AVALANCHE DIODES

When a high reverse voltage is applied to a semiconductor diode, a large avalanche current often flows. Electrons and holes are accelerated to energies so high that they collide with atoms in the crystal and ionize them, creating new electron–hole pairs. The secondary carriers can also initiate ionization, leading to a chain reaction with potentially destructive consequences. The rapid increase of current with voltage is referred to as *avalanche breakdown.* The amplification of the original current is called avalanche *multiplication* or *gain.* As the bias approaches a characteristic *breakdown voltage,* the current will usually rise by many factors of 10, reaching a limit imposed by another mechanism or destroying the device. Several useful semiconductor devices exploit this mechanism to obtain favorable performance. A few of the important types are listed here.

Zener Diode. This is designed to exhibit an abrupt avalanche breakdown at a well-defined voltage, at which the current rises dramatically. Since the voltage stays close to the breakdown value over a wide range of currents, these diodes make effective voltage regulators or surge protectors. The term *Zener diode* can also refer to diodes that break down by an entirely different mechanism, the Zener effect. Most practical Zener diodes, however, are actually avalanche diodes.

Avalanche Photodiode (APD). Here the avalanche effect amplifies the flow of current resulting from light incident on the diode. This is beneficial when the optical sensitivity is limited by noise in the amplifier following the photodiode. The larger signal from an APD helps to overcome amplifier noise. Unfortunately, the avalanche process introduces noise of its own, which leads to degradation in the signal-to-noise ratio when the avalanche gain exceeds an optimum value at which the APD noise and the amplifier noise are comparable.

IMPATT and Other Transit-Time Diodes. The *impact ionization avalanche transit time* (IMPATT) diode employs the physics of impact ionization and transit-time effects to create a high-frequency negative resistance. In one simple structure, the device combines an amplification mechanism, a time delay, and feedback. IMPATTs can be used as solid-state microwave oscillators or amplifiers. A number of variants and related device structures exist including the barrier injection and transit time (BARITT), double velocity transit time (DOVETT), and trapped plasma avalanche triggered transit (TRAPATT) diodes.

OVERVIEW

All avalanche diodes rely on the same physical principles for operation. In the remainder of this article, these principles will be illustrated using the avalanche photodiode (APD) (1–3) as an example. Other avalanche diodes are discussed in more detail in separate articles. For more on Zener diodes, see DIODES FOR POWER ELECTRONICS and SURGE PROTECTION. For more on IMPATTs and related devices, see TRANSIT TIME DEVICES.

Figure 1 shows a one-dimensional cross section of a particular avalanche diode design that will be analyzed later in this article. A junction is formed between two semiconductor re-

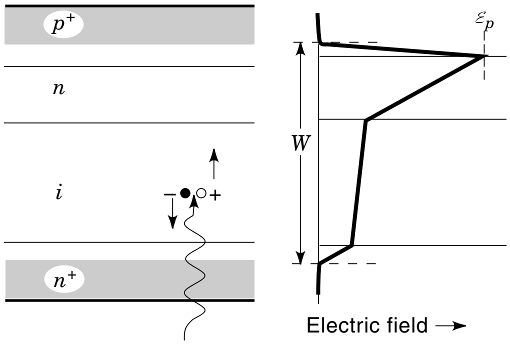

Figure 1. Cross-section of a p^+nin avalanche diode showing the electric field profile and the depletion region of width W. Avalanche multiplication occurs near the peak field at the p^+n junction. An incident photon is absorbed, producing an electron–hole pair. Electrons and holes are accelerated in opposite directions.

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*. Copyright © 1999 John Wiley & Sons, Inc.

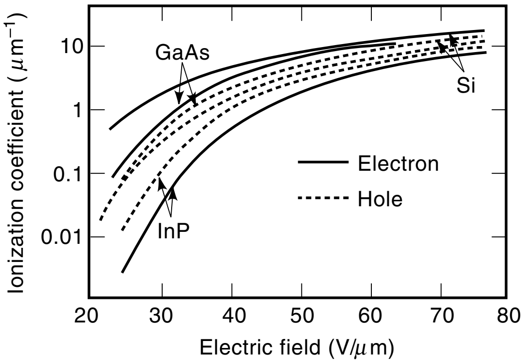

Figure 2. Measured ionization coefficients for Si (4), InP (5), and GaAs (6) as a function of the electric field.

gions that are doped *p* and *n* type, comprising a *pn* junction diode, with electrical contacts made to the *p* and *n* regions. When a negative voltage is applied to the *p* electrode (reverse bias), a region of high field forms near the *pn* junction, sweeping out most of the electrons and holes. This leaves a *depletion region* that is almost free of carriers, but contains a space-charge due to the immobile donor and acceptor ions. The figure also shows the electric field profile under these bias conditions. The peak $\mathscr{E}_p$ occurs at the *pn* junction with the field falling off to zero at both edges of the depletion region. The rate at which the field falls to zero is determined by the density of ions within the space-charge region, and thus on the concentration of dopants in the *p* and *n* regions. For a given bias voltage, higher doping leads to a more rapid change of electric field, a higher peak field, and a thinner depletion region.

If an electron or hole finds its way into the depletion region, it will be accelerated by the electric field and swept out, creating a current in the electrodes as it moves. Holes and electrons can diffuse into the depletion region from the adjacent neutral regions or be generated within the depletion region by thermal excitation or by the absorption of light. Avalanche photodiodes are designed so that light incident on the diode is absorbed within the depletion area, creating electron–hole pairs that produce a photocurrent under the influence of the electric field.

As carriers are accelerated by the electric field, they collide with atoms in the semiconductor material and, on average, reach a terminal velocity, the drift velocity, that is determined by the electric field and by the material's velocity–field relationship. This relationship represents a statistical average over an ensemble of particles distributed within a range of velocities. If the field is high enough, the more energetic carriers in the distribution will collide with neutral atoms with sufficient energy to ionize them, kicking valence electrons into the conduction band and producing secondary electron–hole pairs. Both electrons and holes can initiate this impact ionization process. The creation of secondary electron–hole pairs is inherently random and can be described by an average probability of ionization per unit length for each of the electrons or holes traveling within the depletion region. Not surprisingly, the ionization coefficients depend strongly on the electric field, the type of the initiating carrier, and the semiconductor material. Figure 2 shows the dependence of the electron (α) and hole (β) ionization coefficients for several important semiconductor materials as a function of the electric field (4–6). Notice that in Si and GaAs, electrons are more likely to initiate impact ionization than holes. For InP, the situation is reversed.

