2002

Gain-bandwidth characteristics of thin avalanche photodiodes

Majeed M. Hayat
Oh-Hyun Kwon
Yi Pan
Paul P. Sotirelis
Joe C. Campbell

See next page for additional authors

Follow this and additional works at: https://epublications.marquette.edu/electric_fac

Part of the Computer Engineering Commons, and the Electrical and Computer Engineering Commons
Authors
Majeed M. Hayat, Oh-Hyun Kwon, Yi Pan, Paul P. Sotirelis, Joe C. Campbell, Bahaa E.A. Saleh, and Malvin Carl Teich
Gain-bandwidth characteristics of thin avalanche photodiodes

Majeed M. Hayat
Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87131-1356 USA

Oh-Hyun Kwon
Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87131-1356 USA

Yi Pan
Department of Computer Science, Georgia State University, Atlanta, GA 30303 USA

Paul Sotirelis
National Center for Supercomputer Applications, University of Illinois, Urbana, Urbana, IL, 61801 USA

Joe C. Campbell
Department of Electrical and Computer Engineering, The University of Texas, Austin, TX 78712 USA
Bahaa E. A. Saleh  
Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215-2421 USA  
Malvin Carl Teich  
Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215-2421 USA

Abstract

The frequency-response characteristics of avalanche photodiodes (APDs) with thin multiplication layers are investigated by means of a recurrence technique that incorporates the history dependence of ionization coefficients. In addition, to characterize the autocorrelation function of the impulse response, new recurrence equations are derived and solved using a parallel computer. The mean frequency response and the gain-bandwidth product are computed and a simple model for the dependence of the gain-bandwidth product on the multiplication-layer width is set forth for GaAs, InP, Al\textsubscript{0.2}/Ga\textsubscript{0.8}/As, and In\textsubscript{0.52}/Al\textsubscript{0.48}/As APDs. It is shown that the dead-space effect leads to a reduction (up to 30\%) in the bandwidth from that predicted by the conventional multiplication theory. Notably, calculation of the power-spectral density of the photocurrent reveals that the presence of dead space also results in a reduction in the fluctuations in the frequency response. This result is the spectral generalization of the reduction in the excess noise factor in thin APDs and reveals an added advantage of using thin APDs in ultrafast receivers.

SECTION I.  
Introduction

With the increasing demand for high-speed optical communication, the need for ultrafast and low-noise photo detectors has become greater than ever. Among the semiconductor photo detectors that are commonly used in today’s long-haul and metro-area fiber-optic systems, avalanche photodiodes (APDs) are often preferred over p-i-n photodiodes by virtue of their internal gain, which significantly improves receiver sensitivity and alleviates the need for optical preamplification. Indeed, recent advances in the design and fabrication of APDs (see for example [1]–[2][3]) have allowed these devices to achieve levels of gain-bandwidth product that have made them the photodetectors of choice in many of the current 10-Gbps systems that operate in the silica-fiber window.

Unfortunately, the very process of carrier impact ionization that produces the gain is inherently noisy and results in fluctuations not only in the gain but also in the time response. In particular, the APDs impulse response, which is the response to a single photo excitation event, is a stochastic process with random shape (whose total area is proportional to the random gain) and random duration [4]. Generally, the duration of the impulse response increases with increasing gain. This excess time in the response is often referred to as the avalanche buildup time. While gain uncertainty plays a major role in the performance of power-limited APD-based receivers, it is the avalanche buildup time that gives rise to intersymbol interference (ISI) that limits receiver performance in high-speed systems.
It has been demonstrated that the excess noise factor (which is a measure of the gain fluctuations) and the avalanche buildup time can both be reduced by using thin multiplication layers [1]–[2][3], [5]–[6][7][8][9][10][11][12][13][14][15][16][17]. The reduction of the avalanche buildup time in thin APDs is primarily due to the reduction in the carriers’ transit time across the thin multiplication layer. In contrast, the reduction of the excess noise factor is now known to be primarily due to the effect of a carrier’s past-history on its ability to create a new carrier pair via impact ionization. Accounting for carrier history is important because newly born carriers are incapable of immediately causing impact ionizations: they must first travel a sufficient distance (called the dead space), in the course of which they gain enough energy from the field to permit them to cause an impact ionization. The conventional avalanche multiplication model, first developed by McIntyre [18], does not account for the dead-space effect and does not predict a reduction of the excess noise factor for thin APDs. The effect of dead space on the gain and excess noise factor has been extensively studied and multiplication models that take carrier history into account have been developed and tested against experimental measurements [7]–[8][9][10][11][12][13][14], [19]–[20][21][22][23][24][25][26].

Just as accounting for dead space is essential for correctly predicting the excess noise factor in thin APDs, accurately predicting the bandwidth characteristics of thin APDs necessitates having a time-response analysis of the avalanche multiplication that includes the effect of dead space. The analytical model for the statistics of the impulse response for the history-dependent multiplication model was first developed by Hayat and Saleh [27] and was also recently revisited by the authors and by other groups as well [26], [28]–[29][30]. It has been shown in [27], [30] that dead space tends to elongate the duration of the impulse response when compared with predictions of conventional models for the time response that do not account for dead space [31]–[32][33][34][35][36][37][38][39][40]. This effect therefore tends to reduce the expected transit-time-induced improvement in bandwidth as the multiplication region thickness is reduced. Moreover, knowing that the reduction in the multiplication-layer width is responsible for a reduction in the gain fluctuations (i.e., fluctuations in the area under the impulse response) raises the natural question of whether, more generally, the spectral fluctuations in the photo current are also reduced as a result of reducing the width. If the answer is in the affirmative, this would benefit the receiver bit-error rate (by reducing ISI) in the same way that the reduction in the excess noise factor improves the receiver SNR. Accurate modeling of the photo current and its fluctuations can provide the means for better estimating the contribution of ISI and the bit-error rate in high-speed systems. Such a model can therefore play a key role in device optimization in ultrafast receivers.

