup time as well as the increase in the buildup-time uncertainty at higher realizations of the gain.
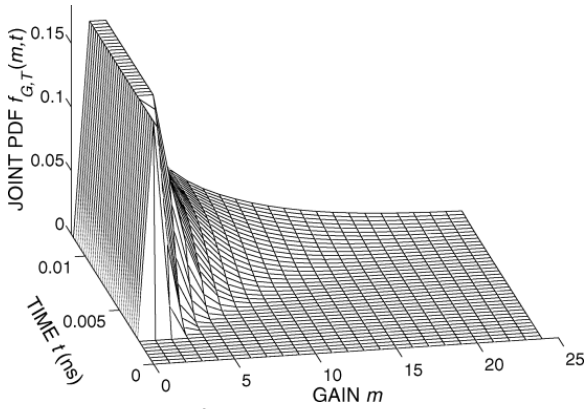

Fig. 2. Joint PDF of the random gain G and the random buildup time T for a GaAs APD with a 160-nm multiplication layer. The applied electric field in the multiplication region is taken as 5.47×10^3 kV/cm, yielding a theoretical average gain of $\langle G \rangle = 10.46$.
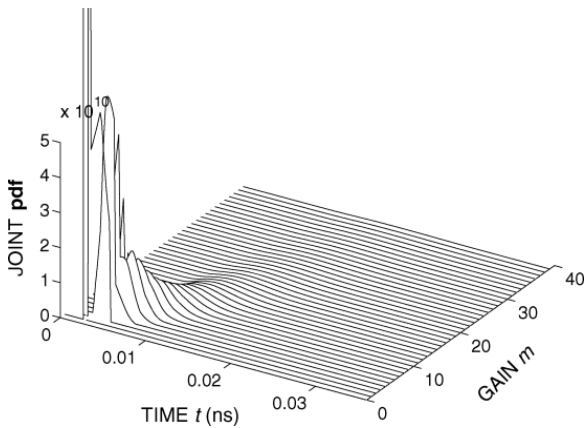

Fig. 3. Joint pdf of the random gain G and the random buildup time T for the APD considered in Fig. 2. In order to better show the details of the pdf, large peaks have been truncated.

In this example, the average gain calculated by the direct approach is 10.46. The marginal distribution of the gain (obtained by averaging the joint PDF over the buildup time) yields the mean of 10.48, which is in good agreement with the direct-approach result. Additionally, if the time t in (7) and (8) is set to infinity, (7) and (8) become the renewal equations governing the generating function of the random gain, and these renewal equations are equivalent to those reported in [17].

Similarly, the probability mass function of G for various realizations t of the random buildup time T shifts toward a higher mean (accompanied by a larger spread) as t increases. This also confirms the correlation between the gain and the buildup time. Fig. 4 (bottom graph) shows the correlation coefficient ρ of G and T as a function of the width of the multiplication region while the mean gain is held fixed at ten. The correlation coefficient is

calculated by $\rho = \dfrac{E[GT]-E[G]E[T]}{\sigma_G \sigma_T}$, and the values of correlation coefficients are approximately 0.9 for the entire range of w.
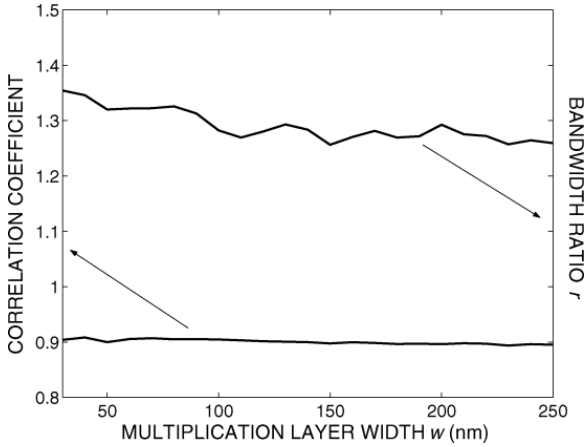

Fig. 4. Bottom: Correlation coefficient ρ of the random gain G and the random buildup time T for a GaAs APD as a function of the multiplication-layer width w. Top: Ratio r of shot-noise-equivalent bandwidth B_{sneq} to the 3-dB bandwidth $B_{\text{3-dB}}$. The mean gain is held constant at $\langle G \rangle = 10$.

We next use the joint PDF of G and T to calculate the statistical properties of the impulse-response function and investigate the effect of the correlation between G and T on ISI and receiver performance in high-speed digital receivers.

SECTION III. Parametric Modeling of the Impulse-Response Function and the Shot-Noise-Equivalent Bandwidth

In this section, we introduce a novel stochastic model for the impulse-response function that facilitates the calculation of its statistics once the joint PDF of the gain and the buildup time has been determined. The rationale is to approximate the impulse-response function $I_{\text{p}}(t)$ by a prespecified shape function of time that is parameterized by the random gain G and the random buildup time T. In this fashion, the randomness of the impulse-response function, as a stochastic process, is lumped in the random parameters G and T. This significantly simplifies the complexity of the impulse-response function while maintaining the key features that govern the excess-noise and speed properties of the APD, namely, randomness in the impulse-response function's area (representing gain uncertainty) and randomness in its duration (representing bandwidth uncertainty). An example of such a shape function is the rectangular RD (RD-rectangular or RD-R) with random height $\dfrac{qG}{T}$ and RD T, as shown in Fig. 5, where q is the electronic charge. Note that the area under this function is qG. Another example is the triangular RD (RD-triangular or RD-T) model, shown in the same figure, for which the RD is again T but a peak occurs at the electron transit time τ_{e} across the multiplication region. The peak in the RD-T model is consistent with the behavior of the true impulse-response function, as all first-generation electrons continue to multiply until they simultaneously exit the multiplication region at precisely τ_{e} units of time past the time of the primary-electron injection. The height of the peak is $\dfrac{2qG}{T}$ so that the area under the shape function is qG. Other shape functions may also be considered.
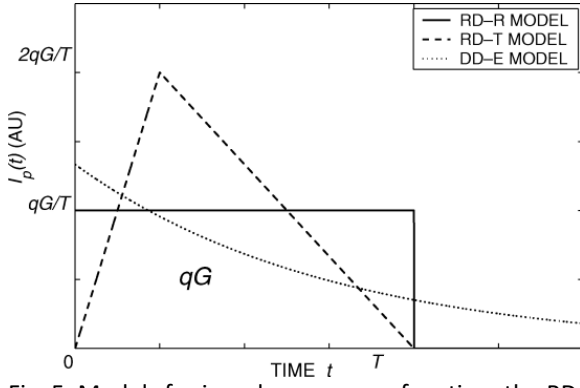
Fig. 5. Models for impulse-response function: the RD-R model (solid curve) and the RD-T model (dotted curve). Note that the area under each of the models is qG. A DD-E is also shown for which the decay rate is deterministic.