Figure 3 shows a simplified picture of the avalanche multiplication that occurs when the electric field is high enough. The top part of the figure is a space–time diagram of the flow of carriers. The carriers are shown traveling at their saturated drift velocity, ignoring the acceleration and deceleration that occur under the influence of the electric field and scattering events. While not quantitatively accurate, this picture helps in understanding the essential physics. At the left side, a photon moves at high speed and produces an initial electron–hole pair. Under the influence of the electric field, holes travel toward the top edge of the depletion region, while electrons travel toward the bottom edge. The carriers are rapidly accelerated to the drift velocity and (on average) travel along straight lines in the space–time diagram. The slope of the upward-moving hole trajectories is slightly less than that of the electrons, reflecting a lower saturation velocity. The primary electron travels toward regions of progressively smaller field and reaches the bottom edge of the space–charge region. The hole moves into the higher field region near the *pn* junction and initiates the first impact ionization event, producing a secondary electron–hole pair. The secondary carriers of both types produce additional ionization events, leading to a chain reaction that produces a considerable number of carriers. If the bias voltage is high enough, the avalanche will continue to grow until self-heating, series resistance, or some other mechanism limits the current. At lower bias, the probability of ionization will be too low to sustain the chain reaction and it will terminate, but the current will be enhanced by the avalanche effect. The bottom part of Fig. 3 shows the flow of current within the diode. A displacement current proportional to the drift velocity flows in the electrodes whenever there is a charge carrier moving within the depletion region. The area under the curve gives the total charge flow within

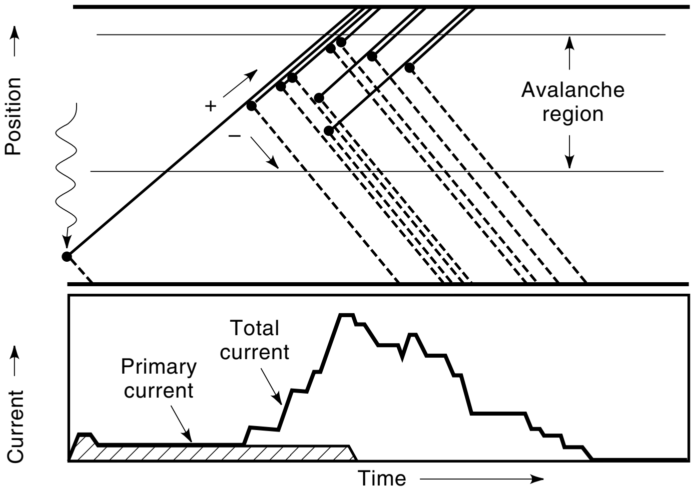

Figure 3. Space–time diagram for carrier multiplication in an avalanche diode like that in Fig. 1. The graph at the bottom shows the flow of primary and secondary current in the diode. Current flows whenever charge is moving within the depletion region. Holes move up toward the *p* side of the diode, electrons move down toward the *n* side.

the diode. In this example, 10 electron–hole pairs—the original primary pair plus nine secondary pairs—contribute to the diode current. The area under the curve is 10 times larger than it would have been had there been no impact ionization. Thus, each ionization produces one additional electronic charge at the electrodes, and this event has an avalanche gain of ten.

When the avalanche diode is used as a photodetector, the avalanche process degrades the signal in two important ways that are both evident in Fig. 3. First, the duration of the electrical pulse is longer than the pulse caused by the primary carriers. Consequently, the frequency response of an avalanche photodiode is worse than that of a conventional photodiode. Second, the shape and size of the pulse produced by each initiating carrier will be different because of the statistical nature of the avalanche multiplication process. The variation of the gain (area under the curve) is especially important. It adds unavoidable noise to the signal and ultimately limits the usable gain. Both of these effects get worse as the avalanche gain increases. Later sections will show that the performance is best when only one type of carrier initiates ionization. The ratio of the ionization coefficients for the two carrier types (k) is one of the most important parameters in determining the characteristics of an APD. To obtain the best performance, one invariably designs the device so that the carrier type that has the highest ionization coefficient is preferentially injected into the region in which avalanche multiplication occurs. The parameter k is usually defined as the ratio of the ionization coefficient of the less strongly ionizing carrier to that of the more ionizing carrier. Much of the history of the development of APDs has been devoted to the quest for materials and device structures that minimize the k ratio.

Figure 4 shows how the avalanche gain M depends on voltage for a simple structure. At low bias, the gain is close to one and increases gradually as the voltage is raised. As the bias approaches a particular voltage, the gain increases rapidly. As long as there is some feedback between the carrier types (i.e., k is not exactly zero), a run-away will occur at a finite breakdown voltage V_b. Without overload protection, the diode will usually be destroyed.

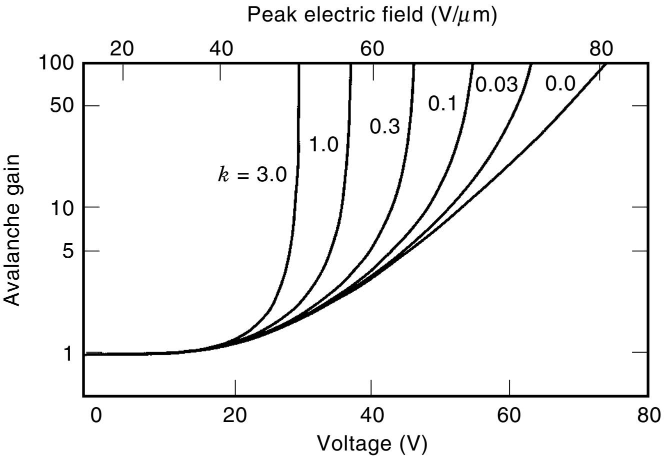

Figure 4. Avalanche gain as a function of voltage for a p^+n InP diode doped at 3×10^{16} cm^{-3} for several assumed values of the k ratio. The upper scale shows the peak value of the electric field at each voltage. The hole ionization coefficient β is from Ref. 5. The electron ionization coefficient is obtained by multiplying β by the assumed value for k.

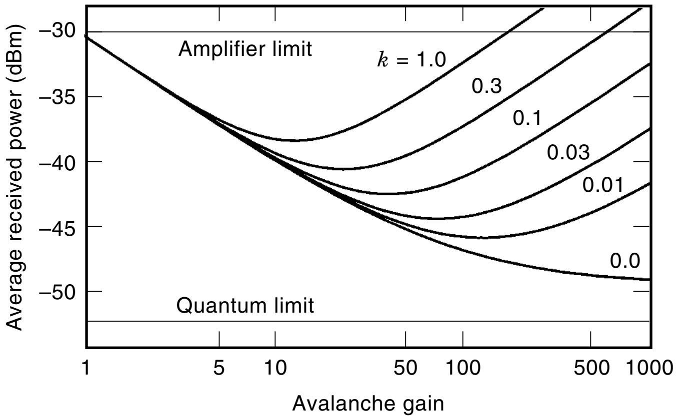

Figure 5. The minimum average receiver power needed to obtain a bit error rate of 10^{-9} at 2.5 Gb/s using an APD with 100% quantum efficiency and an amplifier-limited sensitivity of −30 dBm for several values of the k ratio.

Figure 5 shows the sensitivity of an APD receiver. The optical power required to achieve a given signal-to-noise ratio is plotted versus the avalanche gain for a given value of the amplifier noise. At low gain, the required optical power decreases in inverse proportion to the gain. For example, at a gain of two, only half the optical power is required to achieve a given signal-to-noise ratio. If the gain for each primary electron–hole pair were always the same, the gain could be increased until the effect of the amplifier noise was made negligible and shot noise in the optical signal became the limiting source of noise. The sensitivity would then approach the quantum-limited value shown in the figure. Unfortunately, the excess noise from gain fluctuations in the APD eventually predominates, leading to degradation in performance when the gain exceeds an optimum value. The optimum gain and sensitivity depend on both the APD and the receiver characteristics. For receivers used in long-wavelength telecommunications systems, the improvement is typically close to 10 dB (7).