To date, no theory characterizing the autocorrelation function (or the power spectral density) of APDs has been developed that incorporates the dead-space effect. In this paper, we extend the time-domain analysis of the dead-space multiplication model reported in [27] to compute the autocorrelation function of the APD impulse response under the assumption of a constant electric field. This extension involves developing six recurrence equations, which are derived according to the same renewal-theory rationale used in [27]. Application of the model to actual devices requires knowledge of the ionization coefficients of enabled carries that have traveled the dead space. These material-specific ionization coefficients, which are independent of the multiplication-layer width, have been reported by the authors for GaAs, Al0.2Ga0.8As, InP, and In0.52Al0.48As APDs [8], [9]. In this paper, we use the width-independent model for the ionization coefficients in conjunction with the theory developed here and in [27] to determine the effect of reducing the width of the multiplication region on the gain-bandwidth-product and the photo current spectral fluctuations. Since the excess noise factor is a measure of fluctuation of the dc component of the APDs frequency response, the results of this paper on the photocurrent spectral fluctuations in thin APDs are a generalization of the excess-noise-factor reduction in thin APDs to all operational frequencies.
SECTION II.

History-Dependent Impact Ionization Model

We begin by recalling germane aspects of the dead-space ionization model used in this paper. According to the dead-space multiplication model [8], [10], [22], the probability density function (pdf) of the electron free-path distance \( X_e \), which is the distance from the carrier's creation to the point where it impact ionizes, can be modeled by

\[
h_e(x) = \alpha e^{-\alpha(x-d_e)}u(x-d_e)
\]

where \( d_e \) is the electron dead space, and \( u(x) \) is the unit step function. A similar expression exists for the hole. Assuming the absence of phonon scattering, the electron and hole dead spaces are computed from

\[
d_e = \frac{E_{ie}}{qE} \quad \text{and} \quad d_h = \frac{E_{ih}}{qE},
\]

where \( E_{ie} \) and \( E_{ih} \) are the ionization threshold energies of the electron and hole, respectively, \( q \) is the electron charge, and \( E \) is the applied electric field in the multiplication layer. A model for the electron and hole impact ionization coefficients of enabled carriers has been recently developed by Saleh et al. [8], [10]. For the electrons, the model is given by

\[
\alpha(E) = \Lambda \exp \left[ -\left( \frac{E_C}{E} \right)^m \right].
\]

A similar formula exists for the holes. This model has been shown to correctly predict the excess noise factors independently of the width of the multiplication layer [8], [10]. The width-independent parameters \( \Lambda, E_C, \) and \( m \) for the electron and the hole are given in Table I of [8] and the threshold energies are also given in Table IV of [8]. Later in Section IV, we will compare the results of the dead-space model with the predictions of the conventional multiplication theory, for which the dead space is ignored and the ionization coefficients are those for bulk material, as reported by Bulman et al. [41] and Cook et al. [42] for GaAs and InP, respectively.

In the theory developed in this paper, the form of the pdf of the carrier free-path distance is arbitrary, although we make use of the hard-threshold dead-space model [given by (1)] in our calculations. More realistic soft-threshold ionization models for which the newly created carriers gradually attain ionization capability can also be incorporated by simply modifying the forms of the pdf's of the free-path distance. The specific forms for soft-threshold dead-space ionization coefficients (e.g., \( \alpha(x) \), where \( x \) is the distance from birth location) are typically obtained by means of Monte-Carlo simulation and their use has recently been shown to improve the excess-noise predictions [43], [44]. In general, the pdf of the free-path distance can be obtained from the soft dead-space ionization rate using the simple formula

\[
h_e(x) = \alpha(x)\exp \{- \int_0^x \alpha(y)dy\}.\]

If, for example, \( \alpha(x) \) is a step function, as is the case for a hard dead-space ionization coefficient, then the pdf given by (1) is generated. Unfortunately, however, Monte-Carlo-based calculations of the soft dead-space ionization coefficient have been carried out only for a few materials and are not currently available for all four materials considered in this paper. Nevertheless, it has been shown that with appropriate fine-tuning of the ionization threshold energies, the simple hard dead-space approximation can yield excellent agreement with experimental excess noise factors for thin APDs [8].
Finally, in our formulation of the dead-space model, we adopted the commonly accepted assumption that the dead space is deterministic. In actuality, the dead space should be modeled as a random variable to accommodate the fact that a carrier does not necessarily lose all of its kinetic energy after each impact ionization. This can be incorporated into the pdf of the carrier’s path but requires knowledge of the probability distribution of the dead space, which is not fully characterized in the current state of our knowledge. In general, we expect that such stochastic dead space will not have a significant effect on mean quantities (e.g., the mean gain and bandwidth) since its presence tends to be averaged out. This effect should play a more prominent role in the second-order statistics (e.g., excess noise factor and the power spectral density), since the dead-space randomness will tend to add to the overall uncertainty, in accordance with basic statistical principles.

