2005

Gain-bandwidth product optimization of heterostructure avalanche photodiodes

Oh-Hyun Kwon
University of New Mexico

Majeed M. Hayat
Marquette University, majeed.hayat@marquette.edu

Joe C. Campbell
University of Texas - Austin

Bahaa E.A. Saleh
Boston University

Malvin Carl Teich
Boston University

Gain-bandwidth product optimization of heterostructure avalanche photodiodes

Oh-Hyun Kwon,
Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM
Majeed M. Hayat,
Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM
Joe C. Campbell
Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX
Bahaa E. A. Saleh,
Department of Electrical and Computer Engineering, Boston University, Boston, MA
Malvin C. Teich
Department of Electrical and Computer Engineering, Boston University, Boston, MA

Abstract:

A generalized history-dependent recurrence theory for the time-response analysis is derived for avalanche photodiodes with multilayer, heterojunction multiplication regions. The heterojunction multiplication region considered consists of two layers: a high-bandgap Al/sub 0.6/Ga/sub 0.4/As energy-buildup layer, which serves to heat up the primary electrons, and a GaAs layer, which serves as the primary avalanching layer. The model is used to optimize the gain-bandwidth product (GBP) by appropriate selection of the width of the energy-buildup layer for a given width of the avalanching layer. The enhanced GBP is a direct consequence of the heating of primary electrons in the energy-buildup layer, which results in a reduced first dead space for the carriers that are injected into the avalanche-active GaAs layer. This effect is akin to the initial-energy effect previously shown to enhance the excess-noise factor characteristics in thin avalanche photodiodes (APDs). Calculations show that the GBP optimization is insensitive to the operational gain and the optimized APD also minimizes the excess-noise factor.

SECTION I. Introduction

Avalanche PHOTODIODES (APDs) have the benefit of providing internal optoelectronic gain, which leads to an enhanced receiver sensitivity in comparison to p-i-n diodes in many high-speed lightwave systems. Unfortunately, the avalanche multiplication process, which is the source of the APD gain, degrades the operating speed due to the so-called avalanche buildup time, which is the time required for all the impact ionizations to take place once the APD is triggered by a photogenerated carrier. The buildup time is intrinsically dependent on the material (namely the ionization coefficients and threshold energies), but it also has a strong dependence on the applied electric field (which also controls the mean gain) and the structure of the device. In actuality, the buildup time is random, a nuisance that is a direct consequence of the spatial and temporal stochastic spread of the individual impact-ionization events. It is the finiteness and randomness of the buildup time that is often a primary source for intersymbol-interference (ISI) noise at the receiver in high-speed digital lightwave systems, which limits the maximum bit rate allowed for reliable communication.

For a given material, the buildup time, and hence the speed, can generally be reduced by decreasing the width of the active layer in conventional APDs, and by the width of the multiplication and absorption regions in separate–absorption–multiplication (SAM) APDs [1], [2]. This is because the transit times of primary and offspring carriers are reduced in a thin device. However, this comes at the expense of reduced sensitivity, as the quantum efficiency deteriorates due to the reduced absorption in thin layers. To remedy this sensitivity-versus-speed tradeoff, a number of novel structures have been developed. Lenox et al. reported a gain–bandwidth product (GBP) of 290GHz using a thin resonant-cavity InGaAs–InAlAs APD [1] and Nie et al. showed a GBP of 290 GHz using resonant-cavity separate absorption–charge–multiplication (InGaAs–AlGaAs) APD [3]. Later, Kinsey et al. reported a record GBP of 320 GHz using the edge-coupled waveguide APD [4]. Edge-coupling the light into a thin (<100 nm) layer, however, poses a new challenge and a number of schemes have been lately proposed to enhance the coupling efficiency [5] [6]–[7][8][9] [10].

Aside from the thickness of the multiplication region, there are a number of factors that further affect the avalanche buildup time. As the multiplication region becomes thin (e.g., below 200nm), the dead-

space effect becomes progressively more pronounced because the dead space occupies a larger fraction of the multiplication region [11], [12]. The dead space is the minimum distance a carrier must travel in the high-field multiplication region before reaching the ionization threshold energy [13], [14]. This spatial inhibition between successive ionizations, which is brought about by the dead space, turns out to increase the buildup time, as was initially shown analytically by Hayat and Saleh [15] and subsequently confirmed by others [16], [17]. However, dead space is not the only factor that affects the speed in thin multiplication layers. Monte Carlo simulations have shown that carriers experiencing the high electric field, which is required in thin multiplication region, have the tendency to ionize at early stages of their sojourn in the multiplication region [16], [18]. Moreover, those carriers that impact ionize early on are found to possess speeds well above the saturated drift velocities [18]. In fact, Hambleton et al. showed that this velocity-enhancement effect actually overcompensates for the dead-space effect [19].

In this paper, we show that the buildup time can be further reduced by considering a heterojunction multiplication region. The main mechanism for the enhancement is the so-called initial-energy effect. In particular, when a hot (energized) carrier initiates the avalanche multiplication, its first dead space is reduced depending on the magnitude of the initial energy of the carrier relative to the ionization threshold energy. The reduced first dead space, in turn, enhances the probability of the first ionization occurring sooner than the corresponding probability for the offspring, cold carriers. Roughly speaking, in the event that an early ionization takes place, the resulting buildup time is approximately that for a multiplication process that is initiated by two carriers (the parent and its offspring), each of which would be responsible for generating, on average, half of the required gain.

Clearly, we would expect a significant reduction in the overall build up time if we were to enhance the likelihood of the occurrence of the first ionization as early as possible. This is precisely what the heating of the parent carrier does by reducing the first dead space of the primary carrier. The heating of the primary carrier is accomplished by using a high field, high bandgap undoped layer adjacent to the avalanching region, which serves to energize the carriers to the ionization threshold energy of the avalanching-layer material. At the same time, the high-bandgap property of the energy-buildup layer reduces the possibility of having ionizations in it; thus, the heating of parent carriers is achieved while practically keeping the width of the multiplication region unaltered.

Interestingly, the above effect was shown earlier to reduce the excess noise factor as well in a host of low-noise APDs (termed impact–ionization–engineered, *I*2*E*, APDs) developed at the University of Texas. The low-noise characteristics of these bandgap engineered devices where shown to be a result of the initial energy effect using both analytical techniques [20], [21] as well as Monte Carlo simulation [16], [22]. We show in this paper that by carefully selecting the width of the energy buildup layer, the gain–bandwidth product can also be improved and optimized.