To see the validity of the proposed RD-R and RD-T models, we calculated the mean impulse-response function, denoted by $i_p(t) = E[I_p(t)]$, and compared it to the exact theoretical model for the mean impulse-response function [15]. For example, it can be easily shown that for the rectangular model

$$E[I_p(t)] = \sum_{m=1}^{\infty} \int_t^{\infty} qm\tau^{-1} \frac{\partial}{\partial \tau} f_{G,T}(m,\tau)d\tau, t \geq 0$$

which can be easily computed as the joint distribution $f_{G,T}$ has already been computed. Figs. 6 and 7 depict the mean and the second moment of the random impulse-response function obtained using both the rectangular and triangular parametric models. Indeed, the area under $i_p(t)$ is $10.48q$, which is in excellent agreement with the theoretical average gain of $10.46q$ that is calculated independently using that in [24]. The comparison with the theoretical prediction of the APD mean impulse-response function by Hayat and Saleh [15] is also shown on this figure (dashed curve). The RD-R model predicts GBPs of 274 and 194 GHz for the multiplication width 100 and 200 nm, respectively, which are comparable to 292 GHz and 171 GHz, respectively, reported in [16]. Note that, due to the frequency-response characteristics of the rectangular and triangular waveforms, the rectangular model slightly underestimates the bandwidth while the triangular model overestimates it. Nevertheless, the approximation is very good in both cases.
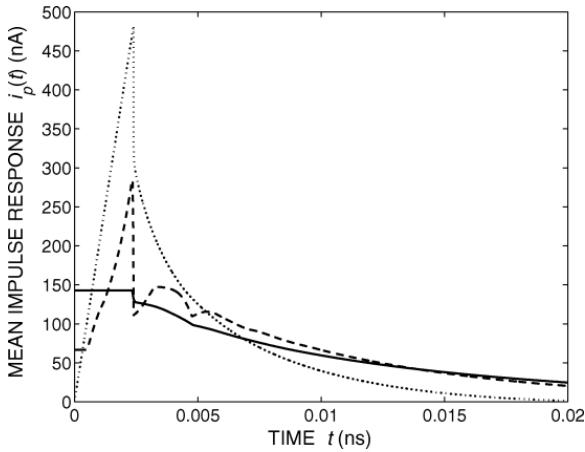

Fig. 6. Mean impulse-response function $i_p(t)$ corresponding to the RD model with a rectangular shape (solid curve), the RD model with a triangular shape (dotted curve), and the exact model [15] (dashed curve). Note that the rectangular-shape RD model slightly underestimates the bandwidth while the triangular-shaped RD model overestimates it.
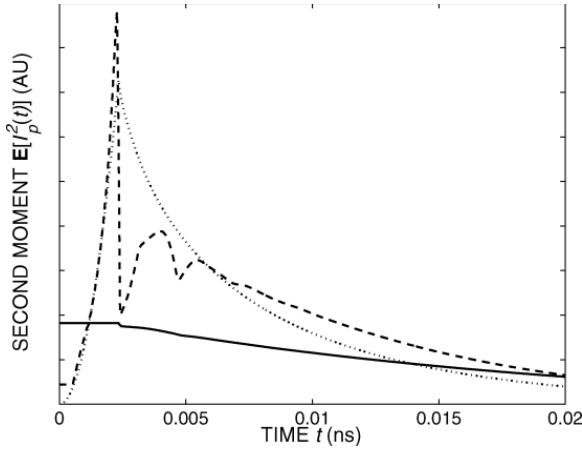
Fig. 7. Second moment of the impulse-response function ip(t) corresponding to the RD model with a rectangular shape (solid curve), the RD model with a triangular shape (dotted curve), and the exact model [15] (dashed curve).

A. Bandwidth

The calculation of the variance of the photocurrent generated by an APD (i.e., the variance of the filtered shot noise [26]) requires knowledge of the second moment of the APD's impulse-response function of time t. In particular, the variance of the photocurrent is given by [27] $q\left(\frac{\eta P}{h\nu}\right)\int_{-\infty}^{\infty}\langle I_{\mathrm{p}}^2(t)\rangle dt$, where P is the optical power, η is the APD's quantum efficiency (the probability that a single photon incident on the device generates a photocarrier pair that contributes to the detector's current), h is Plank's constant, and ν is the photon's frequency. However, calculation of the second-order statistics of the impulse-response function are generally computationally intensive, with no known closed-form expressions available [15], [28]. As a result, simplifying assumptions that ignore the randomness in the shape of the impulse-response function are often practiced in the calculation of the photocurrent variance. For example, one customary approach is to assume that the impulse-response function takes the simplified form of $I_{p,c}(t) = qGh(t)$ [i.e., the deterministic duration (DD) model], where $h(t)$ is a unit-area deterministic function that is proportional to the mean impulse-response function [3]. We emphasize that implicit in such simplifications is the absence of the statistical correlation between the gain and buildup time. For example, the DD model $I_{p,c}(t) = qGh(t)$ implies that both high-gain and low-gain realizations of $I_{p,c}(t)$ yield the same bandwidth, while in actuality, high-gain realizations are accompanied by long buildup times, and hence low bandwidth, and vice versa. Nevertheless, the type of simplifications described above admit the customary closed-form expression for the variance of the photocurrent given by $2B_c q\langle\frac{(G)^2 F\eta P}{h\nu}$, where B_c is the conventional bandwidth of the APD and F is the APD's excess noise factor $\left(F = \frac{\langle G^2\rangle}{\langle G\rangle^2}\right)$. For example, if the DD model $I_{p,c}(t) = qGh(t)$ is used, then $B_c = \frac{\int_{-\infty}^{\infty} h^2(t)dt}{2\left(\int_{-\infty}^{\infty} h(t)dt\right)^2}$ [3], which approximates the 3-dB bandwidth of the APD. The point being made here is that B_c is calculated solely from the mean impulse-response function and it does not capture the statistical correlation between the gain and the buildup time.

In contrast, if the exact expression for the photocurrent variance is used $q\left(\frac{\eta P}{h\nu}\right) \times \int_{-\infty}^{\infty}\langle I_{\mathrm{p}}^2(t)\rangle dt$, then we can continue to conveniently express the shot-noise variance in the customary form of $2qB_{\mathrm{sneq}}\langle G\rangle^2 F\eta P/h\nu$ provided that we use the correct bandwidth B_{sneq}, which we term the shot-noise-equivalent bandwidth. This leads to the definition of B_{sneq} as

$$B_{\text{sneq}} = \frac{\int_{-\infty}^{\infty} \langle I_p^2(t) \rangle dt}{2q^2 \langle G \rangle^2 F}. \quad (9)$$

We emphasize that, in contrast to B_c, B_{sneq} includes the effects of fluctuations in the shape (e.g., gain and duration) of the APD's random impulse-response function. Moreover, a closed-form estimate of the second moment of the APD's impulse-response function can be obtained approximately using the parametric stochastic models for the impulse-response function introduced in Section III. For example, with the RD-R model

$$\langle I_p^2(t) \rangle = \sum_{m=1}^{\infty} \int_t^{\infty} q^2 m^2 \tau^{-2} \frac{\partial}{\partial \tau} f_{G,T}(m, \tau) d\tau, t \geq 0 \quad (10)$$

and the expression for B_{sneq} can be simplified to

$$B_{\text{sneq}} = \frac{\left\langle \frac{G^2}{T} \right\rangle}{2\langle G \rangle^2 F} \equiv \frac{\left\langle \frac{G^2}{T} \right\rangle}{2\langle G^2 \rangle} \quad (11)$$

which can be readily evaluated using our knowledge of the joint distribution $f_{G,T}$ obtained in Section II.

The point of the preceding analysis is that, to incorporate the effect of statistical correlation between the APD's gain and buildup time on the signal-to-noise ratio (SNR) of the photocurrent, we must use B_{sneq} in place of B3-dB in the customary SNR expression $\frac{\eta P}{2h\nu F B_{\text{3-dB}}}$.

We have evaluated the shot-noise-equivalent bandwidth B_{sneq} for GaAs APDs, for various widths of the multiplication layer, and compared it with B_c taken as the 3-dB bandwidth. Fig. 4 (upper curve) shows the ratio $r = \frac{B_{\text{sneq}}}{B_c}$ as a function of the width of the multiplication layer for a fixed mean gain of $\langle G \rangle = 10$. Notably, the shot-noise-equivalent bandwidth exceeds the 3-dB bandwidth by approximately 30%. Thus, our calculations show that the 3-dB bandwidth, which is customarily used in calculating the shot-noise variance, leads to tangible underestimation of the shot-noise variance. This is primarily a result of ignoring the statistical correlation between the buildup time and the gain, or equivalently, it results from ignoring the uncertainty in the shape of the impulse-response function.