The principles introduced here will be covered in more detail in the following sections.

IMPACT IONIZATION

Physical Principles

The behavior of an avalanche diode depends critically on the physics of impact ionization. The fundamental principles are well understood, but analytical and numerical models can be complex since they involve high electric fields in which many of the simplifying approximations of semiconductor transport are not valid. Rather than give a detailed treatment, the essential physics will be described and a simple model will be mentioned that captures some of the important behavior and provides a useful model for empirical modeling of measured data.

When an electron experiences an electric field within a semiconductor, it begins to accelerate. As it moves, it collides with scattering centers and loses energy. On average, an ensemble of these electrons reaches a velocity at which the average energy lost per collision is balanced by the increase in kinetic energy that occurs as the electrons accelerate between collisions. At low fields, this drift velocity rises in direct proportion to the field. The constant of proportionality is called the mobility. For the large values of electric field that are found in avalanche diodes, the drift velocity usually saturates and becomes approximately independent of the electric field. It is useful to describe the behavior in terms of an average scattering length or mean free path λ and an average energy loss per collision E_o. Some of the electrons will have energies much higher than the average value. These *lucky electrons* have experienced a smaller number of collisions or a smaller energy loss per collision than average. Things get interesting when one of these lucky electrons collides with an atom within the crystal with sufficient energy to ionize it, thus creating two new free carriers, an electron and a hole. The initiating electron must have enough energy to guarantee the conservation of energy and momentum in this interaction. Energy conservation alone requires that the ionization threshold be greater than the bandgap energy E_g, since this is the minimum energy required to create an electron–hole pair. Since the momentum of the initiating carrier must also be conserved, the threshold energy is always higher than the bandgap. If the electrons and holes have the same effective mass and if the band energies depend quadratically on the momentum (parabolic bands), it can be shown that the kinematic constraints can be met only when the energy of the primary electron or hole exceeds one and one-half times the bandgap energy ($E_i = 1.5E_g$). The parabolic band approximation is grossly violated for real semiconductors at these large excitation energies, but once the band structure is known, the thresholds are readily computed and typically yield values a bit larger than E_g.

Baraff Theory

There is a long history of attempts to understand the physics of impact ionization in semiconductors (8–10). Baraff (10) developed a general theory of ionization by solving the Boltzmann transport equations for a simplified model for which E_o, E_i, and λ completely describe the material. The model yields a carrier velocity distribution that is intermediate between the diffuse distribution in Wolff's theory (8) and the strongly peaked *lucky electron* distribution in Shockley's theory (9). When the predicted ionization coefficients are plotted against the electric field, the resulting curves depend only on the ratio E_o/E_i when the axes are properly normalized. In Fig. 6 this universal plot is shown for several values of the parameter E_o/E_i. The Baraff theory has been successful in explaining the ionization coefficients in a number of materials including Ge (11) and Si (4), although sometimes requiring unrealistic values for the fit parameters. It provides a simple model for ionization that explains the general shape of the ionization curves and clarifies the underlying physics. A feature of both theoretical and experimental coefficients is an exponential dependence on $1/\mathscr{E}$. Literature values are usually given in the form $\alpha = A \exp(-(\mathscr{E}_a/\mathscr{E})^m)$, where A, $\mathscr{E}_a$, and m are empirical values chosen to fit the data over a restricted range of electric field values (12). More recent models such as the *lucky drift* theory (13,14) have improved the agreement between theory and experiment and aided the physical understanding.

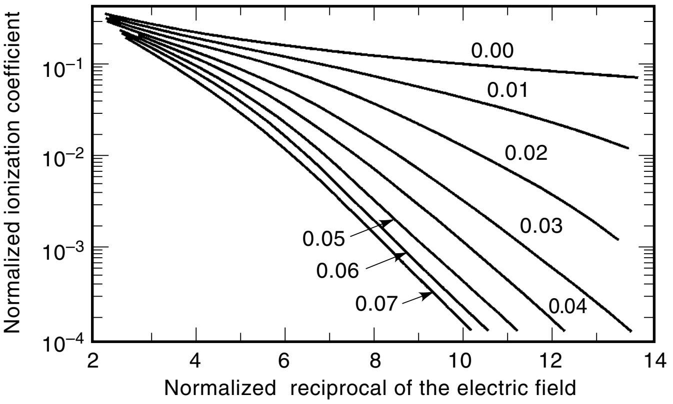

Figure 6. The results of Baraff's (10) numerical calculation of ionization coefficient plotted as universal curves on normalized axes. The ionization coefficient is normalized by the mean free path λ, and the electric field $\mathscr{E}$ is normalized by E_i/λ. The parameter is E_o/E_i.

Since avalanche diodes must work over a range of temperatures, it is important to understand what happens to the ionization coefficients as the temperature is varied. At higher temperatures, the increased density of phonons shortens the mean free path, requiring a higher electric field to achieve the same ionization probability. Experimental results agree with this picture and, in most materials, show an increase in breakdown voltage with increasing temperature, in marked contrast to other mechanisms that cause leakage current in reverse biased diodes. This behavior is so characteristic of impact ionization that it is often used to distinguish true avalanche breakdown from other leakage mechanisms such as Zener breakdown. The positive temperature coefficient is fortunate because it prevents thermal run-away, which could destroy the device or lead to undesired gain nonuniformities.

A fundamental assumption behind the notion of the ionization coefficient is that the probability of ionization is independent of the past history of the initiating carrier. This is obviously approximate since a carrier will be unable to initiate an impact ionization event if it has just lost most of its energy in a prior event. As long as the fields are low enough that the ionization probability per unit mean free path is much less than one, the ionization coefficient is a useful concept. While this condition is often met in practical avalanche diodes, it is not always met and one must be careful in extending the conventional theory of avalanche diodes without modification into this domain.

For an extensive review of the physics of impact ionization, see the article by Capasso (15).

CURRENT TRANSPORT AND AVALANCHE GAIN

To calculate the current flow within an avalanche diode, it is first necessary to solve Poisson's equation to obtain the electric field distribution. In principle, this should be done self-consistently with the equations that determine the density of carriers, taking account of the Fermi–Dirac statistics and carrier diffusion properties. In practice, avalanche diodes are typically biased at voltages that are high enough that the *depletion approximation* is accurate. Here one assumes that the region near the *pn* junction is completely depleted of mobile carriers, thus exhibiting a charge density equal to the density of donors and acceptors, usually assumed to be completely ionized. Outside this depletion region, the semiconductor is electrically neutral. When biased near breakdown, the charge due to the flowing current can become important and lead to nonlinearities such as gain saturation in APDs or negative differential resistance in IMPATTs. In the absence of these effects, one merely solves Poisson's equation within the depletion region, and adjusts the positions of the depletion edges in order to obtain a self-consistent solution. In general, Poisson's equation must be solved in a two or three-dimensional geometry. However, most of the important behavior of these devices can be understood in one dimension. Since many practical devices have lateral dimensions much greater than their vertical dimensions, a one-dimensional analysis is typically an excellent approximation within the active region of the device.