SECTION III.
Statistics of the Impulse Response

Consider an electron-injected APD with a multiplication region of width $w$, with the uniform electric field pointing from $x=w$ to $x=0$. Let $Z_e(t,x)$ be the total number of electrons resulting from an initial parent electron born at location $x$, at $t$ units of time after its birth. Similarly, let $Z_h(t,x)$ be the total number of holes resulting from an initial parent electron, at location $x$ at $t$ units of time after its birth. To see how these quantities can be used in representing the impulse response, consider the case when a photo-generated electron is injected into the edge of the multiplication region (at $x=0$) at time $t=0$. The value of the buildup-time-limited impulse response $I(t)$ can be obtained by adding up the current contributions from all the offspring electrons and holes that are traveling in the multiplication region at time $t$. More precisely, if $v_e$ and $v_h$ are the saturation velocities of the electron and hole, respectively, then Ramo’s Theorem gives

$$I(t) = \frac{q}{w} [v_e Z_e(t,0) + v_h Z_h(t,0)].$$

The statistics of $I(t)$ can therefore be readily calculated from the joint statistics of $Z_e(t,0)$ and $Z_h(t,0)$, which will be discussed next.

A. Rationale for Recurrence Relations

We begin by recalling germane aspects of the recurrence technique developed in [27]. It turns out that it is necessary to first characterize the statistics of $Z_e(t,x)$ and $Z_h(t,x)$ for all $x$ and then specialize the results to $x=0$. To do so, we also need to introduce additional quantities representing cases when a hole initiates the multiplication. In particular, let $Y_e(t,x)$ be the total number of electrons resulting from an parent hole born at location $x$, at $t$ units of time after its birth, and let $Y_h(t,x)$ be defined similarly to $Y_e(t,x)$ but with the number of generated electrons replaced with the number of generated holes.

The key idea that enables us to recursively characterize the statistics of the aforementioned random functions is based on a renewal argument that can be explained as follows. Consider the quantity $Z_e(t,x)$ and the parent electron associated with it (born at position $x$) that initiates the multiplication process. Suppose for the moment that this parent electron first impact ionizes at a certain location, say $\xi$, where $x \leq \xi \leq w$. In such an event, there will be two newly created electrons and a hole at location $\xi$. The two
electrons and the hole will then independently induce an avalanche process at the new location. Let \( \Delta = (\xi - x)/v_e \) be the transit time of an electron from \( x \) to \( \xi \). Hence, under the earlier scenario, \( Z_e(t,x) \) is precisely the sum of the electrons that each of the two offspring electrons and the offspring hole would generate at precisely \( t-\Delta \) units of time after their births. Informally, we can express this observation as \( Z_e(t,x|\xi) = 2Z_e(t-\Delta,\xi) + Y_e(t-\Delta,\xi) \), where the symbol “|\( \xi \)” is used to denote “conditional on \( \xi \).”

“We call the previous relation a conditional renewal relation. Similarly, we can examine \( Y_e(t,x) \) (where we track a parent hole in this case) and obtain \( Y_e(t,x|\xi) = 2Y_e(t-(x/v_h),\xi) + Z_e(t-(x/v_h),\xi) \), where in this case \( 0 \leq \xi \leq x \). Similar conditional renewal expressions can be obtained for \( Z_h(t,x) \) and \( Y_h(t,x) \); these are:

\[
Z_h(t,x|\xi) = 2Z_h(t-\Delta,\xi) + Y_h(t-\Delta,\xi) \quad \text{and} \quad Y_h(t,x|\xi) = 2Y_h(t-(x/v_h),\xi) + Z_h(t-(x/v_h),\xi).
\]

In the next section, we will use the previously mentioned conditional renewal relations to derive integral (recurrence) equations for autocorrelations and crosscorrelations of the quantities \( Z_e(t,x), Y_e(t,x), Z_h(t,x), \) and \( Y_h(t,x) \), which will be used, in turn, to characterize the autocorrelation function of the impulse response \( i(t) \). Recurrence equations for the mean quantities \( z_e(t,x), zh(t,x), ye(t,x), \) and \( y_h(t,x) \), have been previously derived in [27] \((6), (10)-(12)\). When these mean quantities are computed, the mean impulse response function \( \bar{i}(t) = \frac{\Delta}{E[I(t)]} \) can be determined as

\[
i(t) = \frac{q}{w} [v_e z_e(t,0) + v_h z_h(t,0)].
\]

B. Autocorrelation Function of the Impulse Response

Let \( R(t_1,t_2) \) = \( E[I(t_1)I(t_2)] \) be the autocorrelation function of the impulse response. By using \([3]\), expanding terms, and taking averages, we obtain

\[
R(t_1,t_2) = E\left[ \left( \frac{q}{w} \right)^2 \{ v_e Z_e(t_1,0) + v_h Z_h(t_1,0) \} \left( \frac{q}{w} \right)^2 \{ v_e Z_e(t_2,0) + v_h Z_h(t_2,0) \} \right] \\
= \left( \frac{q}{w} \right)^2 \{ v_e^2 C_{Z_e}(t_1,t_2,0) + v_h^2 C_{Z_h}(t_1,t_2,0) + 2v_e v_h C_{Z_e Z_h}(t_1,t_2,0) \}
\]

where the count autocorrelations are defined as follows: \( C_{Z_e}(t_1,t_2,x) = E[Z_e(t_1,x)Z_e(t_2,x)] \) and \( C_{Z_h}(t_1,t_2,x) = E[Z_h(t_1,x)Z_h(t_2,x)] \) and the count crosscorrelation is defined by \( C_{Z_e Z_h}(t_1,t_2,x) = E[Z_e(t_1,x)Z_h(t_2,x)]. \) Hence, \( R(t_1,t_2) \) can be readily computed once the above autocorrelations and crosscorrelations are characterized. However, characterizing the above correlations will involve additional correlations involving the variables \( Y_e \) and \( Y_h \). These are:

\[
C_{Y_e}(t_1,t_2,x) = \]
We will now develop a pair of coupled recurrence relations for the autocorrelations $C_{Ze}(t_1,t_2,x)$ and $C_{Ye}(t_1,t_2,x)$. Other count autocorrelations and crosscorrelations can be addressed similarly. Following the discussion in Section III-A, we will condition on the location of the first ionization and examine the events that transpire thereafter. We also observe that for any correlation times $t_1 \leq t_2$, there are three possibilities for the time of the first impact-ionization event. It could occur before $t_1$, between $t_1$ and $t_2$ or after $t_2$. In each case, we analyze the outcome events and use appropriate conditional renewal expressions from Section III-A to obtain a conditional renewal expression for $C_{Ze}(t_1,t_2,x)$.

\[ C_{Ze}(t_1,t_2,x) = \begin{cases} \end{cases} \]

\[ C_{Ye}(t_1,t_2,x) = \begin{cases} \end{cases} \]

In the Appendix, we average over all possible locations $\xi$ for the first impact ionization [using the pdf (1)], while carefully considering the previously mentioned three possibilities for the time of the first electron ionization and derive the following recurrence relation: For $t_1 \leq t_2$, see (6), shown at the bottom of the page, where $\Delta_1 = (s - x)/v_e$ and where

\[ H_e(x) = \int_{-\infty}^{x} h_e(x')dx' \]

is the cumulative distribution function of the electron life-span random variable $X_e$.

\[ C_{Ye}(t_1,t_2,x) = \begin{cases} \end{cases} \]

\[ A \text{ similar analysis can be carried out to derive a recurrence relation for } C_{Ye}(t_1,t_2,x): \text{ For } t_1 \leq t_2, \text{ see (8), shown at the bottom of the page, where } \Delta_2 = (x - s)/v_h \text{ and} \]
We now state, without proof, the recurrence equations for the remaining count autocorrelations and count
interchange must be computed \textit{a priori} according to [27]. We make the final note that when \( t_1 > t_2 \), we must simply interchange \( t_1 \) and \( t_2 \) in (6) and (8).

\[
C_{Z_h}(t_1, t_2, x) = \int_{\min(x + v_t_2, w)}^{\min(x + v_t_1, w)} \{2z_h(t_2 - \Delta_1, s) + y_h(t_2 - \Delta_1, s)\} h_e(s - x) ds \\
+ \int_x^{\min(x + v_t_1, w)} [2C_{Z_h}(t_1 - \Delta_1, t_2 - \Delta_1, s) + C_{Y_h}(t_1 - \Delta_1, t_2 - \Delta_1, s) \\
+ 2z_h(t_1 - \Delta_1, s) z_h(t_2 - \Delta_1, s) + 2z_h(t_1 - \Delta_1, s) y_h(t_2 - \Delta_1, s) \\
+ 2y_h(t_1 - \Delta_1, s) z_h(t_2 - \Delta_1, s) + y_h(t_1 - \Delta_1, s) y_h(t_2 - \Delta_2, s)] \\
\times h_e(s - x) ds
\]

And

\[
C_{Y_h}(t_1, t_2, x) = u(\left[\frac{x}{v_h}\right] - t_2)[1 - H_h(v_h t_2)] \\
+ \int_{\max(x - v_h t_1, 0)}^{\max(x - v_h t_2, 0)} \{2y_h(t_2 - \Delta_2, s) + z_h(t_2 - \Delta_2, s)\} h_h(x - s) ds \\
+ \int_x^{\max(x - v_t_2, w)} [2C_{Y_h}(t_1 - \Delta_2, t_2 - \Delta_2, s) + C_{Z_h}(t_1 - \Delta_2, t_2 - \Delta_2, s) \\
+ 2y_h(t_1 - \Delta_2, s) y_h(t_2 - \Delta_2, s) + 2z_h(t_1 - \Delta_2, s) y_h(t_2 - \Delta_2, s) \\
+ 2y_h(t_2 - \Delta_1, s) z_h(t_2 - \Delta_2, s) + z_h(t_1 - \Delta_2, s) z_h(t_2 - \Delta_2, s)] \\
\times h_h(x - s) ds
\]

\[
C_{Z_e}(t_1, t_2, x) = \int_{\min(x + v_t_2, w)}^{\min(x + v_t_1, w)} \{2z_e(t_2 - \Delta_1, s) + y_e(t_2 - \Delta_1, s)\} h_e(s - x) ds \\
+ \int_x^{\min(x + v_t_1, w)} [2C_{Z_e}(t_1 - \Delta_1, t_2 - \Delta_1, s) + C_{Y_e}(t_1 - \Delta_1, t_2 - \Delta_1, s) \\
+ 2z_e(t_1 - \Delta_1, s) z_e(t_2 - \Delta_1, s) \{2z_e(t_1 - \Delta_1, s) y_e(t_2 - \Delta_1, s) \\
+ 2y_e(t_1 - \Delta_1, s) z_e(t_2 - \Delta_1, s)\} h_e(s - x) ds
\]

\[
C_{Y_e}(t_1, t_2, x) = \int_{\max(x - v_t_1, 0)}^{\max(x - v_t_2, 0)} \{2y_e(t_1 - \Delta_2, s) + z_e(t_2 - \Delta_2, s)\} h_e(x - s) ds \\
+ \int_x^{\max(x - v_t_2, w)} [2C_{Y_e}(t_1 - \Delta_2, t_2 - \Delta_2, s) + C_{Z_e}(t_1 - \Delta_2, t_2 - \Delta_2, s) \\
+ 2y_e(t_1 - \Delta_2, s) y_e(t_2 - \Delta_2, s) + 2z_e(t_1 - \Delta_2, s) y_e(t_2 - \Delta_2, s) \\
+ 2y_e(t_1 - \Delta_2, s) z_e(t_2 - \Delta_2, s)] h_e(x - s) ds.
\]