SECTION IV. Receiver-Performance Analysis

In this section, we investigate the effect of the correlation between the buildup time and the gain on the performance of an APD-based receiver in a direct-detection OOK optical communication system. Since we are particularly interested in the performance at high transmission speeds, we will assume that the buildup time dominates the *RC* effects and governs the receiver bandwidth characteristics. We first consider the SNR for the output of an integrate-and-dump receiver in the presence of ISI, Johnson noise, and buildup-time and gain uncertainties of the APD. Subsequently, we will examine the receiver sensitivity at high transmission speeds. Our analysis will ignore optical pulse broadening and dispersion effects.

A. Statistics of the Photocurrent and the Integrate-and-Dump Output

Consider an arbitrary bit of duration T_b from a stream of optical pulses, and let $I(t)$ denote the photocurrent in such a bit. The photocurrent $I(t)$ contains two components: $I_c(t)$, the component due to photons occurring in the present bit, and $I_{\text{ISI}}(t)$, the ISI component, which is due to the photons that had arrived in previous bits. More precisely, if we represent the photon arrival times by the sequence $\{t_i\}$ and assume that the current bit extends from 0 to T_b, then for $0 \leq t \leq T_b$, the photocurrent can be written as

$$I(t) = \sum_{t_i<0} I_{p_i}(t - t_i) + \sum_{0 \le t_i < t} I_{p_i}(t - t_i)$$

$$\triangleq I_{\mathrm{ISI}}(t) + I_c(t) \qquad (12)$$

where $I_{p_i}(t - t_i)$ is the impulse-response function induced by a photon absorption at time t_i. Note that $I_c(t)$ and $I_{\mathrm{ISI}}(t)$ are statistically independent since they correspond to disjoint and independent sets of photon absorptions. Suppose that photons are absorbed at a rate of $\phi(\tau)$ photons per second $\phi(\tau) = \frac{\eta P(\tau)}{h\nu}$, where $P(\tau)$ is the received optical power at time τ]. By using standard filtered-shot-noise analysis [26], we can express the mean of I(t) in terms of a convolution as follows:

$$\langle I(t)\rangle = \int_{-\infty}^{0} \langle I_p(t - \tau)\rangle \phi(\tau)d\tau + \int_{0}^{t} \langle I_p(t - \tau)\rangle \phi(\tau)d\tau \qquad .(13)$$

$$\triangleq \langle I_{\mathrm{ISI}}(t)\rangle + \langle I_c(t)\rangle.$$

To see the maximum ISI effect, we consider the case for which all the bits prior to the current bit are on, i.e., $\phi(\tau) \equiv \phi_1$, for all τ, where ϕ_1 is the photon absorption rate in a pulse. It follows that

$$\langle I_c(t)\rangle = \phi_1 \int_{0}^{t} \langle I_p(t - \tau)\rangle d\tau = \phi_1 \int_{0}^{t} \langle I_p(\tau)\rangle d\tau$$

$$\langle I_{\mathrm{ISI}}(t)\rangle = \phi_1 \int_{-\infty}^{0} \langle I_p(t - \tau)\rangle d\tau = \phi_1 \int_{t}^{\infty} \langle I_p(\tau)\rangle d\tau.$$

The photocurrent $I(t)$ is fed into a bit integrator that yields the integral of the photocurrent synchronously over each bit of duration T_b. The integral is used to detect the information (0 or 1) modulated with the incident optical sequences. We denote the output of the integrate-and-dump unit by $\Gamma = \int_{0}^{T_b} I(t)dt$, which can be further decomposed [see (12) –(15)] into two parts: $\Gamma = \Gamma_c + \Gamma_{\mathrm{ISI}}$. The first component is the net signal resulting from the photocurrent due to the absorbed photons in the present bit, and the second component corresponds to ISI resulting from the photocurrent due to the absorbed photons in the previous bits. The means of Γ, Γ_c, and Γ_{ISI} are, respectively

$$\langle \Gamma_c \rangle = \phi_1 \int_{0}^{T_b} \int_{0}^{t} \langle I_p(t - \tau)\rangle d\tau dt$$

$$\langle \Gamma_{\mathrm{ISI}} \rangle = \phi_1 \int_{0}^{T_b} \int_{-\infty}^{0} \langle I_p(t - \tau)\rangle d\tau dt$$

and

$$\langle \Gamma \rangle = \phi_1 \int_{0}^{T_b} \int_{-\infty}^{t} \langle I_p(t - \tau)\rangle d\tau dt. \qquad (18)$$

The variance of Γ is given by

$$\sigma_\Gamma^2 = \phi_1 \int_{0}^{T_b} \int_{0}^{T_b} \int_{-\infty}^{\mu \wedge \nu} R_{I_p}(\mu - \tau, \nu - \tau)d\tau d\mu d\nu \qquad (19)$$

where $\mu \wedge \nu$ is the minimum of μ and ν, $R_{I_p}(\mu, \nu)$ is the autocorrelation function of the impulse-response $I_p(t)$, which is defined as $R_{I_p}(\mu, \nu) = E[I_p(\mu - \xi)I_p(\nu - \xi)]$ [the derivation of (19) utilizes standard analysis of filtered shot-noise processes [26] and will not be presented here].

We now define the maximal-ISI SNR for the integrate-and-dump output as $\langle \Gamma_c \rangle^2 / \sigma_\Gamma^2$. In the next section, we specialize the SNR to two models for the APD's impulse-response function: the RD-R model and the DD-exponential (DD-E) model. This will permit us to perform a comparison that shows the effect of the uncertainty in the shape of the impulse-response function on the receiver SNR.

B. SNR for a DD-E Model

We now use a DD-E model to approximate the APD's impulse-response function: $I_{DD}(t) = qGbe^{-bt}u(t)$ (we ignore the electronic charge). Here, b is the fixed exponential rate consistent with the 3-dB bandwidth B3-dB of the APD [15], [16], namely, $b = 2\pi B_{3\text{-dB}}$. Note that in this model the statistical correlation between the buildup time and the gain is ignored. Clearly, the mean of this impulse-response function is $\langle I_{DD}(t) \rangle = \langle G \rangle be^{-bt}u(t)$, and the autocorrelation function can also be calculated, yielding

$$R_{I_{DD}}(\mu, \nu) = \langle G^2 \rangle b^2 e^{-b(\mu+\nu)} u(\mu)u(\nu).$$

We now substitute $\langle I_{DD}(t) \rangle$ and $R_{I_{DD}}(\mu, \nu)$ into (16) and (19), respectively, and obtain

$$\langle \Gamma_c \rangle = \phi_1 \langle G \rangle \left[T_b - \frac{1}{b}(1 - e^{-bT_b}) \right]$$

$$\sigma_\Gamma^2 = \phi_1 \langle G^2 \rangle \left[T_b - \frac{1}{b}(1 - e^{-bT_b}) \right].$$

Note that only the marginal statistics for the random variable G are used in above equations. Finally, we replace b by $2\pi B_{3\text{-dB}}$ and obtain

$$\text{SNR}_{DD} = \frac{\phi_1 T_b}{F} \left(\frac{2\pi T_b B_{3-dB} - 1 + e^{-2\pi T_b B_{3-dB}}}{2\pi T_b B_{3\text{-dB}}} \right). \quad (20)$$

Note that the first fraction is the SNR of an instantaneous detector and the second fraction is a scaling factor accounting for ISI effects.