Poisson's Equation

The solution to Poisson's equation in one dimension is

$$\mathscr{E}(x) = \frac{q}{\epsilon}\int^{x} \rho(x')\,dx' \tag{1}$$

Here $\mathscr{E}(x)$ is the electric field at a point x, $\rho(x)$ is the charge density, and ϵ is the permittivity of the semiconductor. In n-type material, $\rho(x)$ is positive and equal to the density of ionized donors. Equation (1) is easily evaluated when the structure consists of layers with uniform doping as in Fig. 1. The field is a piecewise linear function that is zero at the edges of the depletion region. The sign and magnitude of the charge density in each region determine the slope. The area under the curve gives the voltage drop across the diode according to

$$V = \int_0^W \mathscr{E}(x)\,dx - V_{\mathrm{bi}} \tag{2}$$

where W is the width of the depletion region and V_{bi} is the built-in voltage, which is comparable to the bandgap voltage and usually much smaller than the bias voltage. Since the avalanche coefficients are strong functions of the electric field, an important parameter is the peak value of the electric field $\mathscr{E}_{\mathrm{p}}$, which occurs at the pn junction itself.

Current Transport Equations

Once the field is known, the current flow can be calculated using the following partial differential equations for the hole and electron current:

$$\frac{1}{v_p}\frac{\partial J_p}{\partial t} = -\frac{\partial J_p}{\partial x} + G(x,t) \tag{3}$$

$$\frac{1}{v_n}\frac{\partial J_n}{\partial t} = \frac{\partial J_n}{\partial x} + G(x,t) \tag{4}$$

Here, J_n and J_p are the electron and hole current densities and $G(x, t)$ is the net rate of charge generation (or recombination). v_n and v_p are the electron and hole drift velocities, which depend on the field and thus on the position x. In an avalanche diode, the most important contributions to the generation current are those due to impact ionization, the photogeneration due to light absorbed in the device, and the various undesired mechanisms that lead to dark current.

Low-Frequency Gain Calculation

The random nature of the avalanche generation current makes Eqs. (3) and (4) difficult to solve without simplifying assumptions. A great deal can be understood by looking at the time independent or steady-state solution. An important approach, which forms the basis of much of the analysis of avalanche diodes, is to take an ensemble average of both sides of Eqs. (3) and (4), thus interpreting these as equations for the average currents.

Setting the time derivatives to zero, the steady-state equations for the average currents in the presence of nonzero avalanche coefficients can be written as

$$-\frac{dJ_p}{dx} = \alpha(x)J_n + \beta(x)J_p + G_0(x) \tag{5}$$

$$\frac{dJ_n}{dx} = \alpha(x)J_n + \beta(x)J_p + G_0(x) \tag{6}$$

where $G_0(x)$ is the part of the generation current that is not due to avalanche effects and α and β are the ionization coefficients of the electrons and holes. By subtracting Eqs. (5) and (6), the total current density $J_{\mathrm{T}} = J_n + J_p$ is seen to be constant. Eliminating J_p, Eqs. (5) and (6) reduce to a single first-order linear differential equation:

$$\frac{dJ_n}{dx} = (\alpha - \beta)J_n + \beta J_{\mathrm{T}} + G_0 \tag{7}$$

By introducing the integration factor $\exp[\int_0^x (\alpha - \beta)\,dx']$, the solution is readily found. The total current is

$$J_{\mathrm{T}} = \int G_0(x)M(x)\,dx \tag{8}$$

where $M(x)$ represents the average avalanche gain for an electron–hole pair introduced at a position x within the avalanche region. $M(x)$ can be written in terms of integrals as (16–20)

$$M(x) = \frac{\exp\left\{-\int_0^x [\alpha(x') - \beta(x')]\,dx'\right\}}{1 - \int_0^W \alpha(x')\exp\left\{-\int_0^{x'} [\alpha(x'') - \beta(x'')]\,dx''\right\}dx'} \tag{9}$$

There are two important limiting cases of this equation. Substituting $x = 0$ gives the gain M_n for an electron injected at the left-hand side of the depletion region. Setting $x = W$ gives the gain M_p for a hole injected at the right-hand side. Note that in either of these limiting cases, Eq. (9) can be written without the complicated numerator. Since it is advantageous to inject only the most ionizing carrier into the avalanche region, the gain of an optimized avalanche diode will be given by the greater of M_n or M_p. Typically, dark current mechanisms will lead to mixed injection so it is useful to have the more general expression above for $M(x)$ at all values of x within the avalanche region.

While Eq. (9) is often quoted in the literature and is useful as a closed-form expression for the avalanche gain, it is not the best formulation for numerical work. As written, it is a double integral requiring a double iteration loop, if naively

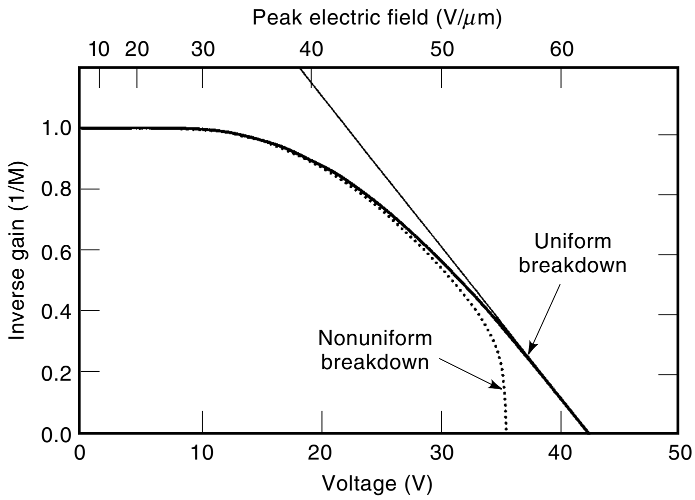

Figure 7. Inverse gain curve for an InP diode. Equation (11) is a straight line that provides an excellent approximation for gains above 2. The dotted curve shows the same diode with a small nonuniformity as described in the text.

implemented. A better approach that requires only a single loop is to start with the differential equation for the electron current, Eq. (7), and numerically integrate in the direction opposite to the current flow (21). As a boundary condition, J_n and J_T can be set to unity at the n side of the diode, yielding an electron current of $1/M_n$ at the p depletion edge. Standard methods can be used to perform the numerical integration. The same approach is useful in calculating other quantities such as the excess noise and gain–bandwidth product. The procedure does not diverge, since the value of J_n simply integrates to zero at breakdown.

A great simplification of the gain equation is obtained by assuming that the ratio of ionization coefficients ($k = \alpha/\beta$) is constant throughout the avalanche region. While this is an oversimplification, it yields analytical results that are useful in developing an intuitive feeling for the device behavior. McIntyre (22) has analyzed this case in detail. The electron gain M_n is easily shown to be

$$M_n = \frac{1-k}{\exp\left[(k-1)\int_0^W \alpha\, dx\right] - k} \tag{10}$$

with a similar result for M_p. M_n and M_p are related by the seldom stated result that $(1 - k) = (M_n - 1)/(M_p - 1)$. Equation (10) leads to especially simple results for $k = 0$, $M_n = \exp(\int \alpha\, dx)$, and for $k = 1$, $M_n^{-1} = 1 - \int \alpha\, dx$.