We now state, without proof, the recurrence equations for the remaining count autocorrelations and count
crosscorrelations. For \( t_1 \leq t_2 \), the coupled recurrence relations for \( C_{Z_h}(t_1, t_2, x) \) and \( C_{Y_h}(t_1, t_2, x) \) are given as in (10)–(11), shown at the bottom of the next page.
Finally, the crosscorrelations \( C_Z(t_1, t_2, x) \) and \( C_Y(t_1, t_2, x) \) obey the following coupled recurrence equations: For \( t_1 \leq t_2 \), see (12)–(13), shown at the bottom of the next page. As before, when \( t_1 > t_2 \), we simply interchange \( t_1 \) and \( t_2 \) in all the previous expressions.

The preceding model can be easily modified to suit hole-injection APDs. This is done by simply interchanging the role of electrons and holes in all the recurrence equations (i.e., interchange the electron and hole ionization coefficients, saturation velocities and dead spaces).

We conclude this section by considering the effects of time-varying and stochastic carrier velocities. For layers with high fields (> 400 kV/cm), carriers do not assume their saturation velocities immediately following impact ionization. Rather, as has been demonstrated by Monte-Carlo simulation [45], there is a velocity overshoot at short distances (up to several multiples of the saturation velocity) following each impact ionization, as well as a random spread in the velocity. The overshoot effect can be readily incorporated into the dead-space model considered above in a straightforward fashion. Moreover, capturing the velocity randomness is also possible with a slight added complexity to the recurrence equations. The details of these extensions will be reported elsewhere.

Finally, with the statistics of the impulse response at hand, the statistics of the frequency response can be readily obtained. We define the stochastic frequency response as the Fourier transform of the random impulse response

\[
\mathcal{I}(f) \equiv \int_{-\infty}^{\infty} I(t) e^{-j2\pi ft} dt.
\]

The mean frequency response, \( \bar{\mathcal{I}}(f) \equiv \mathbb{E}[\mathcal{I}(f)] \), is then the Fourier transform of the mean impulse response \( i(t) \), defined in (4), which can be computed by solving the recurrence equations (6), (10)–(12) in [27]. The 3-dB bandwidth is then determined from the \( \mathcal{I} - \mathcal{I} \) curve. The APD power-spectral density, which is a measure of fluctuations in the frequency response, is defined by \( S(f) \equiv \mathbb{E}[\mathcal{I}^2(f)] \) and can be related to the autocorrelation function by using (14)

\[
S(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(t_1, t_2) e^{-j2\pi(f(t_1 + t_2))} dt_1 dt_2.
\]

SECTION IV.

Results

We now proceed to use the theory developed above to investigate the effect of reducing the width of the multiplication region on two fundamental performance characteristics: 1) the APD gain-bandwidth product; and 2) the fluctuations in the APDs frequency response. These characteristics are studied for four multiplication-region materials: GaAs, Al\(_{0.2}\)Ga\(_{0.8}\)As, InP, and In\(_{0.52}\)Al\(_{0.48}\)As, which are used in APDs of various multiplication-region widths. Following Anselm et al. [17], we use the uniform saturation velocities \( v_e = 0.8 \times 10^7 \text{cm/s} \) and \( v_h = 0.6 \times 10^7 \text{cm/s} \) for electrons and holes, respectively, in Al\(_{0.2}\)Ga\(_{0.8}\)As. For the other
materials, we use the saturation velocities $v_e=1.0\times10^7$ cm/s and $v_h=0.5\times10^7$ cm/s, for electrons and holes, respectively. Although these saturation velocities may vary according to the material, the previous values seem to serve well as a convenient approximation. All the devices considered operate on the basis of electron-injection, with the exception of InP, which is a hole-injection APD. Whenever needed, the mean gain is computed in accordance with the dead-space multiplication theory of the gain reported in [8], [10], [22].

A. Gain-Bandwidth Characteristics: Significance of Dead Space

To demonstrate the effect of dead space on the bandwidth, we compute the mean impulse response of a thin GaAs APD with a 100-nm multiplication region and compare it to the prediction provided by conventional multiplication theory, as shown in Fig. 1. Note that the dead-space multiplication theory (DSMT) predicts an elongated tail in the mean impulse response, an attribute that is consistent with our earlier work [27] and it arises because dead space tends, on average, to increase the time separation between consecutive impact ionizations. The predicted increase in the response time will, in turn, result in a bandwidth that is lower than that predicted by the conventional multiplication model, as can be seen from the frequency-response curves shown in Fig. 2. In this example, the conventional theory predicts a bandwidth of 37 GHz while the dead-space model predicts a value of 30 GHz, a reduction of 23%.