SECTION II. Modeling the Impulse Response in Heterostructure APDs

The buildup time of an APD is best characterized by the width of the mean impulse response, which is the average response to a single photoexitation. To calculate the mean impulse response, we will essentially use the recurrence technique reported in [15]. However, in this paper we will introduce two key generalizations to our recurrence technique: 1) we modify the recurrence technique to account for

the possibility of a reduced first dead space for a hot carrier that is injected into the multiplication region, and 2) we extend the theory to accommodate heterojunctions (i.e., multiple layers) within the high-field multiplication region.[1] The first generalization will be primarily used to study the extent of GBP enhancement that the initial-energy effect can bring about in the ideal situation when hot carriers are injected into a single-layer multiplication region from a doped (low-field) layer. The second generalization will be used to assess the anticipated GBP enhancement in the realistic case where the heating is performed by an additional undoped, high-bandgap layer (viz., the energy-buildup layer), in which case, the multiplication region will consist of the totality of the energy-buildup layer and the avalanching layer.

In the analysis that follows, we assume that carriers in the multiplication region drift with a constant velocity equal to their saturation velocity. This is primarily done for the simplicity of the analysis, as the main thrust of this work is to show the GBP enhancement rendered by a bandgap-engineered, heterojunction multiplication region. Further generalization to account for the position-dependent velocity of carriers (from the location of birth) can be done similarly to the technique developed by Ng et al. [17].

A. Recurrence Theory for Heterojunction Multiplication Regions

Consider an APD with a heterojunction multiplication region of total width w with a certain position-dependent electric field. Following the notation in [15], for a parent electron created in (or injected into) the multiplication region at location x, we define $Ze(t,x)$ as the total random number of electrons resulting from the parent electron t units of time after its creation. Similarly, $Zh(t,x)$ is defined as the total number of holes resulting from a parent electron at a location x at a time t after its birth. Now consider a case when a photogenerated electron is injected into the edge of the multiplication region (x=0) at time t=0. The value of the buildup-time-limited random impulse response $I(t)$ can be obtained by summing up the current contributions from all the offspring electrons and holes that are traveling in the multiplication region at time t. Assuming ve and vh as the saturation velocities of electrons and holes, respectively, Ramo's Theorem gives

$$I(t) = \frac{q}{w}[v_e Z_e(t,0) + v_h Z_h(t,0)]. \tag{1}$$

In accordance with the recurrence theory [15], in order to calculate the statistics of $Z_e(t,x)$ and $Z_h(t,x)$, we will need to introduce additional quantities $Y_e(t,x)$ and $Y_h(t,x)$. Here, $Y_e(t,x)$ is the total number of electrons resulting from a parent hole born at location x at time t after its birth, and $Y_h(t,x)$ is the total number of holes resulting from the parent hole. In [15], recurrence equations for the mean quantities $z_e(t,x)=\langle Z_e(t,x)\rangle$, $z_h(t,x)=\langle Z_h(t,x)\rangle$, $y_e(t,x)=\langle Y_e(t,x)\rangle$, and $y_h(t,x)=\langle Y_h(t,x)\rangle$ were developed and used to compute the impulse response of a homojunction APD under the assumption that the field in the multiplication region is uniform.

Generalization of the above recurrence method to accommodate heterojunctions can be obtained by modifying the probability density functions of the distance to ionization (carrier's free-path distance) in [15] so as to reflect the inhomogeneous nature (in terms of the electric field and material) of the

multiplication region. More precisely, let $h_e(\xi|x)$ denote the probability density function (pdf) of the net distance ξ (from x) to the first ionization resulting from an electron born at x. Similarly, we have $h_h(\xi|x)$ as the pdf of the net distance ξ to the first ionization resulting from a hole that is born at x. The generalized recurrence equations for $z_e(t,x)$ and $y_e(t,x)$ are

TABLE I Parameters for Ionization Coefficients (α for Electrons and β for Holes) and Ionization Threshold Energies of In0.52 Al0.48 As, GaAs, and Al0.6 Ga0.4 As

$$\alpha, \beta(\mathcal{E}) = A\exp[-(E_c/\mathcal{E})^m]$$

		A (cm^{-1})	E_c (V/cm)	m	E_{th} (eV)
GaAs [24]	α	6.01×10^6	2.31×10^6	0.901	1.9
	β	3.59×10^6	2.26×10^6	0.92	1.55
$Al_{0.6}Ga_{0.4}As$ [25]	α	3.47×10^6	2.82×10^6	1	3.4
($\mathcal{E} < 600$ kV/cm)	β	3.08×10^6	2.86×10^6	1	3.6
$Al_{0.6}Ga_{0.4}As$ [25]	α	1.39×10^6	2.27×10^6	1	3.4
($\mathcal{E} > 600$ kV/cm)	β	1.48×10^6	2.38×10^6	1	3.6
$In_{0.52}Al_{0.48}As$ [24]	α	4.17×10^6	2.09×10^6	1.2	2.15
	β	2.65×10^6	2.79×10^6	1.07	2.30

$$\alpha, \beta(\mathcal{E}) = Aexp[-(E_c/\mathcal{E})^m]$$

		$A(cm^{-1})$	E_c(V/cm)	m	$E_{th}(eV)$
GaAs[24]	α	6.01×10^6	6.01×10^6	0.901	1.9
	β	6.01×10^6	6.01×10^6	0.92	1.55
$Al_{0.6}Ga_{0.4}As$[25]	α	6.01×10^6	6.01×10^6	1	3.4
($\mathcal{E} < 600$kV/cm)	β	6.01×10^6	6.01×10^6	1	3.6
$Al_{0.6}Ga_{0.4}As$[25]	α	6.01×10^6	6.01×10^6	1	3.4
($\mathcal{E} > 600$kV/cm)	β	6.01×10^6	6.01×10^6	1	3.6
$In_{0.52}Al_{0.48}As$[24]	α	6.01×10^6	6.01×10^6	1.2	2.15
	β	6.01×10^6	6.01×10^6	1.07	2.30

(2)

$$\begin{aligned} z_e(t,x) &= u\left(\frac{w-x}{v_e} - t\right)[1 - H_e(v_e t|x)] \\ &\quad + \int_0^{\min(w-x, v_e t)} [2z_e(t - \xi/v_e, x+\xi) \\ &\quad + y_e(t - \xi/v_e, x+\xi)]h_e(\xi|x)d\xi \\ y_e(t,x) &= \int_0^{\min(x, v_h t)} [2y_e(t - \xi/v_h, x-\xi) \\ &\quad + z_e(t - \xi/v_h, x-\xi)]h_h(\xi|x)d\xi \end{aligned}$$

and those for $zh(t,x)$ and $yh(t,x)$ are

(3)

$$
\begin{aligned}
z_h(t,x) &= \int_0^{\min(w-x,v_e t)} [2z_h(t-\xi/v_e, x+\xi) \\
&\quad + y_h(t-\xi/v_e, x+\xi)]h_e(\xi|x)d\xi \\
y_h(t,x) &= u\left(\frac{x}{v_h}-t\right)[1-H_h(v_h t|x)] \\
&\quad + \int_0^{\min(x,v_h t)} [2y_h(t-\xi/v_h, x-\xi) \\
&\quad + z_h(t-\xi/v_h, x-\xi)]h_h(\xi|x)d\xi
\end{aligned}
$$

where $u(y)$=1 if $y \geq 0$ and 0 otherwise. The quantities

$$
\begin{aligned}
H_e(\xi|x) &= \int_{-\infty}^{\xi} h_e(x'|x)dx' \\
H_h(\xi|x) &= \int_{-\infty}^{\xi} h_h(x'|x)dx'
\end{aligned}
$$

are the cumulative probability distribution functions corresponding to the pdfs $he(\xi|x)$ and $hh(\xi|x)$, respectively.