Dependence of Gain on Voltage

With this background, it is straightforward to calculate the current–voltage characteristic of an avalanche diode. For a given structure and applied voltage, the electric field is calculated using Eqs. (1) and (2). With the known values for the electric field, the avalanche coefficients and generation current are calculated for each point within the diode. The total current is then calculated using Eqs. (8) and (9).

A complete treatment of I–V characteristics would require an examination of all important sources of leakage current, a topic well beyond the scope of this article. Instead, the photocurrent I–V characteristics will be emphasized. Figure 4 shows the calculated normalized photocurrent (gain) versus voltage for several hypothetical diodes with constant k ratio. The curves were calculated for a one-sided abrupt p^+n diode in InP, using experimental values for β and setting $\alpha = k\beta$. The peak value of electric field is also shown. For nonzero k, there is a voltage at which the current theoretically goes to infinity. This breakdown voltage occurs at the bias point at which the denominator is equal to zero in Eq. (9) or (10) and is the result of the feedback between the two carrier types. Note that as k gets larger, the breakdown becomes more and more abrupt, with the avalanche gain depending sensitively on the bias, one of many reasons that it is desirable to have low k ratio.

Several equations have been used to make an approximate fit to the voltage dependence of the avalanche gain (9,16,23,24). The shape of the curve is especially important near breakdown where the device is operated. For finite k, Shockley (9) and Muehlner (24) have both noted that the denominator in Eq. (9) goes smoothly through zero and should be well approximated by a Taylor expansion near the breakdown V_b, leading to the following simple expression:

$$M = \frac{A}{V_b - V} \tag{11}$$

The coefficient A has units of voltage and establishes the scale over which breakdown occurs. Shockley and Muehlner suggested plotting the reciprocal of the gain as a way of observing localized breakdown and other deviations from the expected linear behavior. Figure 7 shows the so-called *inverse gain* calculated for an InP p^+n diode. Equation (11) accurately models the gain for all gains greater than 2. The dotted line is the inverse gain for the same diode except that 5% of the light falls on an area that has a slightly lower breakdown voltage. This leads to an abrupt change in the A coefficient at high gains. It is easy to spot this change on the inverse gain characteristic, but it would be difficult to see it on a conventional plot of gain versus voltage. A sudden change or lower-than-expected value for A is a sign of premature breakdown.

The breakdown voltage can be calculated by finding the bias voltage at which the denominator in Eq. (9) goes to zero. Numerically, this involves evaluating the inverse gain as a function of bias and iterating to find the zero. Figure 8 shows the result for uniformly doped one-sided abrupt junctions in

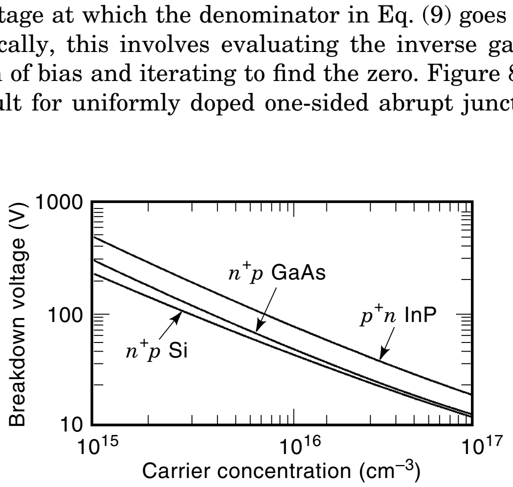

Figure 8. Calculated breakdown voltage versus doping for one-sided abrupt diodes in Si, GaAs, and InP using the ionization coefficients of Fig. 2. The type of the low doped side was chosen so that the predominant carrier has the higher ionization coefficient.

Si, GaAs, and InP as a function of the doping level on the lower doped side. Because of the rapid increase in avalanche coefficients with field, the peak field rises only weakly with the doping level. The depletion width decreases rapidly with doping, leading to a rapidly decreasing breakdown voltage as the doping is increased. Several empirical formulae have been proposed for estimating the avalanche breakdown voltage (11). In practice, it is often more useful to perform the numerical calculation.

AVALANCHE MULTIPLICATION NOISE

The physical effects discussed above are important for any type of avalanche diode. The multiplication noise is of special importance for avalanche photodiodes, since their primary use is to improve the signal-to-noise ratio of optical receivers. This noise arises from the statistical nature of the multiplication process. Each time a primary carrier initiates an avalanche chain, the gain and pulse shape will be unpredictable. Again, averaging over an ensemble of events gives useful results. When we let M_i represent the gain for the ith event in the ensemble, the two most important parameters that characterize the noise are the first and second moments of the distribution of M_i, expressed by the average gain $M = \langle M_i\rangle$, and the mean-square gain $\langle M_i^2\rangle$. At frequencies that are well below any characteristic response frequency of the diode, the spectral density of the noise is white. The statistics of the noise are nongaussian unless the avalanche is initiated by a large number of primary carriers (25,26). This latter fact is usually ignored in analyzing the performance of APD receivers because of the difficulty of analyzing non-Gaussian noise sources. In the white noise part of the spectrum, it is straightforward to compute the power spectrum of the current fluctuations in an APD. The result is

$$S_i = 2eI\langle M_i^2\rangle = 2eIM^2F(M), \qquad \text{where } F(M) \equiv \frac{\langle M_i^2\rangle}{\langle M_i\rangle^2} \tag{12}$$

This is identical to the expression for the shot noise in a diode, except for the factor $M^2F(M)$. If the APD were an ideal amplifier, it would have a gain M_i that was always equal to M. Since $\langle M_i^2\rangle$ would equal $\langle M_i\rangle^2$, the factor $F(M)$ would be one and the noise spectral density would be larger than that of a conventional diode by a factor M^2. This is exactly the shot noise that would be expected for a diode in which the individual "shots" are larger by a factor of M. For a real device with gain fluctuations, $F(M)$ is always larger than one and represents the factor by which the noise is greater than shot noise. Appropriately, it is called the *excess noise factor.*

Composition Law for Avalanche Regions

In predicting the behavior of avalanche diodes, it is helpful to break the device into smaller regions that can be analyzed separately. If a gain region is broken into two parts with electron gains M_{1n} and M_{2n}, and hole gains M_{1p} and M_{2p}, the overall electron gain is just

$$M_n = \frac{M_{1n}M_{2n}}{1-(M_{2n}-1)(M_{1p}-1)} \tag{13}$$

when electrons travel from region 1 to region 2. Using similar notation, the excess noise factor for the composite gain region can be expressed in terms of the excess noise factors for the subregions (21,27):

$$F_n = F_{1n} + \frac{M_{1p}^2 M_n}{M_{1n}^2 M_{2n}}[(F_{1p}-1)(M_{2n}-1)+F_{2n}-1] \tag{14}$$

These expressions are useful in analyzing some of the more exotic multilayer APDs that have been proposed. For example, they can be used to show that when a diode consists of multiple identical stages, the excess noise factor of the composite structure always has the form $F - 1 = A(M - 1) + B(1 - 1/M)$, where A and B depend in a simple way on the properties of a single stage. Equations (13) and (14) can be applied to an infinitesimal gain region to obtain a pair of coupled differential equations for the gain M and F (21). These can be integrated to obtain the gain and noise for arbitrary continuous structures, reproducing the theory first derived by McIntyre (22). In the case of constant k ratio, the excess noise factor is independent of the distribution of gain within the device and is given by the celebrated McIntyre formula:

$$F = kM + (1-k)(2-1/M) \tag{15}$$

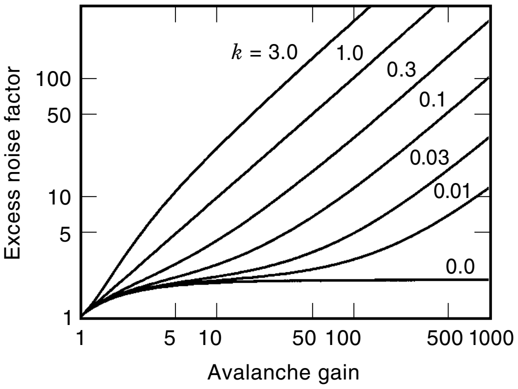

Figure 9. Excess noise factor versus gain for an avalanche diode with constant k ratio, for several values of k.