As expected, the effect of dead space on the bandwidth becomes progressively more important as the multiplication-region thickness decreases and may be neglected for thick devices. This is a consequence of the well-known fact that dead-space occupies a larger fraction of the multiplication layer as the thickness of the multiplication layer is reduced. To see this effect, we have generated plots of the bandwidth as a function of the mean gain for devices with different multiplication-region thicknesses in the range 100–800 nm. The results for GaAs and InP are depicted in Figs. 3 and 4, respectively. For example, it is seen from Fig. 4 that the conventional multiplication model overestimates the gain-bandwidth product by as much as 29% for the 100-nm InP device at gain value of 30; however, the discrepancy between the DSMT and conventional model predictions is significantly less (approximately 8.7%) for the 582-nm InP APD. It is also seen that the role of dead space becomes more significant as the gain increases and the buildup time becomes longer. For example, it is seen from Fig. 4 that the conventional multiplication model overestimates the gain-bandwidth product by as much as 32% for the 100-nm InP device at the high gain value of 37. In contrast, we see a discrepancy of 9% at a gain of 5.

In general, the presence of velocity overshoot (which we have neglected in our calculations) will tend to increase the bandwidth, which, in turn will serve to counteract the deleterious effect of the dead space on the bandwidth. We expect that the spread in the velocity will not have a significant effect on the mean impulse response and bandwidth (since its effect should be averaged out), but it is likely to have a noticeable effect on the power spectral density. A definitive analysis of the role of velocity overshoot and spread on the accuracy of the simple dead-space model is currently underway.

To compare our calculations to measurements, we considered bandwidth measurements for a resonant-cavity-enhanced (RCE) APD with 200-nm and 400-nm In$_{0.52}$Al$_{0.48}$As multiplication layers, reported by Lenox et al. [3]. These APDs were demonstrated to have gain-bandwidth products (at high gains) of 290 GHz and 130 GHz, respectively. (The experimental unity-gain bandwidths were measured as 10 GHz and 24 GHz for the 400-nm and 200-nm multiplication layers, respectively.) The theoretical bandwidth predictions are strictly buildup-time limited and hence they deviate from the experimental bandwidths at low gains (< 10), where RC effects dominate the bandwidth. In general, the agreement with the experimental bandwidth vs. gain plots is good in the range of high gains (> 15) with an approximate average error of 10%. The predicted
high-gain gain-bandwidth products are 253 GHz and 120 GHz for the 200-nm and 400-nm APDs, respectively. The larger error in the 200-nm device may be partially attributable to neglecting the velocity overshoot and/or to the nonuniform distribution of the electric field in the multiplication layer (as a result of background doping). This could further confine the location of impact ionizations, which would, in turn, cause a reduction in the buildup time in comparison with a uniform electric field. Part of the error could also be attributable to the inaccuracy of the assumed saturation electron and hole velocities in the model.

Fig. 1.

Mean impulse response of an APD with a 100-nm GaAs multiplication layer. The electric field was selected to be $6.7 \times 10^5$ V/cm, resulting in a mean gain of 9.4. Solid and dashed curves represent the predictions of the dead-space and the conventional models, respectively.
Fig. 2.

Normalized frequency response of the impulse response shown in Fig. 1. Solid and dashed curves represent the predictions of the dead-space and the conventional models, respectively. The dead-space prediction of the 3-db bandwidth is approximately 30 GHz; the conventional-model prediction is 37 GHz.
Bandwidth-versus-gain characteristics for GaAs APDs with various multiplication-layer widths. Solid and dashed curves represent the predictions of the dead-space model and the conventional model, respectively. Note that the conventional model overestimates the bandwidth and that the dead-space effect becomes more significant as the thickness is reduced. The range of the multiplication-layer electric field is $4 \times 10^5 - 7 \times 10^5$ V/cm for the 100-nm apd and $2.7 \times 10^5 - 3.5 \times 10^5$ V/cm for the 800-nm device.
Same as Fig. 3, but for InP multiplication layers. The range of the multiplication-layer electric field is $6 \times 10^5$–$9 \times 10^5$ v/cm for the 100-nm APD and $4 \times 10^5$–$5 \times 10^5$ v/cm for the 582-nm device.

B. Modeling the Dependence of the Gain-Bandwidth Product on the Multiplication-Layer Width

The dependence of the gain-bandwidth product on the multiplication-region width is shown in Figs. 5 and 6 for GaAs and InP, respectively. Similar curves (not shown) were also generated for In$_{0.52}$Al$_{0.48}$As and Al$_{0.2}$Ga$_{0.8}$As. Note that for each material, the gain-bandwidth characteristic curves are almost independent of the operational gain. This is a very desirable feature from a device-engineering perspective, since it implies that a single characteristic gain-bandwidth-product (GB) model can be used for all operational gains. Indeed, we can infer from the graphs of Figs. 5 and 6 that for each material, there is a pair of parameters, $c$ and $\kappa$, for which

\begin{equation}
\text{GB} = cw^{-\kappa}.
\end{equation}

These parameters, for each of the four materials, are provided in Table I.
Gain-bandwidth product as a function of the multiplication-layer width for GaAs. Different curves correspond to different operational gains. Note that the curves are almost overlapping.
Fig. 6.

Same as Fig. 5, but for inp multiplication layers.

Table I. Material-specific parameters for the model [given by (16)] describing the dependence of the gain-bandwidth-product on the multiplication-layer width.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>InP</th>
<th>In$<em>{0.52}$Al$</em>{0.48}$As</th>
<th>GaAs</th>
<th>Al$<em>{0.2}$Ga$</em>{0.8}$As</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>GHz/nm$^k$</td>
<td>14598.50</td>
<td>18844.21</td>
<td>16288.48</td>
<td>37639.44</td>
</tr>
<tr>
<td>$\kappa$</td>
<td></td>
<td>0.8234</td>
<td>0.8445</td>
<td>0.8690</td>
<td>1.0168</td>
</tr>
</tbody>
</table>

To exhibit the validity of the model given in (16), we apply the model to devices whose gain-bandwidth product were recently measured. Kinsey et al. [2] recently reported a record 320-GHz measurement of the gain-bandwidth product for a waveguide InGaAs/InAlAs APD, with a 150-nm InAlAs multiplication region. The use of (16) and the parameters for In$_{0.52}$Al$_{0.48}$As from Table I yields a predicted gain-bandwidth product of 274 GHz. Again, we expect that the error (14%) is attributable to the combined effects of the nonuniform electric field, errors in the assumed saturation velocities and neglecting velocity overshoot.
Fig. 7.