In order to account for the material and field inhomogeneity in the multiplication region, the pdfs $he(x)$ and $hh(x)$ must incorporate: 1) the appropriate dead-space profile, which would accommodate the abrupt bandgap transition at the heterojunction and 2) the position-dependent ionization coefficients, which, in turn, depend on both the field value and the material at any specific location. In a hard-threshold dead-space model, the expressions for the pdfs are given by [20]

$$h_e(\xi|x) = \begin{cases} \alpha(x+\xi)\exp\left(-\int_{d_e(x)}^{\xi} \alpha(x+y)dy\right), & \xi \geq d_e(x) \\ 0, & \xi < d_e(x) \end{cases} \quad (4)$$

and

$$h_h(\xi|x) = \begin{cases} \beta(x-\xi)\exp\left(-\int_{d_h(x)}^{\xi} \beta(x-y)dy\right), & \xi \geq d_h(x) \\ 0, & \xi < d_h(x) \end{cases}. \quad (5)$$

The position- and material-dependent electron dead space $de(x)$ is calculated using the following implicit equation [20]:

$$q \int_{x}^{x+d_e(x)} \mathcal{E}(y)dy = E_{\mathrm{th},e}\left(x + d_e(x)\right) \quad (6)$$

where $E_{\mathrm{th},e}(x+d_e(x))$ is the electron ionization threshold energy for the material occupying the location $x+d_e(x)$. Similarly, the hole dead space $dh(x)$ is obtained using

$$q \int_{x}^{x-d_h(x)} \mathcal{E}(y)dy = E_{\mathrm{th},h}\left(x - d_h(x)\right). \quad (7)$$

In this paper, the model used to calculate the impact ionization coefficients, α and β, along with the threshold energies for electron and for holes (viz., A,Ee,m, and Eth) correspond to those developed by Saleh et al. [24] for GaAs and Plimmer et al. [25] for $Al_{0.6}$ $Ga_{0.4}$ As, as shown in Table I.

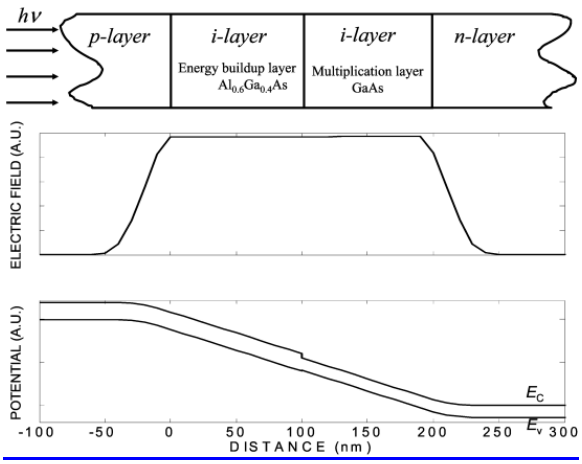

Fig. 1. Schematic of the heterostructure APD considered in this paper. The electric field distribution and the band diagram are also illustrated.

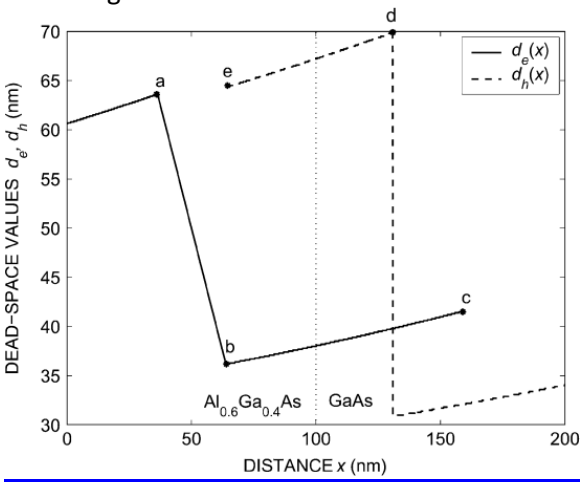

Fig. 2. Corresponding dead-space profiles for electrons (solid) and holes (dashed) in a heterojunction multiplication region under nonconstant electric field. The vertical dotted line represents the boundary of the two material.

Fig. 1 shows the schematic of the device structure to be examined. The device is a p-i-n heterostructure APD and the i-region consists of two layers: 100-nm $Al_{0.6}$ $Ga_{0.4}$ As and 100-nm GaAs. The $Al_{0.6}$ $Ga_{0.4}$ As layer serves as the energy-buildup layer, and the GaAs is the actual multiplication layer. The figure also includes the electric field distribution and and band diagram of the device. Fig. 2 shows the electron (solid curve) and hole (dashed curve) dead-space profiles for a representative heterojunction multiplication region shown in the Fig. 1. The dotted vertical line at x=100 nm indicates the boundary of the material. The applied electric field in the multiplication region is modeled to decrease linearly (in the direction pointing from $Al_{0.6}$ $Ga_{0.4}$ As to GaAs) due to unintentional doping [20]. Notice how the electron and hole dead spaces anticipate the bandgap boundary and change their values well before the boundary. This is because in calculating the dead space, we use the threshold energy that coincides with the material at the anticipated location at which the dead-space is reached. This behavior is represented in the dead-space (6) and (7) by the terms $x+de(x)$ and $x-dh(x)$ in the arguments of Eth,e and Eth,h, respectively. Let us now examine the electron dead-space profile shown in Fig. 2. Due to the linearly decreasing nature of the electric field, the electron dead space increases linearly in the $Al_{0.6}$ $Ga_{0.4}$ As layer (Eth,e=3.4 eV) up to the point "a." An electron born between positions

"a" and "b," however, cannot complete traveling its dead space in the $Al_{0.6}$ $Ga_{0.4}$ As layer. Nonetheless, as soon as it reaches the GaAs layer, which has a lower threshold energy (Eth,e=1.90 eV) than $Al_{0.6}$ $Ga_{0.4}$ As, it will have fulfilled the dead-space requirement for GaAs and becomes capable of ionizing immediately after crossing the boundary. Consequently, the dead space for this class of electrons is the distance from their location of birth to the boundary, which warrants the negative slope of the electron dead-space curve between points "a" and "b." Furthermore, in calculating the dead space for an electron that is born to the right of the point "b," the ionization threshold of GaAs is used, which results in the increasing behavior of the dead space beyond the point "b." Finally, note that the electron dead-space curve ends at the point "c" (x=159 nm in this particular case). This is because electrons born beyond the point "c" do not have a sufficient distance left to travel the dead space for GaAs and therefore escape the multiplication region without being able to impact ionize.