C. SNR for the RD-R Model

As introduced in Section III, the RD-R parametric model for the random impulse-response function is

$$I_{RD}(t) = \left(\frac{G}{T} \right) \{ u(t) - u(t - T) \}.$$

Since the first moment $\langle I_{RD}(t) \rangle$ and the second moment $\langle I_{RD}^2(t) \rangle$ of the impulse-response function (below avalanche breakdown) are known to decay exponentially at the same rate [15], b_{se} is referred to the decay rate (we will see later that b_{se} is related with B_{sneq}). We approximate them by

$$\langle I_{RD}(t) \rangle \approx ae^{-b_{se}t} \text{ and } \langle I_{RD}^2(t) \rangle \approx ce^{-b_{se}t}. \quad (21)$$

If we substitute these approximations in (16), we obtain

$$\langle \Gamma_c \rangle = \phi \frac{a}{b_{se}} \left(T_b - \frac{1}{b_{se}} (1 - e^{-b_{se}T_b}) \right). \text{(22)}$$

Next, to calculate the second moment of the photocurrent in (19), we need to first determine the autocorrelation function $R_{I_{RD}}(\mu, \nu) = E[I_{RD}(\mu)I_{RD}(\nu)]$, which can be approximated by (given in Appendix A)

$$R_{I_{RD}}(\mu, \nu) \approx c e^{-b_{se}(\mu \vee \nu)}. \text{(23)}$$

If we now substitute the above autocorrelation function into (19), we obtain the expression for the variance of the receiver output

$$\sigma_\Gamma^2 \approx \phi_1 \frac{2c}{b_{se}^2} \left[T_b - \frac{1}{b_{se}} (1 - e^{-b_{se}T_b}) \right] \text{(24)}$$

and the SNR according to the RD-R model can be cast as

$$\text{SNR}_{RD} = \frac{\langle \Gamma_c \rangle^2}{\sigma_\Gamma^2} = \frac{\phi \frac{a^2}{b_{se}^2} \left[T_b - \frac{1}{b_{se}} (1 - e^{-b_{se}T_b}) \right]}{\frac{2c}{b_{se}^2}}. \text{(25)}$$

Our next step is to explore the physical meaning of the parameters a, b_{se}, and c based on our rectangular model. To do so, we note that

$$\int_0^\infty \langle I_{RD}(t) \rangle \, dt \approx \quad \int_0^\infty a e^{-b_{se}t} dt = \frac{a}{b_{se}} \text{ and}$$

$$\left\langle \int_0^\infty I_{RD}(t) dt \right\rangle = \quad \left\langle \int_0^\infty \frac{G}{T} \{u(t) - u(t - T)\} dt \right\rangle = \langle G \rangle \text{ (26)(27)}$$

and use the fact that the left sides of the above two equations are equal to conclude that $\frac{a}{b_{se}} \approx \langle G \rangle$. Similarly

$$\int_0^\infty \langle I_{RD}^2(t) \rangle \, dt \approx \quad \int_0^\infty c e^{-b_{se}t} dt = \frac{c}{b_{se}} \text{ and}$$

$$\left\langle \int_0^\infty I_{RD}^2(t) dt \right\rangle = \quad \left\langle \int_0^\infty \frac{G^2}{T^2} \{u(t) - u(t - T) dt\} \right\rangle = \left\langle \frac{G^2}{T} \right\rangle \text{ (28)(29)}$$

which results in $\frac{c}{b_{se}} \approx \left\langle \frac{G^2}{T} \right\rangle$.

We now proceed to characterize b_{se}. Note that the approximation of $\langle I_{RD}(t) \rangle \approx a e^{-b_{se}t}$ is based on the rectangular parameterized model, which takes into account the correlation of G and T. Thus, the effective bandwidth rendered by its shape (which is $\frac{b_{se}}{4}$) should be identical to the shot-noise-equivalent bandwidth B_{sneq}. We therefore have $b_{se} = 4B_{sneq}$. By substituting the approximation $\frac{a}{b_{se}} \approx \langle G \rangle$ and $b_{se} = 4B_{sneq}$ in (25), we can recast the SNR as

$$\text{SNR}_{RD} = \frac{\phi_1 \langle G \rangle^2}{\left\langle \frac{G^2}{T} \right\rangle} \frac{\left(4T_b B_{sneq} - 1 + e^{-4T_b B_{sneq}} \right)}{2}. \text{(30)}$$

Note that the joint statistics of G and T are used in the above equation. Finally, recall that $B_{\mathrm{sneq}} = \langle G^2/T \rangle / 2\langle G \rangle^2 F$ (see Section III-A) and define a correction factor $\kappa = \frac{4B_{\mathrm{sneq}}}{2\pi B_{\text{3-dB}}}$ to obtain

$$\mathrm{SNR}_{\mathrm{RD}} = \frac{\phi_1 T_{\mathrm{b}}}{F} \left(\frac{2\pi\kappa T_{\mathrm{b}} B_{\text{3-dB}} - 1 + e^{-2\pi\kappa T_{\mathrm{b}} B_{\text{3-dB}}}}{2\pi\kappa T_{\mathrm{b}} B_{\text{3-dB}}} \right). \tag{31}$$

In the above formula, the first fraction is the SNR for an instantaneous detector for which case the ISI is absent; the second fraction represents the correction factor due to ISI. Note that the statistical correlation between the buildup time and the gain is incorporated through the parameter B_{sneq}.

More general expressions for the SNR, taking into account the randomness of the sequence of 1's and 0's (as in an OOK setting), are derived in Appendix B. Results corresponding to these generalized expressions for the SNR will be in the next section.

D. Results: ISI and Gain–Buildup-Time Correlation Effects on the SNR

We now examine the ISI correction factor of the SNR expression given by (31) (the quantity in parenthesis) and observe that it monotonically varies from 1 to 0 as the product $T_{\mathrm{b}} B_{\mathrm{sneq}}$, which is a measure of the transmission speed relative to the APD's buildup time, varies from ∞ to 0. Recall that it was shown in Section III-A that the ratio $r = \frac{B_{\mathrm{sneq}}}{B_{\text{3-dB}}}$ is approximately 1.3 (for the APD parameters considered), which implies that $\kappa < 1$.

Consequently, by comparing the SNR expression given in (20)(which corresponds to the case when the correlation between the gain and the buildup time is ignored) and $\mathrm{SNR}_{\mathrm{RD}}$, and by using the monotonicity of the ISI factor, we conclude that $\mathrm{SNR}_{\mathrm{DD}} < \mathrm{SNR}_{\mathrm{RD}}$. Namely, the correlation between the gain and buildup time adversely affects the SNR. In what follows, we will show that this effect becomes more significant at high transmission speeds.