This equation is plotted in Fig. 9 for several values of k. For $k = 0$, the excess noise factor rises only slightly with gain, approaching $F = 2$ asymptotically as the gain goes to infinity. When k is nonzero, F rises linearly with gain and in direct proportion to k at high multiplication. Thus, it is desirable to keep k as low as possible by choice of materials, by design of the electric field profile, and by injection of the carrier with the highest ionization coefficient. For a real APD, the dependence of F on M will be different than that predicted by Eq. (15), since the k ratio will not be constant. It is straightforward to compute the excess noise factor by direct numerical integration of the differential equation for the excess noise factor. For sufficiently high gains, the excess noise will still rise in direct proportion to the gain as long as there is some ionization by the undesired carrier type. The proportionality constant in this limit is sometimes called the effective ionization ratio k_{eff}.

NONUNIFORM BREAKDOWN

To this point, the discussion has assumed that the breakdown is uniform over the active region of the diode. One of the biggest challenges in APD design and fabrication is to guarantee that the gain is highest and is uniform throughout the region where the photoinduced carriers flow. The consequences are severe if the gain is not uniform. If a small part of the diode has a breakdown voltage just slightly less than that of the active area, it may not be possible to get any useful gain in the active area at all. Even if the desired gain can be achieved, the undesired region will have a gain much larger than that of the active region and will contribute noise proportional to M^3 as evident from Eqs. (12) and (15). For example, if 10% of the light were falling on a region with a gain that was higher by a factor of 10, it would contribute about 100 times as much noise as that from the 90% of the light that experienced the desired gain. Even a small nonuniformity can lead to a significant increase in the noise. Most straightforward designs lead to undesired nonuniformities, often exhibiting breakdown at an edge. For example, a junction formed by diffusion through a mask into a more lightly doped substrate will have a concentration of electric field at the curved periphery of the junction. Another problem is the concentration of the electric field at the surface of the diode, where breakdown or high leakage often occurs at lower fields than in the bulk. Many schemes have been proposed and implemented for eliminating these gain nonuniformities, including guard rings, beveled junctions, implanted regions, and specially tailored doping profiles.

FREQUENCY RESPONSE

Since APDs are often used in high-frequency optical receivers, it is important to understand the physical effects that limit the frequency response. They are typically used to detect small signals, so it is usually accurate to treat them as linear devices. A complete description must then include both an equivalent circuit model and the frequency response or impulse response of all elements within the model. The dominant element in the equivalent circuit is the diode capacitance shunted by a source representing the photocurrent, whose frequency response depends on the applied optical signal and on various intrinsic properties of the photodiode. For a planar geometry, the capacitance is given by $C_d = \epsilon A/W$, where A is the area of the diode. To get the lowest capacitance, it is desirable to make the area as small as possible and to keep the depletion width as thick as other constraints allow. The circuit model must also include elements that model the capacitance and resistance of the contact, the inductance of the bond wire, and any other important parasitics.

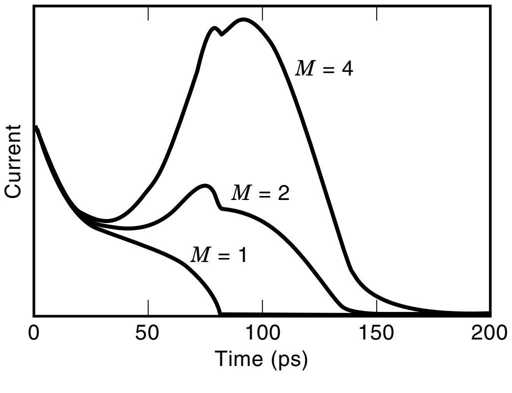

Figure 10. Response of the avalanche diode in Fig. 1 to an impulse of light incident from the bottom for gains of 1, 2, and 4.

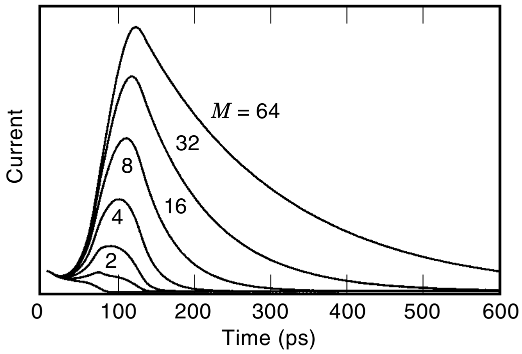

Figure 11. Same as Fig. 10 but showing the onset of an exponential tail at higher gains.

Several transit-time effects limit the frequency response of the photocurrent. These are best understood in the time domain. Figure 10 shows a calculation of the impulse response of an APD like that in Fig. 1 for three values of the avalanche gain. The simulated device is a so-called "separated absorption and multiplication" (SAM) APD (28–31) in which the absorption is predominantly in the i region of Fig. 1. The calculations are done by a matrix solution of the transport Eqs. (3) and (4), which again represent an ensemble average over many primary events (32). First, consider the behavior at a gain of one, when no avalanche multiplication occurs. The light is absorbed predominantly near the n side of the device, creating electron–hole pairs. Current flows in the electrodes whenever charge is moving within the depletion region (33). The current induced by each charge carrier is just qv/W, where q is the charge of the electron and v is the velocity of the carrier. At the initial instant, all the primary electrons and holes are flowing toward their respective electrodes and the current is at its maximum. The electrons reach the n electrode first, and the current falls to roughly half the initial value, where it remains until the holes reach the p electrode and the current drops to zero. The rounding-out of the pulse shape reflects the fact that not all the electron–hole pairs are created in the same position. Next, consider what happens at a gain of 2. In addition to the flow of primary carriers, which gives exactly the same pulse shape, the current due to secondary carriers must also be added. The secondary carriers are produced close to the pn junction as the holes reach the high-field region and initiate the avalanche. They produce a secondary pulse that begins near the end of the primary pulse and extends in time until the secondary electrons reach the n electrode. As the gain is increased to 4, notice that the secondary pulse grows in size.