The behavior of the hole dead-space profile can be explained similarly. The important observation to make is that holes which are born to the right of point "d" have to overcome the dead space associated with the $Al_{0.6}$ $Ga_{0.4}$ As layer (Eth,h=3.6 eV). Unfortunately, the holes born between the position "d" and "e" are able to complete their dead space and become capable of ionizing in the $Al_{0.6}$ $Ga_{0.4}$ As layer. This will cause an undesirable hole-feedback effect in the $Al_{0.6}$ $Ga_{0.4}$ As layer. In a GBP-optimized structure, ionization of electrons and holes in the $Al_{0.6}$ $Ga_{0.4}$ As must be minimized. Details of the optimization will be discussed in Section IV.

B. Incorporating the Initial Energy of Parent Carriers

If the multiplication process is initiated by a hot parent carrier, the energy that the injected carrier needs to build up for impact ionization is reduced significantly by an amount equal to the carrier's initial energy. In the case of electron injection at location x, if the initial energy of the parent electron is Ei, then the first dead space that the electron will experience $de,i(x)$ is characterized by an equation similar to (6) but with Eth,$e(x+de(x))$ replaced by Eth,$e(x+de,i(x))-Ei$. (In cases when $Ei \geq E$th, we set $de,i(x)=0$.) As a result of the reduced first dead space for the parent electron, the pdf $he,i(\xi|x)$ of the distance to the first impact ionization for the parent electron will be similar to (4) but with $de(x)$ replaced with $de,i(x)$.

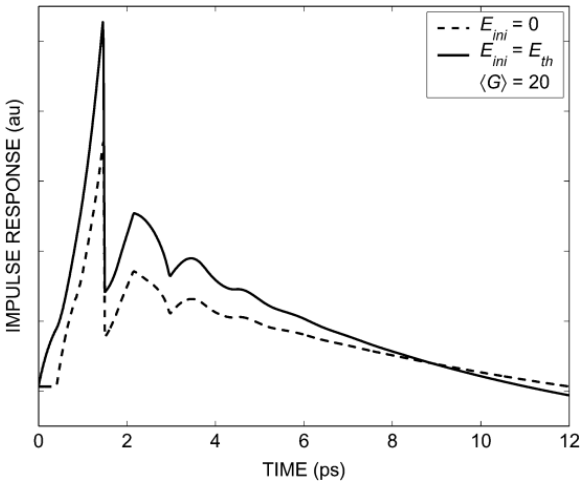

Fig. 3. Mean impulse response of a GaAs homojunction with a 100-nm multiplication layer with a mean gain of 20. The solid curve represents the impulse response with an initial energy equal to the ionization threshold

energy (E=682.8 kV/cm), and the dashed curve represents the impulse response with zero initial energy (E=689.5 kV/cm).

We now generalize the recurrence theory for the mean impulse response [15] to account for the initial-energy effect. Let $Ze,i(t,x)$ be defined as $Ze(t,x)$ with the exception that for the parent electron at x, the distance ξ to the first impact ionization has a pdf $he,i(\xi|x)$, as described previously. The key observation here is that upon the first ionization of the parent electron, the two newly created electrons and hole are assumed to have zero initial energy, independently of the initial energy of their parent electron. Consequently, conditional on the first ionization occurring at ξ, two independent copies of $Ze(t-\xi/ve,x+\xi)$ and one copy of $Ye(t-\xi/ve,x+\xi)$ are generated. Now by averaging over all possibilities for ξ, we obtain the following modified recurrence equation for the mean value $ze,i(t,x)=\langle Ze,i(t,x)\rangle$ and $zh,i(t,x)=\langle Zh,i(t,x)\rangle$, as follows:

(8) Top

(9) down

$$z_{e,i}(t,x) = u((w-x)/v_e - t)[1 - H_{e,i}(v_e t|x)] + \int_0^{min(w-x,v_e t)} [2z_e(t-\xi/v_e, x+\xi) + y_e(t-\xi/v_e, x+\xi)]h_{e,i}(\xi|x)d\xi$$

and

$$z_{h,i}(t,x) = \int_0^{min(x,v_h t)} [2z_e(t-\xi/v_h, x-\xi) + y_h(t-\xi/v_h, x-\xi)]h_{e,i}(\xi|x)d\xi.$$

Note that the quantities $ze(t,x)$ and $ye(t,x)$ must be computed *a priori*, according to (2), and then used to calculate $ze,i(t,x)$, and a similar statement holds in regards to $zh,i(t,x)$. Also note that if the parent electron has no initial energy, then (8) and (9) collapse to the top equations in (2) and (3), respectively. Finally, once the quantities $ze,i(t,x)$ and $zh,i(t,x)$ are computed, they are used to calculate the mean impulse response as usual. For example, in the case of electron edge injection (x=0), the mean impulse response function $i(t)\equiv E[I(t)]$ is determined using

(10)

$$i(t) = \frac{q}{w}\left[v_e z_{e,i}(t,0) + v_h z_{h,i}(t,0)\right].$$

SECTION III. Gain–Bandwidth Product

To see the extent of the effect of the initial energy of injected carrier, we consider a homojunction GaAs APD and examine its impulse responses and bandwidth with and without an initial energy equal to the ionization threshold of GaAs (i.e., either *E*ini=*E*th or *E*ini=0). Fig. 3 shows the impulse responses of a 100-nm homojunction GaAs APD. The electric field was chosen as 689.5 and 682.8 kV/cm for the cases *E*ini=0 and *E*ini=*E*th, respectively. This choice yields the same mean gain of 20 in both cases. (In calculating the mean gain for each APD, the recurrence dead-space multiplication theory [20], [21] was used.) In our calculations, we have assumed that the saturation velocities are 1.0×107 cm/s for electrons and 0.5×107 cm/s for holes. As expected, the impulse response for the *E*ini=0 case (dashed curve) exhibits a flat region, from 0 to 0.4 ps, which corresponds to the dead space, while the multiplication process initiated by fully energized carriers (solid curve) does not exhibit the initial dead-space region. Thus, carriers can initiate the first impact ionization earlier (possibly at the boundary) and have more chances to cause impact ionization before leaving the multiplication region. The solid curve reaches a higher value at the peak than the dashed curve does. Consequently, the solid curve decays at a faster rate than the dashed curve (since they both generate the same mean gain), which results in an improved bandwidth in the case when *E*ini=*E*th.