Fig. 8 depicts $\mathrm{SNR}_{\mathrm{DD}}$ and $\mathrm{SNR}_{\mathrm{RD}}$ as a function of the transmission rates. The signal-to-noise expressions are given in Appendix B [see (53), (54), (59), and (60)]. It is assumed that a GaAs APD is being used with multiplication-layer width of $w = 100$nm and an average gain of $\langle G \rangle = 10$. Moreover, the average number of photons absorbed by the APD is assumed to be 1000 per "1" bit, and the Johnson-noise parameter is $\sigma_{\mathrm{J}} = 500$ noise counts per bitnoise counts per bit. We observe that as the transmission rate increases, the effect of ISI becomes progressively more detrimental to the SNR. It is also shown that the relative separation between $\mathrm{SNR}_{\mathrm{RD}}$ and SNR increases with the transmission rate.
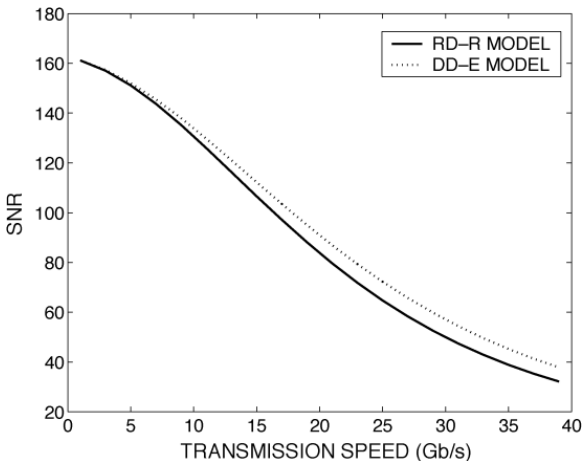

Fig. 8. SNR of the output of the integrate-and-dump receiver as a function of the digital transmission speed for a GaAs APD with a 100-nm multiplication layer and a theoretical mean gain of 10. The solid curve corresponds to the case where the RD-R model is used for the APD's impulse-response function, while the dashed curve corresponds to the DD-E model. The average number of photons is fixed to 1000 photons/bit and the Johnson noise is selected so that $\sigma_{\mathrm{J}} = 500$ noise counts/bit.

SECTION V. BER and Receiver Sensitivity

In this section, we investigate the effect of the statistical correlation between the gain and buildup time on the receiver performance. We use the same APD parameters as those used in the previous section. The APD's 3-dB bandwidth is found to be $B_{3\text{-dB}} = 29$ GHz. We first consider the RD-R model. Let $\mu_{0,\text{RD}}$, $\sigma_{0,\text{RD}}^2$, $\mu_{1,\text{RD}}$, and $\sigma_{1,\text{RD}}^2$ denote the mean and variance corresponding to the cases for which the current bit is 0 and 1, respectively. These parameters are (see Appendix B for derivation)

$$\mu_{0,\text{RD}} = \frac{1}{2}\frac{n_0\langle G\rangle}{\kappa\lambda}\left(1 - e^{-\kappa\lambda}\right)$$

$$\sigma_{0,\text{RD}}^2 = \frac{1}{4}\frac{n_0^2\langle G\rangle^2}{\kappa\lambda^2}\frac{(1 - e^{-\kappa\lambda})^4}{(1 - e^{-2\kappa\lambda})}$$
$$+ \frac{n_0\langle G^2\rangle}{2\kappa\lambda}\left(1 - e^{-\kappa\lambda} - \kappa\lambda e^{-\kappa\lambda}\right) + \sigma_{\text{J}}^2$$

$$\mu_{1,\text{RD}} = \mu_{0,\text{RD}} + \frac{n_0\langle G\rangle}{\kappa\lambda}\left(\kappa\lambda - 1 + e^{-\kappa\lambda}\right)$$

$$\sigma_{1,\text{RD}}^2 = \sigma_{0,\text{RD}}^2 + \frac{n_0\langle G^2\rangle}{\kappa\lambda}\left(\kappa\lambda - 2 + 2e^{-\kappa\lambda} + \kappa\lambda e^{-\kappa\lambda}\right)$$

where $\lambda = 2\pi T_{\text{b}}B_{3\text{-dB}}$ and κ is the correction factor defined earlier. Note that $\mu_{0,\text{RD}}$ and $\sigma_{0,\text{RD}}^2$ (less σ_{J}^2) are entirely due to ISI.

In contrast to the above statistics, the corresponding quantities for the DD-E model for which the correlation between the gain and the buildup time is ignored can be found to be (we omit the derivation)

$$\mu_{0,\text{DD}} = \frac{1}{2}\frac{n_0\langle G\rangle}{\lambda}\left(1 - e^{-\lambda}\right)$$

$$\sigma_{0,\text{DD}}^2 = \frac{1}{4}\frac{n_0^2\langle G\rangle^2}{\lambda^2}\frac{(1 - e^{-\lambda})^4}{(1 - e^{-2\lambda})}$$
$$+ \frac{n_0\langle G^2\rangle}{\lambda}\left(1 - e^{-\lambda}\right)^2 + \sigma_{\text{J}}^2$$

$$\mu_{1,\text{DD}} = \mu_{0,\text{DD}} + \frac{n_0\langle G\rangle}{\lambda}\left(\lambda - 1 + e^{-\lambda}\right)$$

$$\sigma_{1,\text{DD}}^2 = \sigma_{0,\text{DD}}^2 + \frac{n_0\langle G^2\rangle}{2\lambda}\left(2\lambda - (1 - e^{-\lambda})^2\right).$$

As before, $\mu_{0,\text{DD}}$ and $\sigma_{0,\text{DD}}^2$ (less σ_{J}^2) are contributions of ISI.

To simplify the analysis, we will approximate the output of the receiver by a Gaussian random variable. The BER is then obtained according to

$$\text{BER} \approx \frac{1}{2} \left\{ \text{erfc} \left(\frac{\theta - \mu_{0,\text{RD}}}{\sqrt{2}\sigma_{0,\text{RD}}} \right) + \text{erfc} \left(\frac{\mu_{1,\text{RD}} - \theta}{\sqrt{2}\sigma_{1,\text{RD}}} \right) \right\} \quad (40)$$

where θ is the optimal decision threshold. Fig. 9 depicts the BER as a function of the transmission rate calculated for both the DD-E and RD-R models. Here, n_0 is assumed 1000 photons/bit and the Johnson noise is held constant at $\sigma_J = 500$noise counts/bit. The adverse effect of the correlation between the gain and the buildup time is evident in this plot, as seen by the elevated BER curve in the RD-R model case compared to that obtained for the DD-E model. This behavior is also consistent with the SNR plots shown in Fig. 8.
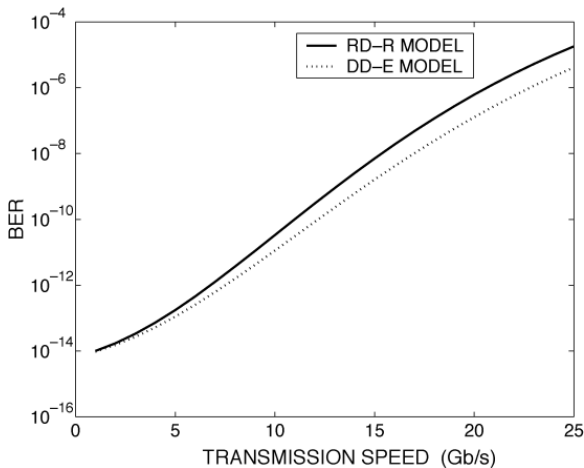

Fig. 9. BER as a function of the transmission speed using the RD-R model (solid curve) and the DD-E model (dashed curve). The mean gain of the APD is assumed as 10 and the average number of photons per bit is fixed to 1000 photons/bit. The Johnson noise is selected so that $\sigma_J = 500$noise counts/bit.