Figure 11 shows what happens as the gain is raised to higher values. Notice that the secondary pulse develops an exponential tail, which gets longer as the gain is increased. In fact, the peak current tends to level off, with additional

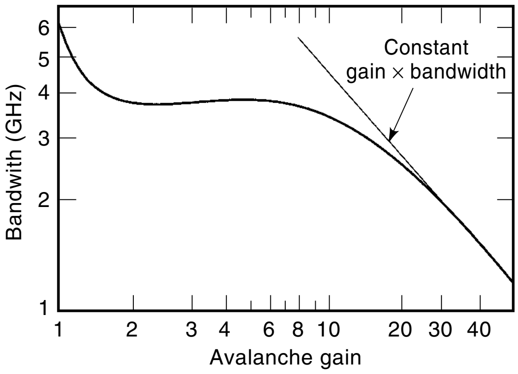

Figure 12. Bandwidth of the avalanche diode in Fig. 1 as a function of gain. At high gains, the curve approaches a line with constant gain–bandwidth product.

gain lengthening the pulse rather than increasing its height significantly. This behavior is caused by the finite transit time of carriers within the avalanche region. The linear relationship between gain and pulse width leads to an inverse relation between gain and bandwidth characterized by a constant gain–bandwidth product for sufficiently high gain. This is made clear by Fig. 12, which shows the calculated 3 dB bandwidth as a function of gain. The bandwidth is highest at a gain of 1 and drops significantly as the gain reaches 2 because of the lengthening of the pulse by the transit of secondary carriers. As the gain is increased beyond 2, the bandwidth actually improves as the secondary carriers dominate the response. Finally, the avalanche buildup effect predominates and the bandwidth falls off inversely with gain.

An analysis of the frequency response of an avalanche region with a uniform electric field and ionization ratio k has been done by Emmons (34). From his result, it is possible to derive a simple expression for the gain–bandwidth product:

$$Mf_{3\mathrm{dB}} = \frac{(1-k)^2 \ln k}{4\pi k \tau_0 [2(1-k) + (1-k)\ln k]} \qquad (16)$$

where $f_{3\mathrm{dB}}$ is the 3 dB bandwidth and $\tau_0 = (l/2)(1/v_n + 1/v_p)$ is the average transit time across the avalanche region of length l, with carrier velocities v_n and v_p. To achieve a high gain–bandwidth product, it is desirable to have a low k ratio and to operate the device at high fields so that the ionization coefficients are as high as possible, thus minimizing the effective length of the avalanche region. It is desirable to use a material in which both carriers have high saturated drift velocities.

SATURATION EFFECTS

Several effects cause the response of an APD to saturate at high current levels (20,35). These include thermal heating, series resistance, and space–charge saturation. All can be modeled approximately by incorporating a series resistance into the equivalent circuit. The effect of the series resistance is to drop the effective voltage bias, and hence the gain, at high currents. Consider, for example, the effect of the flowing charge carriers on the gain. At low current levels, the density of mobile carriers has a negligible effect relative to the donor and acceptor ions. At sufficiently high current, the mobile charge has to be included when integrating Poisson's equation to obtain the electric field profile. In a one-sided diode in which most of the avalanche multiplication takes place at one side of the depletion layer, it is simple to show that the space-charge of secondary carriers tends to lower the electric field just as if the voltage had been dropped by an amount δV given by $\delta V = IR_\mathrm{s}$, where the space charge resistance R_s is just $W^2/(2A\epsilon v)$, where v is the drift velocity of the secondary carriers (11). Similarly, an effective resistance can be calculated to account for the increase of breakdown voltage that occurs due to self-heating. This resistance is proportional to the junction thermal resistance and to the temperature coefficient of the breakdown voltage. The various saturation effects can be combined to give an overall effective resistance R_e. This resistance can then be used in combination with Eq. (11) to estimate the effect on gain:

$$M = A/(V_\mathrm{b} + IR_\mathrm{e} - V) \qquad (17)$$

RECEIVER SENSITIVITY

Most optical receivers are dominated by noise in the preamplifier that measures the photocurrent. An APD amplifies the current before it reaches the preamp and thus diminishes the effect of amplifier noise. In the process, it introduces its own noise caused by fluctuations in the gain and dark current. The analysis of dark current noise is complicated by the fact that not all of the primary carriers experience the full gain. Often, the biggest source of dark current is surface leakage that does not experience significant gain. The situation is usually modeled by defining a *primary* dark current I_p that receives the full gain, and an *unmultiplied* dark current I_u that receives a gain of 1. The total dark current is then $I_\mathrm{T} = MI_\mathrm{p} + I_\mathrm{n}$. When such a diode is coupled to an amplifier, the spectral density of fluctuations at the output can be expressed as equivalent fluctuations in the input photocurrent given by

$$S_\mathrm{i} = 2q[(I_\mathrm{s} + I_\mathrm{p})M^2F(M) + I_\mathrm{u}] + S_\mathrm{a} \qquad (18)$$

where S_a is the spectral density of the equivalent input current noise of the amplifier and I_s is the average signal current. Since the electrical signal power is proportional to $I_\mathrm{s}^2M^2$, the frequency-dependent noise-to-signal ratio at the amplifier output is proportional to

$$\frac{\text{Noise}}{\text{Signal}} \propto \frac{1}{I_\mathrm{s}^2}\left\{\frac{S_\mathrm{a}}{M^2} + 2qF(M)\left[I_\mathrm{s} + I_\mathrm{p} + \frac{I_\mathrm{u}}{M^2F(M)}\right]\right\} \qquad (19)$$

Typically, the first term in Eq. (19) predominates for low values of M. The noise-to-signal ratio improves with gain like $1/M^2$ until the first term becomes as small as the second. For higher gains, the excess noise factor $F(M)$ leads to an increase in the magnitude of the second term and a consequent degradation in the noise-to-signal ratio. The primary dark current I_p is unimportant as long as it is small compared to the signal current I_s. The unmultiplied dark current I_u is reduced in importance by a factor $M^2F(M)$, which can easily be as high as 500 at a gain of 10. If the frequency dependence of S_a, the shape of the optical pulse, and the response of the receiver filter are known, Eq. (19) can be used to calculate the sensitivity of a photoreceiver. The details of the frequency response lead to a set of weighting factors called *Personick integrals* (36–39), which can be used to express the result. When the dark current is small enough to neglect, the minimum average receiver power P needed to achieve a specified signal-to-noise ratio (or bit error rate) is given by

$$P = \frac{P_a}{M} + P_q F(M), \qquad \text{where } P_q = \frac{h\nu}{2\eta} Q^2 B \tag{20}$$

Here P_a is the amplifier-limited receiver sensitivity that would be obtained in the absence of quantum noise if a conventional photodiode with the same capacitance and quantum efficiency were used in place of the APD, and P_q is the quantum-limited sensitivity for this bit error rate. Figure 5 is a plot of Eq. (20) for a hypothetical receiver operating at 2.5 Gb/s with an amplifier-limited sensitivity P_a of -30 dBm and several values for the k ratio. For any finite k, an optimum gain M_o exists with sensitivity P_o somewhere between the amplifier limit and the quantum limit. As the k ratio decreases, the optimum gain moves to higher values and the sensitivity approaches an asymptote that is 3 dB higher than the quantum limit, reflecting the fact that, for $k = 0$, $F(M)$ approaches 2 at high gain. If the optimum gain is high compared to $1/k$, McIntyre's formula for the excess noise factor can be approximated as $F(M) = kM$, and the optimum gain and receiver power are

$$M_o = (P_a/kP_q)^{1/2} \quad \text{and} \quad P_o = 2(kP_aP_q)^{1/2} \tag{21}$$

The optimum receiver sensitivity is twice the geometric mean of the quantum limit and the amplifier limit, with an additional factor of $k^{1/2}$, again demonstrating the importance of a low k ratio.