Mean frequency response (bold solid curve) and the square-root of the power spectral density (bold dash-dot curve) for a 100-nm GaAs APD. Thin curves represent predictions provided by the conventional model. Note the reduction in the frequency-response fluctuations predicted by the dead-space model.
Fig. 8.

Spectral SNR for three GaAs devices of different multiplication-layer widths. Note that the thinner devices produce higher SNRs.

C. Reduced Fluctuations in the Frequency Response for Thin APDs

The autocorrelation functions for the APDs considered in this paper have been computed by solving the recurrence equations (6), (8), (10)–(13). The numerical solutions are obtained using a simple iterative algorithm (as in [27]) which was implemented on a parallel computer. (The details of the computational aspects and the parallel computing will be reported elsewhere.) The two-dimensional (2-D) Fourier transform of the autocorrelation function was numerically computed and its diagonal values where obtained to yield the power-spectral density according to (15). Fig. 7 shows the power spectral density of the 100-nm GaAs (solid bold curve) overlayed with the mean frequency response (dashed bold curve). For comparison, the predictions of the mean frequency response and the power-spectral density corresponding to the conventional model are also shown in the same figure (shown as thin solid and dashed curves). Notably, the plots in Fig. 7 demonstrate that the spectral fluctuations in the APD frequency response are reduced as a result of the reduction in the multiplication-region width. This is a very promising result and it is the spectral generalization of the reduction of excess noise in thin APDs.
Same as Fig. 8 but the predictions are obtained using the conventional model. Unlike the dead-space-model predictions, the spectral SNR is independent of the multiplication-layer width.
Fig. 10.

Spectral SNR for three InP devices of different multiplication-layer widths. Note that thinner devices produce higher SNRs. The graph for each device is terminated at the corresponding bandwidth.

To establish this spectral-noise reduction more definitively, we define the spectral SNR

\[
\rho(f) = \frac{\bar{I}(f)}{S^1(f)}
\]

which is a measure of the relative spectral fluctuations as a function of frequency. Fig. 8 shows the prediction of the spectral SNR for GaAs for devices with different multiplication-region widths. For comparison, the predictions by the conventional model are shown in Fig. 9. Note that the conventional multiplication theory gives almost the same \( \rho \) value regardless of the width of the device, just as it predicts the same excess-noise factor regardless of the width. In contrast, the dead-space model clearly shows an increase in the spectral SNR \( \rho \) as the multiplication-region width decreases. For example in the case of InP, as shown in Fig. 10, \( \rho \) varies from 0.37 to 0.6 as the thickness varies from 800-nm to 100-nm (the prediction by the conventional model for InP is approximately 0.4 for all widths). The significance of the reduction in spectral fluctuation is that as the device becomes thinner, the contribution of ISI noise may decrease.
Finally, in all cases studied it was observed that the spectral noise is reduced at frequencies beyond the bandwidth of the device, as can be seen in Fig. 8 for the 800-nm GaAs APD. Although operational frequencies beyond the bandwidth of the device may not be of interest in applications, it is interesting to understand why this seemingly strange behavior occurs. It can be attributed to the fact that the photocurrent at such high frequencies is very low and the avalanche mechanism, which is the very source of multiplication uncertainty, is weak.

SECTION V.
Conclusion

In this paper, we investigated the effect of reducing the width of the multiplication region on the bandwidth, gain-bandwidth product and spectral SNR for thin III-V APDs. The theory developed for the time response specifically incorporates dead-space effects and extends the dead-space recurrence theory of Hayat and Saleh [27] to the autocorrelation function of the APDs impulse response. Our predictions showed that although the gain-bandwidth-product increases with the reduction of multiplication-region width, as a result of the reduction of the carrier transit time in the multiplication region, this increase is accompanied, to a less degree, by an excess buildup time resulting from the dead-space effect due to the inhibition (or dead time) between successive impact ionizations. The discrepancy between the bandwidth predictions of the conventional theory and the dead space-based theory can be high in certain thin APDs, depending on the gain and the width of the multiplication layer. As an example, for an InP APD with a 100-nm wide multiplication layer operating at a gain of 30, the discrepancy in bandwidth prediction is 29%. Our theory shows that the gain-bandwidth product characteristics of APDs can be represented by a simple hyperbolic-type model that depends only on the material. Comparison of the theoretical results with experiments indicate that the theory underestimates the gain-bandwidth product in certain devices. We suspect that this error is due to factors such as nonuniformity of the electric field, inaccuracy in the saturation velocities and ignoring the carrier velocity overshoot.

Analysis of the frequency-response fluctuations showed that the dead-space-induced excess buildup time is accompanied by an inherent benefit: the fluctuations in the frequency response, represented by the ratio of the mean frequency response to the power-spectral density, are reduced as a result of the dead-space effect. Hence, the reduction in photocurrent fluctuations is not limited to the dc component of the frequency response (i.e., the excess-noise factor) but generalizes to all operational frequencies within the device bandwidth. Because dead space tends to make the spectral characteristics of the output of thin-APD-based receivers “less random,” there is the potential that equalization techniques may become more effective in combating intersymbol interference.