Finally, Fig. 10 shows the receiver sensitivity, defined as the minimum n_0 needed to achieve a BER of 10^{-9}. As shown in the figure, the entire transmission range can be split into two parts: a Johnson-noise-limited regime at low transmission speeds and an ISI-limited regime at high transmission speeds. The adverse effect of the correlation between the gain and the buildup time is observed in the ISI-limited regime through the higher sensitivity predicted by the RD-R model compared to that obtained for the DD-E model. For instance, we observe that, in comparison to the predictions of the RD-R model, the DD-E model results in underestimating the sensitivity by 3.4% at a transmission speed of 10 Gb/s, by 7.2% at 15 Gb/s, and by 26% at 20 Gb/s. Thus, the effect of neglecting the statistical correlation between build time and the gain in the receiver-performance analysis leads to an underassessment of ISI, which, in turn, leads to an overly optimistic assessment of the receiver performance. Moreover, this trend intensifies progressively as transmission rate increases.
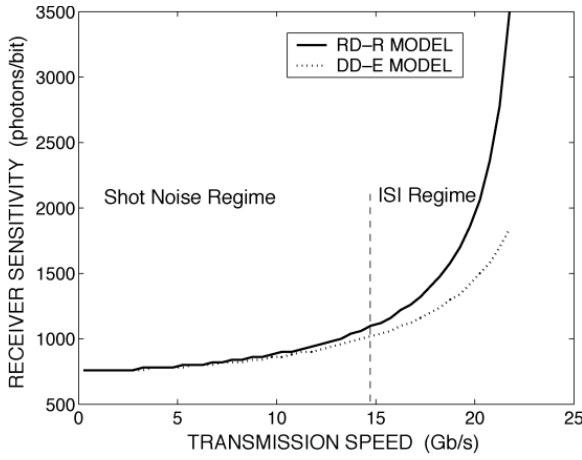
Fig. 10. Receiver sensitivity as a function of the transmission speed using the RD-R model (solid curve) and the DD-E model (dashed curve). The parameters of the APD and Johnson noise are the same as in Fig. 9.

SECTION VI. Conclusion

The statistical correlation between the random gain and the random avalanche buildup time in APDs is determined for the first time, and the effect of the correlation on the receiver performance is established. We have shown that, as a result of this correlation, the photocurrent SNR is diminished by a factor $r = \frac{B_{\text{sneq}}}{B_{\text{3-dB}}}$, where $B_{\text{3-dB}}$ is the APD's 3-dB bandwidth and B_{sneq} is the shot-noise-equivalent bandwidth, which takes into account the statistical correlation between the gain and the buildup time and defined by $B_{\text{sneq}} = \frac{\left\langle \frac{G^2}{T} \right\rangle}{2\langle G^2 \rangle}$. Thus, B_{sneq} is precisely the bandwidth that needs to be used in calculating the shot-noise variance when photodetection is achieved by means of an APD. For example, our calculations show that $\frac{B_{\text{sneq}}}{B_{\text{3-dB}}}$ is approximately 1.3 for a GaAs APD with a multiplication region in the range 30–250 nm.

Additionally, we have derived compact expressions for the output of an integrate-and-dump receiver in an OOK direct-detection system that includes the effects of ISI and the statistical correlation between the gain and the buildup time. This expression is similar to its counterpart when the gain–buildup-time correlation is ignored, but it includes a correction factor, $\kappa = \frac{4B_{\text{sneq}}}{2\pi B_{\text{3-dB}}}$, that explicitly quantifies the SNR degradation due to this correlation. Notably, our results predict that this correlation adversely affects ISI noise, and hence receiver sensitivity. Moreover, such an adverse effect becomes progressively more significant as the transmission rate increases.

The analysis in this paper is carried out under the simplifying assumptions of a constant electric field in the APD's multiplication region and a constant drift velocity for carriers. The analysis, however, includes the effects of dead space and utilizes nonlocalized ionization coefficients. Generalization to nonuniform fields, variable carrier drift speed, and heterostructure APDs can be carried out in a straightforward fashion using the techniques recently reported in [20] and [29].

Appendix Autocorrelation Function of the Rd Parametric Model

For the RD-R model, the autocorrelation function can be expressed as

$$R_{I_{RD}}(\mu, \nu)$$

$$= \left\langle \frac{G}{T}\{u(\mu) - u(\mu - T)\} \frac{G}{T}\{u(\nu) - u(\nu - T)\} \right\rangle$$

$$= \sum_{m=1}^{\infty} \int_0^{\infty} \frac{m^2}{\tau^2}\{u(\mu) - u(\mu - \tau)\}$$

$$\times \{u(\nu) - u(\nu - \tau)\} \frac{\partial}{\partial \tau} f_{G,T}(m, \tau) d\tau.$$

Note that the integrand is zero unless $\tau \geq \mu \vee \nu$ (the notation $\mu \vee \nu$ denotes the maximum of μ and ν). Therefore, the above integral can be rewritten compactly as

$$R_{I_{RD}}(\mu, \nu) = \sum_{m=1}^{\infty} \int_{\mu \vee \nu}^{\infty} \frac{m^2}{\tau^2} \frac{\partial}{\partial \tau} f_{G,T}(m, \tau) d\tau. \quad (41)$$

In particular, the mean of $I_{RD}^2(t)$ is given by

$$\left\langle I_{RD}^2(t) \right\rangle = R_{I_{RD}}(t, t). \quad (42)$$

We now recall the approximation $\left\langle I_{RD}^2(t) \right\rangle \approx c e^{-b_{se} t}$ and compare (41) and (42), and we conclude that $R_{I_{RD}}(\mu, \nu)$ can be approximated by

$$R_{I_{RD}}(\mu, \nu) \approx c e^{-b_{se}(\mu \vee \nu)}. \quad (43)$$

Statistics of Receiver Output in the Presence of Isi

We assume that 0 and 1 occur with equal probability in the transmitted binary sequence $I_b(n)$, and for convenience, we extend $I_b(n)$ to $I_b(t)$ (where t is a continuous variable) by

$$I_b(t) = I_b(n), \text{if} n T_b \leq t < (n+1) T_b$$

and the (absorbed) optical signal at the receiver can be represented by $\phi(t) = \phi_1 I_b(t)$. Note that the covariance function and the mean of $I_b(t)$ are

$$\sigma_{I_b}^2(\xi_1, \xi_2) = \begin{cases} \frac{1}{4}, & \text{if } n T_b \leq \xi_1, \xi_2 < (n+1) T_b, \\ & \text{for } n = -1, -2, -3, \dots \\ 0, & \text{otherwise} \end{cases}$$

$$\langle I_b(t) \rangle = \begin{cases} \frac{1}{2}, & \text{if } t \leq 0 \\ 0, & \text{if } t > 0 \text{ and current bit is } 0 \\ 1, & \text{if } t > 0 \text{ and current bit is } 1. \end{cases}$$

The mean value of the receiver output can be calculated by

$$\langle \Gamma \rangle = \phi_1 \int_0^{T_b} \int_{-\infty}^{t} \langle I_p(t - \xi)\rangle \langle I_b(\xi)\rangle d\xi dt.$$

If the current bit being sent is zero, then $\langle I_b(\xi)\rangle = 0$ for $0 \le \xi < T_b$, and

$$\mu_0 = \frac{\phi_1}{2} \int_0^{T_b} \int_{-\infty}^{0} \langle I_p(t - \xi)\rangle \, d\xi dt. \tag{44}$$

On the other hand, if the current bit being sent is one, then $\langle I_b(\xi)\rangle = 1$ for $0 \le \xi < T_b$, and

$$\begin{aligned}
\mu_1 &= \frac{\phi_1}{2} \int_0^{T_b} \int_{-\infty}^{0} \langle I_p(t - \xi)\rangle \, d\xi dt + \phi_1 \int_0^{T_b} \int_0^{t} \langle I_p(t - \xi)\rangle \, d\xi dt \\
&= \mu_0 + \phi_1 \int_0^{T_b} \int_0^{t} \langle I_p(t - \xi)\rangle \, d\xi dt.
\end{aligned} \tag{45}$$