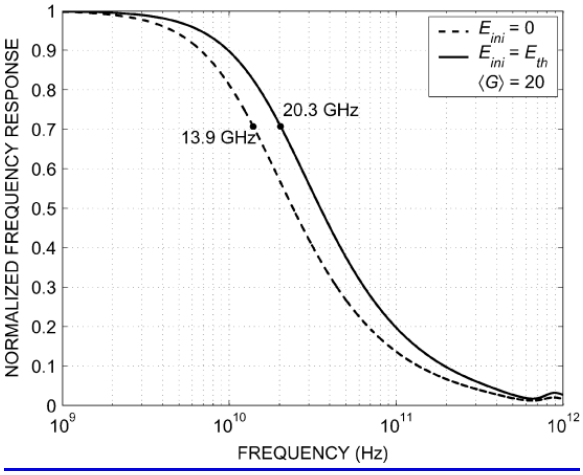

Fig. 4. Frequency response of a GaAs homojunction with a 100-nm multiplication layer with a mean gain of 20. The predicted 3-dB bandwidth for the APD with the initial energy is 20.3 GHz compared with 13.9 GHz for the zero-initial-energy case.

It is interesting to observe the oscillatory behavior of the impulse response. Within the transit time of the injected electron across the multiplication region, the parent electron and its offspring electrons and holes continue to impact ionize, thereby adding continuously (on average) to the number of carrier present in the multiplication region. As a result, the impulse response monotonically increases in this period until the “first-generation” group of electrons synchronously escape the multiplication region, at which point the mean impulse response drops sharply and abruptly. Subsequently, as the residual carriers yet present in the multiplication region continue to ionize, the impulse response increases. However, as the first-generation holes begin to exit the multiplication region, the impulse response gradually drops; thus, the time to the second peak is precisely the hole transit time, relative to the injection time of the parent electron. Evidently, these holes do not reach the end synchronously, and hence, the drop in the impulse response is no longer abrupt. Similarly, after the last of the first-

generation holes have exited the multiplication region, residual carriers within the region continue to ionize causing another rise in the impulse response until second-generation electrons start exiting the multiplication region, resulting in a subsequent valley, and so on. It is these interlaced waves of carrier buildup and carried departures that lead to the oscillatory behavior apparent in Fig. 3. The oscillations fade away, however, as the effects higher-order generations become indistinguishable in time from one another and as the number of carriers diminish.

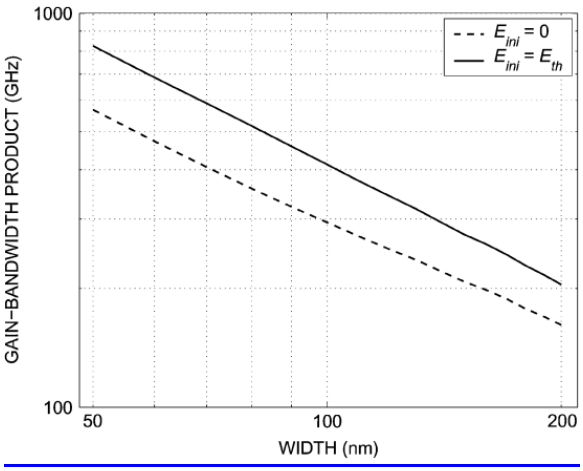

Fig. 5. Gain–bandwidth products of homojunction GaAs APDs with (solid) and without (dashed) the initial-energy effect as a function of the multiplication-layer width.

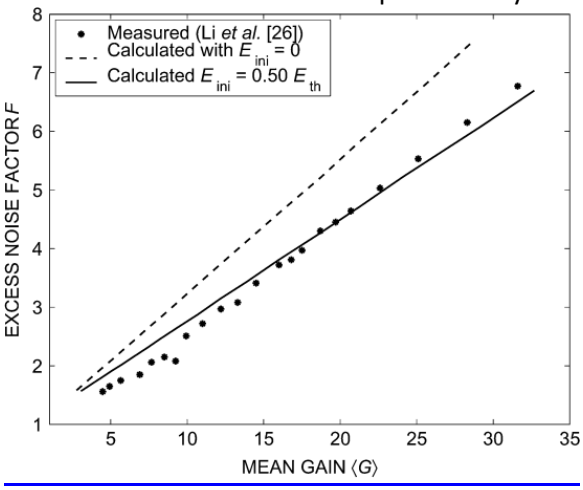

Fig. 6. Computed and measured excess noise factor versus the mean gain of a 150-nm InAlAs APD developed by Li et al. [26]. The symbols represent the measured data reported by Li et al..

The corresponding frequency response curves are shown in Fig. 4 depicting a predicted bandwidth improvement from 13.9 GHz (in the *E*ini=0 case) to 20.3 GHz (in the *E*ini=*E*th case). Fig. 5 shows the calculated GBPs of homojunction GaAs APDs as a function of the multiplication region width. The figure also demonstrates the improvement in the presence of the initial energy. In addition, the improvement is more significant as the width of the multiplication layer decreases. For example, the GBP improvement is 27% and 40% for the widths of 200 and 100 nm, respectively. This effect is most likely to be due to the increase in the significance of the dead space as the multiplication-region width decreases.

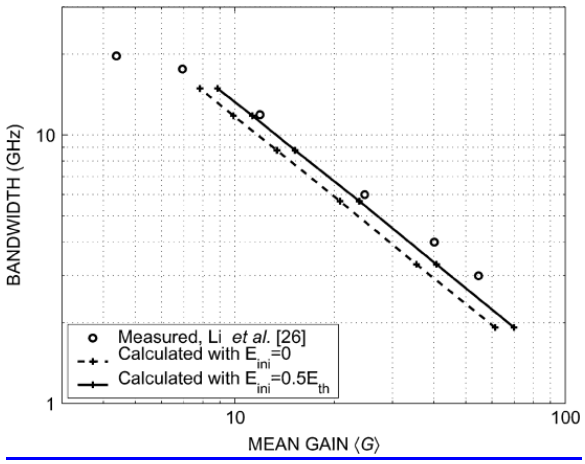

Fig. 7. Computed and measured bandwidth as a function of the mean gain. The symbols represent the measured data reported by Li et al.. The The velocities used for the calculation of bandwidth are 0.5×107 cm/s for electrons and 0.2×107 cm/s for holes.

As a validation of the our model, we computed the GBP of a 150-nm InAlAs APD developed by Li et al. [26]. We used the ionization coefficients for $In_{0.52}$ $Al_{0.48}$ As reported by Saleh et al. [24], as described in Table I. We first estimated the initial energy of the injected carriers, as a function of the applied electric field, by applying our noise recursive model [20] (which captures the initial-energy effect) to experimental measurements. This was done by fitting the excess-noise prediction (as functions of the mean gain) to the measured excess-noise factors while using the initial energy as a free parameter. It was found that the best fit is obtained when *E*ini=0.50*E*th, as shown in Fig. 6. This initial energy reduces the excess noise by approximately 20% compared with the no-initial energy case. Note that we observed a disagreement between calculation and experimental data for low values of gain. This may be due to trapping and other low-bias effects [27] that are not captured in our model. (The sensitivity of the excess-noise-factor versus mean-gain characteristics to the choice of *E*ini has been discussed extensively in [21].) We then applied the estimated initial energy to the recurrence theory developed in Section II-B and calculated the bandwidth. We used the saturation velocities in InAlAs reported by Ma et al. [28] (0.5×107 cm/s for electrons and 0.2×107 cm/s for holes). The calculated bandwidth, as a function the mean gain, is shown in Fig. 7. As shown in the figure, the calculated bandwidth with the initial energy is in better agreement to the measurement than those without the initial energy. The presence of 0.50*E*th of initial energy improves the bandwidth by approximately 17 percent compared to no initial energy is assumed. Albeit, the mere fact that predicted GBP, when the initial-energy effect is considered, is close to experimental data should not be literally used to establish an absolute accuracy for our model. In fact, no such claim is made here since a number of other factors that affect the speed, such as the velocity-enhancement effect [16] [17] [18], softness of ionization [29] [30]–[31] [32], phonon scattering, etc., have not been specifically included in our model. Nonetheless, the point of the present situation is that the inclusion of the initial energy, which was shown by our noise model [20] to yield an improved excess-noise prediction, also moved the GBP prediction in the correct direction.