APPENDIX Derivation Of The Renewal Equation For CZe(t,x)

We provide the details of the analysis that yields the development of the recurrence equation given in (6). Consider a parent electron at x (born at time t = 0) and suppose that it first impact ionizes at position $\xi > x$. To this parent electron, we associate the random count $Z_e(t, x)$, as defined earlier. Now let the electron counts $Z'_e(t_1 - (\xi - x)/v_e, \xi)$ and $Z''_e(t_1 - \xi - x/v_e, \xi)$ correspond to the first and second offspring electrons, respectively, born at position $\xi$. Also, let $Y_e(t_1 - (\xi - x)/v_e, \xi)$ correspond to the offspring hole born at $\xi$. According to the discussion directly preceding (6), which described the three possibilities for the time of the first impact ionization, the events associated with these cases can be further analyzed as follows:
SECTION A.

Case 1. The First Ionization Occurs Before \( t_1 \)

In this case, the position of the first ionization \( \xi \) satisfies \( (\xi-x)/v_e < t_1 \) since the distance to the first ionization is \( \xi-x \). Hence

\[
Z_e(t_1, x) = Z'_e(t_1 - \frac{\xi-x}{v_e}, \xi) + Z''_e(t_1 - \frac{\xi-x}{v_e}, \xi)
\]

\[+ Y_e(t_1 - \frac{\xi-x}{v_e}, \xi)\]

and

\[
= Z_e(t_2, x) = Z'_e(\frac{\xi-x}{v_e}, \xi) + Z''_e(t_2 - \frac{\xi-x}{v_e}, \xi)
\]

\[+ Y_e(t_2 - \frac{\xi-x}{v_e}, \xi).\]

In this case, the conditional correlation, given the knowledge of \( \xi \), becomes the equation shown at the top of the next page. Note that the range of \( \xi \) for which this case occurs is \( x \) to \( \min(x+v_e t_1, w) \). Also note that the processes \( Z'_e(\cdot, \cdot), Z''_e(\cdot, \cdot) \) and \( Y_e(\cdot, \cdot) \) are mutually independent, as they are generated by the independently-acting two electrons and the hole. Moreover, \( Z'_e(\cdot, \cdot) \) and \( Z''_e(\cdot, \cdot) \) are identically distributed.

\[
E[Z_e(t_1, x)Z_e(t_2, x)|\xi] = E[(Z'_e(t_1 - \frac{\xi-x}{v_e}, \xi) + Z''_e(t_1 - \frac{\xi-x}{v_e}, \xi)
\]

\[+ Y_e(t_1 - \frac{\xi-x}{v_e}, \xi)) \times (Z'_e(t_2 - \frac{\xi-x}{v_e}, \xi)
\]

\[+ Z''_e(t_2 - \frac{\xi-x}{v_e}, \xi) + Y_e(t_2 - \frac{\xi-x}{v_e}, \xi))].\]

SECTION B.

Case 2. The First Ionization Occurs Between \( t_1 \) and \( t_2 \)

In this case, the position of the first ionization \( \xi \) satisfies \( t_1 \leq (\xi-x)/v_e < t_2 \). Note that in this case \( Z_e(t_1, x) = 1 \) but
\[ Z_e(t_2, x) = Z'_e(t_2 - \frac{\xi - x}{v_e}, \xi) + Z''_e(t_2 - \frac{\xi - x}{v_e}, \xi) + Y_e(t_2 - \frac{\xi - x}{v_e}, \xi). \]

Hence

\[
E[Z_e(t_1, x)Z_e(t_2, x) | \xi] = E[(Z'_e(t_2 - \frac{\xi - x}{v_e}, \xi) + Z''_e(t_2 - \frac{\xi - x}{v_e}, \xi) + Y_e(t_2 - \frac{\xi - x}{v_e}, \xi))].
\]

Note that the range of \( \xi \) for which this case occurs is \( \min(x + v_e t_1, w) \) to \( \min(x + v_e t_2, w) \).

**SECTION C.**

**Case 3. The First Ionization Occurs After \( t_2 \)**

In this case, \((\xi - x)/v_e > t_2\) and hence necessarily \( Z_e(t_1, x) = Z_e(t_2, x) = 1 \). Hence

\[
E[Z_e(t_1, x)Z_e(t_2, x) | \xi] = 1, \quad t_2 < \frac{\xi - x}{v_e}
\]

and

\[
E[Z_e(t_1, x)Z_e(t_2, x) | \xi] = 0, \quad t_2 \geq \frac{\xi - x}{v_e}.
\]

Note that the probability that the first ionization occurs after \( t_2 \) is \( 1 - H_e(\min\{w - x, v_e t_2\}) \).

The final step is to average the conditional correlation \( E[Z_e(t_1, x)Z_e(t_2, x) | \xi] \) over all possible \( \xi \) in the interval \( [x, \min(x + v_e t, w)] \) with the careful consideration of the three cases mentioned earlier. This yields the recurrence relation given in (6).

**References**

1. G. S. Kinsey, C. C. Hansing, A. L. Holmes, Jr., B. G. Streetman, J. C. Campbell, A. G. Dentai, "Waveguide $\text{In}_{0.53} \text{Ga}_{0.47} \text{As-In}_{0.52} \text{Al}_{0.48} \text{As}$ avalanche photodiode", *IEEE Photon. Technol. Lett.*, vol. 12, pp. 416-418, Apr. 2000.