To calculate the variance of Γ, we will utilize standard analysis of filtered shot-noise processes [26] while considering the stochastic nature of the transmitted signal $I_b(t)$. This yields

$$\begin{aligned}
\langle \Gamma^2 \rangle &= \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{\mu} \int_{\substack{-\infty \\ \xi_1 \ne \xi_2}}^{\nu} \langle I_p(\mu - \xi_1)\phi_1(\xi_1)I_p(\nu - \xi_2)\phi_1(\xi_2)\rangle \\
&\quad \times d\xi_1 d\xi_2 d\mu d\nu \\
&\quad + \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{\mu \wedge \nu} \langle I_p(\mu - \xi)I_p(\nu - \xi)\phi_1(\xi)\rangle \, d\xi d\mu d\nu \\
&= \phi_1^2 \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{\mu} \int_{\substack{-\infty \\ \xi_1 \ne \xi_2}}^{\nu} \langle I_p(\mu - \xi_1)\rangle \langle I_p(\nu - \xi_2)\rangle \\
&\quad \times \langle I_b(\xi_1)I_b(\xi_2)\rangle d\xi_1 d\xi_2 d\mu d\nu \\
&\quad + \phi_1 \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{\mu \wedge \nu} \langle I_p(\mu - \xi)I_p(\nu - \xi)\rangle \langle I_b(\xi)\rangle d\xi d\mu d\nu.
\end{aligned}$$

Since $\langle I_b(\xi_1)I_b(\xi_2)\rangle = \langle I_b(\xi_1)\rangle\langle I_b(\xi_2)\rangle + \sigma_{I_b}^2(\xi_1, \xi_2)$, and since the set $\{(\xi_1, \xi_2): \xi_1 \ne \xi_2\}$ has zero area, we obtain

$$\begin{aligned}
\sigma_\Gamma^2 &= \phi_1^2 \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{\mu} \int_{-\infty}^{\nu} \langle I_p(\mu - \xi_1)\rangle \langle I_p(\nu - \xi_2)\rangle \\
&\quad \times \sigma_{I_b}(\xi_1, \xi_2) d\xi_1 d\xi_2 d\mu d\nu \\
&\quad + \phi_1 \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{\mu \wedge \nu} R_{I_p}(\mu - \xi, \nu - \xi)\langle I_b(\xi)\rangle d\xi d\mu d\nu.
\end{aligned} \tag{46}$$

We first consider the case for which the current bit is zero; in this case

$$\begin{aligned}
\sigma_0^2 &= \frac{\phi_1^2}{4} \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{0} \int_{\substack{-\infty \\ (\xi_1, \xi_2) \in \mathbb{D}}}^{0} \langle I_p(\mu - \xi_1)\rangle\langle I_p(\nu - \xi_2)\rangle d\xi_1 d\xi_2 d\mu d\nu \\
&\quad + \frac{\phi_1}{2} \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^{0} R_{I_p}(\mu - \xi, \nu - \xi) d\xi d\mu d\nu
\end{aligned} \tag{47}$$

where $\mathbb{D} = \{(\xi_1, \xi_2): nT_b \le \xi_1, \xi_2 < (n + 1)T_b, n = -1, -2, -3, \dots\}$.

Alternatively, if the current bit is one, then $\sigma_{I_b}^2(\xi_1, \xi_2) = 0$ for $0 \le \xi_1, \xi_2 < T_b$, and we obtain

$$
\begin{aligned}
\sigma_1^2 =\ & \frac{\phi_1^2}{4} \int_0^{T_b} \int_0^{T_b} \underset{(\xi_1,\xi_2)\in\mathbb{D}}{\int_{-\infty}^0 \int_{-\infty}^0} \langle I_p(\mu - \xi_1)\rangle\langle I_p(\nu - \xi_2)\rangle d\xi_1 d\xi_2 d\mu d\nu \\
& + \frac{\phi_1}{2} \int_0^{T_b} \int_0^{T_b} \int_{-\infty}^0 R_{I_p}(\mu - \xi, \nu - \xi) d\xi d\mu d\nu \\
& + \phi_1 \int_0^{T_b} \int_0^{T_b} \int_0^{\mu \wedge \nu} R_{I_p}(\mu - \xi, \nu - \xi) d\xi d\mu d\nu \\
=\ & \sigma_0^2 + \phi_1 \int_0^{T_b} \int_0^{T_b} \int_0^{\mu \wedge \nu} R_{I_p}(\mu - \xi, \nu - \xi) d\xi d\mu d\nu.
\end{aligned}
\tag{48}
$$

We next specialize these results to the DD-E model and RD-R models. For the DD-E model, we substitute the impulse-response function $I_{DD}(t) = qGbe^{-bt}$ and the autocorrelation function $R_{I_{DD}}(\mu, \nu) = \langle G^2\rangle b^2 e^{-b(\mu+\nu)}$ into (44), (45), (47), and (48), use $b = 2\pi B_{3\text{-dB}}$, and define $\lambda = 2\pi T_b B_{3\text{-dB}}$ to obtain

$$
\begin{aligned}
\mu_{0,DD} =\ & \frac{1}{2} \frac{n_0\langle G\rangle}{\lambda} \left(1 - e^{-\lambda}\right) \\
\sigma_{0,DD}^2 =\ & \frac{1}{4} \frac{n_0^2\langle G\rangle^2}{\lambda^2} \frac{(1-e^{-\lambda})^4}{(1-e^{-2\lambda})} \\
& + \frac{n_0\langle G^2\rangle}{\lambda}(1 - e^{-\lambda})^2 + \sigma_J^2 \\
\mu_{1,DD} =\ & \mu_{0,DD} + \frac{n_0\langle G\rangle}{\lambda}\left(\lambda - 1 + e^{-\lambda}\right) \\
\sigma_{1,DD}^2 =\ & \sigma_{0,DD}^2 + \frac{n_0\langle G^2\rangle}{2\lambda}\left(2\lambda - (1 - e^{-\lambda})^2\right)
\end{aligned}
\tag{49)(50)(51)(52}
$$

where n_0 is the average number of absorbed photons per "1" bit $\left(\phi_1 = \frac{n_0}{T_b}\right)$ and σ_J is the Johnson-noise parameter. The general expressions of SNR and ISI for this RD-exponential model are given by

$$
\begin{aligned}
SNR_{DD} =\ & \frac{(\mu_{1,DD} - \mu_{0,DD})^2}{\sigma_{1,DD}^2} \\
ISI_{DD} =\ & \sigma_{0,DD}^2 - \sigma_J^2.
\end{aligned}
\tag{53)(54}
$$

For the RD-R model, in particular, we substitute the approximations (21), (43), $\frac{a}{b_{se}} \approx \langle G\rangle$, and $\frac{c}{b_{se}} \approx \left\langle \frac{G^2}{T}\right\rangle$ into (44), (45), (47), and (48), use $b_{se} = 4B_{sneq}$, $\kappa = \frac{4B_{sneq}}{2\pi B_{3\text{-dB}}}$, and $\lambda = 2\pi T_b B_{3\text{-dB}}$, and obtain

$$
\begin{aligned}
\mu_{0,RD} =\ & \frac{1}{2} \frac{n_0\langle G\rangle}{\kappa\lambda} \left(1 - e^{-\kappa\lambda}\right) \\
\sigma_{0,RD}^2 =\ & \frac{1}{4} \frac{n_0^2\langle G\rangle^2}{\kappa\lambda^2} \frac{(1-e^{-\kappa\lambda})^4}{1-e^{-4\pi\kappa T_b B_{3\text{-dB}}}} \\
& + \frac{n_0\langle G^2\rangle}{2\kappa\lambda}\left(1 - e^{-\kappa\lambda} - \kappa\lambda e^{-\kappa\lambda}\right) + \sigma_J^2 \\
\mu_{1,RD} =\ & \mu_{0,RD} + \frac{n_0\langle G\rangle}{\kappa\lambda}\left(\kappa\lambda - 1 + e^{-\kappa\lambda}\right) \\
\sigma_{1,RD}^2 =\ & \sigma_{0,RD}^2 + \frac{n_0\langle G^2\rangle}{\kappa\lambda}\left(\kappa\lambda - 2 + 2e^{-\kappa\lambda} + \kappa\lambda e^{-\kappa\lambda}\right).
\end{aligned}
\tag{55)(56)(57)(58}
$$

Finally, the general expressions of SNR and ISI for this RD-R model are given by

$$\text{SNR}_{\text{RD}} = \frac{(\mu_{1,\text{RD}} - \mu_{0,\text{RD}})^2}{\sigma_{1,\text{RD}}^2} \quad (59)(60)$$

$$\text{ISI}_{\text{DD}} = \sigma_{0,\text{RD}}^2 - \sigma_{\text{J}}^2.$$

References

1. R. J. McIntyre, "Multiplication noise in uniform avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. ED-13, no. 1, pp. 164-168, Jan. 1966.