SECTION IV. Optimization of the Gain–Bandwidth Product

In this section, we consider a heterostructure, two-layer multiplication-region APD in which the initial-energy mechanism is generated without the requirement that injected carriers have an initial energy prior to entering the high-field region. The schematic of the proposed heterostructure APD to be optimized is shown in the top portion of Fig. 2. (The very same structure was recently utilized to minimize the excess noise factor [21].) In the proposed structure, a high bandgap intrinsic $Al_{0.6}$ $Ga_{0.4}$ As layer, called the energy-buildup layer, is placed between the p-layer and the GaAs multiplication layer. The width of the $Al_{0.6}$ $Ga_{0.4}$ As energy-buildup layer plays a key role in improving the GBP. On one hand, its width must be large enough to allow the injected carrier to ballistically build up an energy which exceeds the ionization threshold energy of GaAs, where the multiplications are desired to take place. On the other hand, the width of the $Al_{0.6}$ $Ga_{0.4}$ As layer should be below the dead space associated with the $Al_{0.6}$ $Ga_{0.4}$ As layer. In this way, electrons cannot impact ionize in the $Al_{0.6}$ $Ga_{0.4}$ As layer; however, they enter the GaAs with an energy equal to or greater than the ionization threshold energy of GaAs.

Thus, the parent electron's first dead space inside the GaAs layer vanishes and at the same time no multiplication by it is permitted in the energy-buildup layer. With the latter property, the effective multiplication region of the APD is essentially the GaAs layer. Finally, note that for a given range of mean gains, any increase of the width of the GaAs multiplication layer corresponds to an inversely proportional reduction of the electric field. Since the function of the energy-buildup layer is strongly dependent on the range of the operational electric field (as it governs the energy buildup), it is conceivable that the optimum width of the $Al_{0.6}$ $Ga_{0.4}$ As layer would have a dependence on the width of the GaAs layer.

To rigorously establish the optimality of the proposed structure, we used the theory described in Section 2.1 to calculate the mean impulse response function of the type of heterostructure APDs shown in Fig. 2. No initial energy is assumed for carriers injected into the intrinsic $Al_{0.6}$ $Ga_{0.4}$ As layer; thus, we used (2) –(7). We used the saturation velocities of GaAs (1.0×107 cm/s for electrons and 0.5×107 cm/s for holes) throughout the device to simplify the computations. However, it is known that the velocity in Al_x Ga_{1-x} As drops as x, the Al composition, increases [33]. Our use of uniform saturation velocities can be justified by the fact that the difference of velocities between GaAs and $Al_{0.6}$ $Ga_{0.4}$ As is not very significant. For example, when x=0.5 (the highest composition available in the reference), the electron saturation drift velocity approaches to 0.9×107 cm/s, which is about 10% less than that in GaAs. The high-field hole saturation drift velocity when x=0.55 is also approximately equal to one for GaAs.

Fig. 8 shows the calculated GBP as a function of the width of the $Al_{0.6}$ $Ga_{0.4}$ As energy-buildup layer. For each width of the GaAs multiplication layer, a unique optimum width of the $Al_{0.6}$ $Ga_{0.4}$ As energy buildup layer exists. Note that the improvement in the GBP becomes more significant for thick devices. For example, the GBP improvement for a 200-nm GaAs APD is 17.9%, while the improvement is 15.5% for a 100-nm APD and 11.7% for a 70-nm APD. We believe that this is due to the overhead cost caused by an increase in the transit time of carriers, due to the energy-buildup layer, which becomes relatively more significant as the width of the GaAs multiplication layer decreases. In addition, the smaller GBP improvement observed in thinner heterostructures is also believed to be caused by the larger hole-to-

electron ionization coefficient ratio at high fields, which offsets the benefit of the larger relative dead space (relative to the width of the multiplication region).

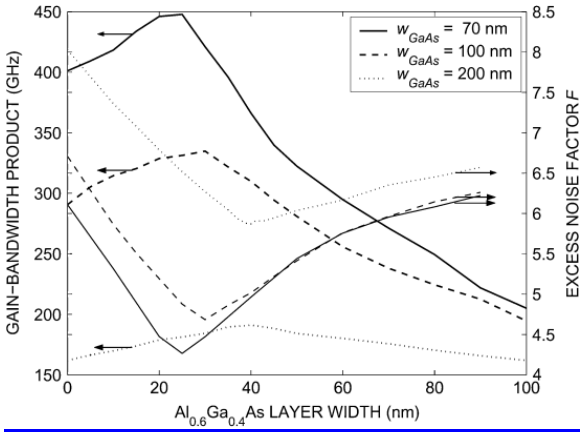

Fig. 8. GBPs (left vertical axis) of the proposed heterostructure APDs as a function of the width of the $Al_{0.6}$ $Ga_{0.4}$ As energy-buildup layer. The curves are parameterized by the widths of the GaAs multiplication layer. The corresponding excess-noise factors are also shown (right vertical axis).

We have also observed that the configuration that optimizes the GBP coincides with the configuration that minimizes the excess noise factor reported in [21]. This property is seen from the excess-noise graphs, also shown in Fig. 8 (right vertical axis). We expect this property to be a result of the fact that a minimum-noise configuration occurs when the probability of the occurrence of high gains is minimized [17] (i.e., by suppressing the tail of the probability mass function of the gain). However, since large gain realizations are precisely the ones that cause long buildup times, reducing the probability of such multiplication events therefore aids to enhance the decay of the tail of the mean impulse response as well. Finally, we have observed (not shown here) that the GBP is almost independent of the value of mean gain within a reasonable range (<30 in our calculations); thus, the optimal configuration, for a given GaAs multiplication layer, is almost uniformly optimal for all practical operational gains. Next, we will discuss the dead-space profiles and the ionization probability density functions for the optimized structures.

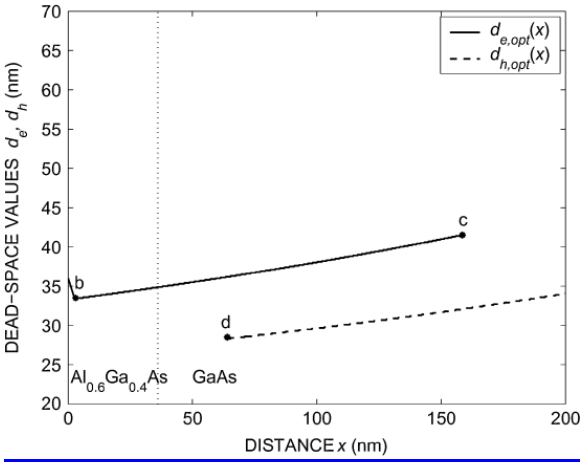

Fig. 9. Electron (solid) and hole (dashed) dead-space profiles for an optimum structure (36-nm $Al_{0.6}$ $Ga_{0.4}$ As/164-nm GaAs).

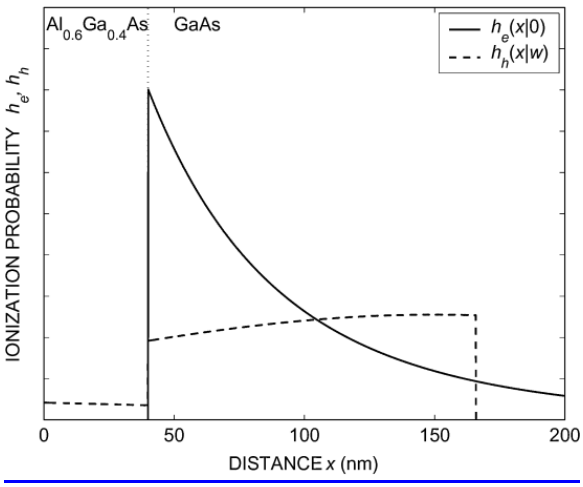

Fig. 10. Ionization pdfs for electrons (solid) and holes (dashed) for the optimal structure considered in Fig. 9. The dotted vertical line at x=40 nm indicates the boundary of the heterojunction.

A. Optimal Dead-Space Profile and Ionization Probability

The dead-space profiles for a representative optimal structure (36-nm $Al_{0.6}$ $Ga_{0.4}$ As/164-nm GaAs) is shown in Fig. 9. The dead space for the electrons born between x=0 and x=b is simply the distance to the heterojunction boundary, as explained in Section 2.1. (In contrast to the suboptimal structure shown in Fig. 2, the point "a" in Fig. 2 is located outside the GaAs avalanching region.) Electrons born beyond the point "b" complete their dead space within the GaAs layer. Thus, no electron is able to impact ionize within the $Al_{0.6}$ $Ga_{0.4}$ As layer. On the other hand, holes born to the left of point "d" will not be able to impact ionize in either layers. This is because the dead space is longer than the distance from their birth location to the left edge of the multiplication region. However, it is possible for holes born to the right of point "d" to travel the GaAs layer without ionizing and continue to accumulate the threshold energy of $Al_{0.6}$ $Ga_{0.4}$ As. These holes may therefore ionize in the $Al_{0.6}$ $Ga_{0.4}$ As layer. However, the probability of this latter scenario is low since it requires that the holes should travel through the GaAs layer, beyond their dead space, without ionizing. The fact that the ionization coefficient for GaAs is higher than that for $Al_{0.6}$ $Ga_{0.4}$ As makes the occurrence of this event unlikely.

The probability density functions (pdf) of electrons born at x=0 (solid curve) and holes born at x=w (dashed curve) are shown in Fig. 10. Note that the pdf of electron impact ionization before the boundary (x<36 nm) is zero due to the high ionization threshold energy of $Al_{0.6}$ $Ga_{0.4}$ As. The distance of the $Al_{0.6}$ $Ga_{0.4}$ As layer is not sufficient for electrons to acquire the energy to overcome the dead space. The pdf reaches its highest at the boundary (x=36 nm) due to the sudden drop in the ionization threshold energy in the GaAs layer. On the other hand, holes travelling to the left (starting at x=w) behave normally in the GaAs layer until they encounter the $Al_{0.6}$ $Ga_{0.4}$ As layer, which has a lower ionization coefficient, and therefore the pdf drops.

SECTION V. Conclusion

When the ionization process is initiated by a hot parent carrier, the first dead space associated with it is reduced by an amount that depends on the magnitude of the initial energy of the heated carrier relative to the ionization threshold energy of the material. Intuitively, the existence of such an initial energy enhances the probability of the first ionization occurring sooner than expected. For example,

consider the extreme case when the first ionization of the heated parent electron occurs almost immediately after injection at the edge of the avalanche multiplication region. In this simplified scenario, it is conceivable to think of the resulting multiplication process as two simultaneous multiplication processes generated by the two cold parent electrons. In a practical situation, however, it is unlikely that the heated carrier ionizes immediately after injection. However, its lifetime remains much shorter than that of a cold carrier due to its reduced dead space, and the analogy to the simplified “two-cold-for-one-hot” picture would still be meaningful. Further, due to the parallel nature of the resulting multiplication process corresponding to the cold parents, it would be reasonable to expect a reduction in the noise as well as the avalanche buildup time in structures that exhibit such an initial-energy effect.

Indeed, it has been shown previously that the initial-energy effect does serve to reduce the excess noise factor. In this paper, we showed that the initial-energy effect serves to improve the bandwidth as well. To do so, we developed a recurrence theory for calculating the mean impulse response of the avalanche photodiode (APD), which generalizes the existing theories to handle heterojunction multiplication regions and hot-carrier injection. We showed analytically that the initial-energy effect can naturally occur in a heterojunction multiplication region if the layers' widths and materials are designed carefully to enhance this effect. Particularly, we considered a two-layer heterojunction multiplication region consisting of a high-bandgap $Al_{0.6}$ $Ga_{0.4}$ As energy-buildup layer, which serves to heat up the primary electrons, and a GaAs layer, which serves as the primary avalanching layer. We used our analytical model to design a bandgap-engineered heterostructure APD which has optimal gain–bandwidth product (GBP). Notably, our calculations show that for the structure considered, maximizing the GBP is obtained by an optimal structure that also approximately minimizes the excess noise factor.

ACKNOWLEDGMENT

The authors would like to thank Dr. F. Ma and Dr. J. P. R. David for their valuable suggestions.

References

1. C. Lenox, H. Nie, P. Yuan, G. Kinsey, A. L. Holmes Jr., B. G. Streetman, J. C. Campbell, "Resonant-cavity InGaAs-InAlAs avalanche photodiodes with gain-bandwidth product of 290 GHz", *IEEE Photon. Technol. Lett.*, vol. 11, no. 9, pp. 1162-1164, Sep. 1999.

2. 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.