2. P. Bhattacharya, Semiconductor Optoelectronic Devices, NJ, Upper Saddle River:Prentice-Hall, 1997.

3. B. E. A. Saleh, M. C. Teich, Fundamentals of Photonics, New York:Wiley, 1991.

4. M. C. Teich, K. Matsuo, B. E. A. Saleh, "Excess noise factors for conventional and superlattice avalanche photodiodes and photomultiplier tubes", *IEEE J. Quantum Electron.*, vol. QE-22, no. 8, pp. 1184-1193, Aug. 1986.

5. C. Hu, K. A. Anselm, B. G. Streetman, J. C. Campbell, "Noise characteristics of thin multiplication region GaAs avalanche photodiodes", *Appl. Phys. Lett.*, vol. 69, no. 24, pp. 3734-3736, Dec. 1996.

6. K. A. Anselm, P. Yuan, C. Hu, C. Lenox, H. Nie, G. Kinsey, J. C. Campbell, B. G. Streetman, "Characteristics of GaAs and AlGaAs homojunction avalanche photodiodes with thin multiplication regions", *Appl. Phys. Lett.*, vol. 71, no. 26, pp. 3883-3885, Dec. 1997.

7. C. Lenox, P. Yuan, H. Nie, O. Baklenov, C. Hansing, J. C. Campbell, J. A. L. Holmes, B. G. Streetman, "Thin multiplication region InAlAs homojunction avalanche photodiodes", *Appl. Phys. Lett.*, vol. 73, no. 6, pp. 783-784, Aug. 1998.

8. 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 $p^{+} - i - n^{+}$ diodes ", *IEEE Trans. Electron Devices*, vol. 45, no. 10, pp. 2102-2107, Oct. 1998.

9. K. F. Li, D. S. Ong, J. P. R. David, P. N. Robson, R. C. Tozer, G. J. Rees, R. Grey, "Low excess noise characteristics in thin avalanche region GaAs diodes", *Electron. Lett.*, vol. 34, no. 1, pp. 125-126, Jan. 1998.

10. D. S. Ong, K. F. Li, G. J. Rees, G. M. Dunn, J. P. R. David, P. N. Robson, " A Monte Carlo investigation of multiplication noise in thin $p^{+} - i - n^{+}$ GaAs avalanche photodiodes ", *IEEE Trans. Electron Devices*, vol. 45, no. 8, pp. 1804-1810, Aug. 1998.

11. D. S. Ong, K. F. Li, G. J. Rees, J. P. R. David, P. N. Robson, " Monte Carlo estimation of multiplication noise in thin $p^{+} - i - n^{+}$ GaAs diodes ", *Appl. Phys. Lett.*, vol. 72, no. 2, pp. 232-234, Jan. 1998.

12. S. A. Plimmer, J. P. R. David, R. Grey, G. J. Rees, " Avalanche multiplication in $\hbox{Al}_{x}\hbox{Ga}_{1-x}\hbox{As}$ ($x = 0$ to 0.60) ", *IEEE Trans. Electron Devices*, vol. 47, no. 5, pp. 1089-1097, May 2000.

13. K. F. Li, D. S. Ong, J. P. R. David, R. C. Tozer, G. J. Rees, S. A. Plimmer, K. Y. Chang, J. S. Roberts, "Avalanche noise characteristics of thin GaAs structures with distributed carrier generation", *IEEE Trans. Electron Devices*, vol. 47, no. 5, pp. 910-914, May 2000.

14. C. H. Tan, J. C. Clark, J. R. P. David, G. J. Rees, S. A. Plimmer, R. C. Tozer, " Avalanche noise measurement in thin Si $p^{+} - i - n^{+}$ diodes ", *Appl. Phys. Lett.*, vol. 76, no. 26, pp. 3926-3928, Jun. 2000.

15. M. M. Hayat, B. E. A. Saleh, "Statistical properties of the impulse response function of double-carrier multiplication avalanche photodiodes including the effect of dead space", *J. Lightw. Technol.*, vol. 10, no. 10, pp. 1415-1425, Oct. 1992.

16. M. M. Hayat, O. Kwon, Y. Pan, P. Sotirelis, J. C. Campbell, B. E. A. Saleh, M. C. Teich, "Gain-bandwidth characteristics of thin avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 49, no. 5, pp. 770-781, May 2002.

17. M. M. Hayat, B. E. A. Saleh, M. C. Teich, "Effect of dead space on gain and noise of double-carrier-multiplication avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 546-552, Mar. 1992.

18. X. Li, X. Zheng, S. Wang, F. Ma, J. C. Campbell, " Calculation of gain and noise with dead space for GaAs and \$hbox{Al}_{x}hbox{Ga}_{1-x}hbox{As}\$ avalanche photodiode ", *IEEE Trans. Electron Devices*, vol. 49, no. 7, pp. 1112-1117, Jul. 2002.

19. M. M. Hayat, G. Dong, "A new approach for computing the bandwidth statistics of avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 47, no. 6, pp. 1273-1279, Jun. 2000.

20. M. M. Hayat, O. Kwon, S. Wang, J. C. Campbell, B. E. A. Saleh, 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.

21. N. R. Das, M. J. Deen, "A new model for avalanche build-up of carriers in a SAGCM avalanche photodiode", *IEEE Trans. Electron Devices*, vol. 49, no. 12, pp. 2362-2366, Dec. 2002.

22. O.-H. Kwon, M. M. Hayat, J. C. Campbell, B. E. A. Saleh, M. C. Teich, "Effect of stochastic dead space on noise in avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 51, no. 5, pp. 693-700, May 2004.

23. J. A. Gubner, M. M. Hayat, "A method to recover counting distributions from their characteristic functions", *IEEE Signal Process. Lett.*, vol. 3, no. 6, pp. 184-186, Jun. 1996.

24. M. A. Saleh, M. M. Hayat, P. 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, no. 12, pp. 2722-2731, Dec. 2001.

25. S. Hava, M. Auslender, "Velocity-field relation in GaAlAs versus alloy composition", *J. Appl. Phys.*, vol. 73, pp. 7431-7434, Jun. 1993.

26. D. Snyder, M. Miller, Random Point Processes in Time and Space, New York:Springer-Verlag, 1991.

27. M. M. Hayat, B. E. A. Saleh, J. A. Gubner, "Bit-error rates for optical receivers using avalanche photodiodes with dead space", *IEEE Trans. Comm.*, vol. 43, no. 1, pp. 99-107, Jan. 1995.

28. G. Kahraman, B. E. A. Saleh, W. L. Sargeant, M. C. Teich, "Time and frequency response of avalanche photodiodes with arbitrary structure", *IEEE Trans. Electron Devices*, vol. 39, no. 3, pp. 553-560, Mar. 1992.

29. 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.