3. H. Nie, K. A. Anselm, C. Lenox, P. Yuan, C. Hu, G. Kinsey, B. G. Streetman, J. C. Campbell, "Resonant-cavity separate absorption charge and multiplication avalanche photodiodes with high-speed and high gain-bandwidth product", *IEEE Photon. Technol. Lett.*, vol. 10, pp. 409-411, Mar. 1998.

4. G. S. Kinsey, J. C. Campbell, A. G. Dentai, " Waveguide avalanche photodiode operating at 1.55 \$mu\$ m with a gain-bandwidth product of 320 GHz ", *IEEE Photon. Technol. Lett.*, vol. 13, no. 8, pp. 842-844, Aug. 2001.

5. F. J. Effenberger, A. M. Joshi, "Ultrafast dual-depletion region InGaAs/InP pin detector", *J. Lightw. Technol.*, vol. 14, no. 8, pp. 1859-1864, Aug. 1996.

6. K. Kato, S. Hata, K. Kawano, J. I. Yoshida, A. Kozen, "A high-efficiency 50 GHz InGaAs multimode waveguide photodetector", *IEEE J. Quantum Electron.*, vol. 28, no. 12, pp. 2728-2735, Dec. 1992.

7. N. Shimizu, N. Watanabe, T. Furuta, T. Ishibashi, "InP̄InGaAs uni-traveling-carrier photodiode with improved 3-dB bandwidth of over 150 GHz", *IEEE Photon. Technol. Lett.*, vol. 10, no. 3, pp. 412-414, Mar. 1998.

8. K. J. William, R. D. Esman, "Large-signal compression-current measurements in high-power microwave pin photodiode", *Electron. Lett.*, no. 35, pp. 82-84, Jan. 1999.

9. S. Demiguel, L. Giraudet, L. Joulaud, J. Decobert, F. Blache, V. Coupé (Coupe), J. Jorge, P. Ragnod-Rossiaux, E. Boucherez, M. Achouche, F. Devaux, "Evanescently coupled photodiodes integrating a double stage taper for 40 Gb/s applications-Compared performance with side-illuminated photodiodes", *J. Lightw. Technol.*, vol. 12, no. 12, pp. 2004-2014, Dec. 2002.

10. F. Xia, J. K. Thomson, M. R. Gokhale, P. V. Studentkov, J. Wei, W. Lin, S. R. Forrest, "An asymmetric twin-waveguide high-bandwidth photo diode using a lateral taper coupler", *IEEE Photon. Technol. Lett.*, vol. 13, no. 8, pp. 845-847, Aug. 2001.

11. 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.

12. M. A. Saleh, M. M. Hayat, B. E. A. Saleh, M. C. Teich, "Dead-space-based theory correctly predicts excess noise factor for thin GaAs and AlGaAs avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 47, no. 3, pp. 625-633, Mar. 2000.

13. B. E. A. Saleh, M. M. Hayat, M. C. Teich, "Effect of dead space on the excess noise factor and time response of avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 37, no. 10, pp. 1976-1984, Oct. 1990.

14. 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.

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. P. J. Hambleton, J. P. R. David, G. J. Rees, "Enhanced carrier velocity to early impact ionization", *J. Appl. Phys.*, vol. 95, pp. 3561-3564, Apr. 2004.

17. 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.

18. S. A. Plimmer, J. P. R. David, B. Jacob, G. J. Rees, "Impact ionization probabilities as functions of two-dimensional space and time", *J. Appl. Phys.*, vol. 89, pp. 2742-2751, Mar. 2001.

19. P. J. Hambleton, B. K. Ng, S. A. Plimmer, J. P. R. David, G. J. Rees, "The effects of nonlocal impact ionization on the speed of avalanche photodiodes", *IEEE Trans. Electron Devices*, vol. 50, no. 2, pp. 347-351, Feb. 2003.

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. O. Kwon, M. M. Hayat, S. Wang, J. C. Campbell, A. L. Holmes Jr., B. E. A. Saleh, M. C. Teich, " Optimal excess-noise reduction in thin heterojunction Al \$_{0.6}\$ Ga \$_{0.4}\$ As-GaAs avalanche photodiodes ", *IEEE J. Quantum Electron.*, vol. 39, no. 10, pp. 1287-1296, Oct. 2003.

22. F. Ma, S. Wang, X. Li, A. Anselm, X. G. Zheng, A. L. Holmes Jr., J. C. Campbell, "Monte Carlo simulation of low-noise avalanche photodiodes with heterojunctions", *J. Appl. Phys.*, vol. 92, pp. 4791-4795, Oct. 2002.

23. A. Bandyopadhyay, M. J. Deen, L. E. Tarof, W. Clark, "A simplified approach to time-domain modeling of avalanche photodiodes", *IEEE J. Quantum Electron.*, vol. 34, no. 4, pp. 691-699, Apr. 1998.

24. M. A. Saleh, M. M. Hayat, P. P. Sotirelis, A. L. Holmes Jr., 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. A. Plimmer, J. P. R. David, R. Grey, G. J. Rees, " Avalanche multiplication in Al \$_x\$ Ga \$_{i-x}\$ As (\${rm x} = 0\$ to 0.60) ", *IEEE Trans. Electron Devices*, vol. 47, no. 5, pp. 1089-1097, May 2000.

26. N. Li, R. Sidhu, X. Li, F. Ma, X. Zheng, S. Wang, G. Karve, S. Demiguel, A. L. Holmes Jr., J. C. Campbell, "InGaAs/InAlAs avalanche photodiode with undepleted absorber", *Appl. Phys. Lett.*, vol. 82, pp. 2175-2177, Mar. 2003.

27. N. R. Das, M. J. Deen, "Low-bias performance of avalanche photodetectors-A time-domain approache", *IEEE J. Quantum Electron.*, vol. 37, no. 1, pp. 69-74, Jan. 2001.

28. F. Ma, X. Li, J. C. Campbell, J. D. Beck, C.-F. Wan, M. A. Kinch, " Monte Carlo simulations of Hg \$_{0.7}\$ Cd \$_{0.3}\$ Te avalanche photodiodes and resonance phenomenon in the multiplication noise ", *Appl. Phys. Lett.*, vol. 83, pp. 785-787, Jul. 2003.

29. N. Sano, T. Aoki, A. Yoshii, "Soft and hard ionization thresholds in Si and GaAs", *Appl. Phys. Lett.*, vol. 55, pp. 1418-1420, Oct. 1989.

30. D. S. Ong, K. F. Li, S. A. Plimmer, G. J. Rees, J. P. R. David, P. N. Robson, " Pull band Monte Carlo modeling of impact ionization avalanche multiplication and noise in submicron GaAs \$p^{+}{-}i{-}n^{+}\$ diodes ", *J. Appl. Phys.*, vol. 87, pp. 7885-7891, Jun. 2000.

31. C. H. Tan, J. R. R. David, G. J. Rees, R. C. Tozer, "Treatment of soft threshold in impact ionization", *J. Appl. Phys.*, vol. 90, pp. 2538-2543, Sep. 2001.

32. O. 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.

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