ADVANCED APD STRUCTURES

Many advanced structures have been developed to improve and optimize the performance of APDs. An important variation is the separated absorption and multiplication (SAM) design in which different materials are used for the avalanche multiplication region and the optical absorption region (28–31). This is necessary, for example, in photodiodes that use an InGaAs absorption region lattice-matched to InP because the InGaAs material becomes leaky at fields high enough for significant multiplication. Wafer bonding techniques have also been used to combine the avalanche properties of one material with the optical absorption properties of another (40). Many devices have been made using superlattice regions to improve the noise behavior or response speed of APDs. Other structures such as the staircase APD (41) and the ballistic APD (42,43) have been proposed to yield nearly noise-free operation. A discussion of these and the many practical details of device fabrication is beyond the scope of this article.

SUGGESTED READING

For a more in-depth treatment of avalanche photodiodes, see the excellent review articles by Stillman and Wolfe (20) and by Müller (44). The article by Capasso (15) gives an extensive review of the physics of impact ionization. For a discussion of other types of avalanche diodes see the articles entitled DIODES FOR POWER ELECTRONICS, SURGE PROTECTION, and TRANSIT TIME DEVICES.

BIBLIOGRAPHY

1. K. G. McKay and K. B. McAfee, Electron multiplication in silicon and germanium, *Phys. Rev.,* **91**: 1079–1084, 1953.
2. K. M. Johnson, Photodiode signal enhancement at avalanche breakdown voltage, *1964 Int. Solid-State Circuits Conf., Dig. Tech. Papers,* 1964, pp. 64–65, also in *IEEE Trans. Electron Devices,* **ED-12**: 55, 1965.
3. L. K. Anderson et al., Microwave photodiodes exhibiting microplasma-free carrier multiplication, *Appl. Phys. Lett.,* **6**: 62–64, 1965.
4. W. N. Grant, Electron and hole ionization rates in epitaxial silicon at high electric fields, *Solid-State Electron.,* **16**: 1189–1203, 1973.
5. L. W. Cooke, G. E. Bulman, and G. E. Stillman, Electron and hole impact ionization coefficients in InP determined by photomultiplication measurements, *Appl. Phys. Lett.,* **40**: 589, 1982. The parameterization used in calculationg the results in this article were taken from G. E. Bulman's Ph.D. thesis.
6. G. E. Bulman, V. M. Robbins, and G. E. Stillman, The determination of impact ionization coefficients in (100) gallium arsenide using avalanche noise and photocurrent multiplication measurements, *IEEE Trans. Electron Devices,* **ED-32**: 2454–2466, 1985.
7. B. L. Kasper and J. C. Campbell, Multigigabit per second avalanche photodiode lightwave receivers, *J. Lightwave Technol.,* **LT-5**: 1351, 1987.
8. P. A. Wolff, Theory of electron multiplication in silicon and germanium, *Phys. Rev.,* **95**: 1415–1420, 1954.
9. W. Shockley, Problems related to *p-n* junctions in silicon, *Solid-State Electron.,* **2**: 35–67, 1961.
10. G. A. Baraff, Distribution functions and ionization rates for hot electrons in semiconductors, *Phys. Rev.,* **128**: 2507–2517, 1962.
11. S. M. Sze, *Physics of Semiconductor Devices,* New York: Wiley, 1981.
12. A.G. Chynoweth, Ionization rates for electrons and holes in silicon, *Phys. Rev.,* **109**: 1537–1540, 1958.
13. B. K. Ridley, Lucky-drift mechanism for impact ionisation in semiconductors, *J. Phys. C, Solid State Phys.,* **16**: 3373–3388, 1983.
14. J. S. Marsland, A lucky drift model, including a soft threshold energy, fitted to experimental measurements of ionization coefficients, *Solid-State Electron.,* **30**: 125–132, 1987.
15. F. Capasso, Physics of avalanche photodiodes, *Semicond. Semimetals,* **22** (D): 1–172, 1985.
16. S. L. Miller, Avalanche breakdown in germanium, *Phys. Rev.,* **99**: 1234–1241, 1955.
17. N. R. Howard, Avalanche multiplication in silicon junctions, *J. Electron. Control,* **13**: 537–544, 1962.
18. C. A. Lee et al., Ionization rates of holes and electrons in silicon, *Phys. Rev.,* **134**: A761–A773, 1964.
19. J. L. Moll, *Physics of Semiconductors,* New York: McGraw-Hill, 1964, p. 225.
20. G. E. Stillman and C. M. Wolfe, Avalanche photodiodes, *Semicond. Semimetals,* **12**: 29, 1977.
21. J. N. Hollenhorst, A theory of multiplication noise, *IEEE Trans. Electron Devices,* **ED-37**: 781–788, 1990.
22. R. J. McIntyre, Multiplication noise in uniform avalanche diodes, *IEEE Trans. Electron Devices,* **ED-13**: 164, 1966.

23. S. L. Miller, Ionization rates for holes and electrons in silicon, *Phys. Rev.,* **105**: 1246–1249, 1957.
24. D. J. Muehlner, private communication, ca. 1984.
25. S. D. Pesonick, Statistics of a general class of avalanche detectors with applications to optical communication, *Bell Syst. Tech. J.,* **50**: 3075, 1971.
26. R. J. McIntyre, The distribution of gains in uniformly multiplying avalanche photodiodes: Theory, *IEEE Trans. Electron Devices,* **ED-19**: 703, 1972.
27. S. Rakshit, N. B. Chakraborti, and R. Sarin, Multiplication noise in multiheterostructure avalanche photodiodes, *Solid-State Electron.,* **26**: 999, 1983.
28. K. Nishida, K. Taguchi, and Y. Matsumoto, InGaAsP heterojunction avalanche photodiodes with high avalanche gain, *Appl. Phys. Lett.,* **35**: 251, 1979.
29. S. R. Forrest, R. G. Smith, and O. K. Kim, *IEEE J. Quant. Electron.,* **QE-18**: 2040, 1982.
30. K. Yasuda et al., InP/InGaAs buried-structure avalanche photodiodes, *Electron. Lett.,* **20**: 158–159, 1984.
31. J. C. Campbell et al., High-performance avalanche photodiode with separate absorption "grading" and multiplication regions, *Electron. Lett.,* **19**: 818–820, 1983.
32. J. N. Hollenhorst, Frequency response theory for multilayer photodiodes, *J. Lightwave Technol.,* **LT-8**: 531–537, 1990.
33. W. T. Read, Jr., A proposed high-frequency, negative-resistance diode, *Bell Syst. Tech. J.,* **37**: 401, 1958.
34. R. B. Emmons, Avalanche-photodiode frequency response, *J. Appl. Phys.,* **38**: 3705, 1967.
35. H. Melchior and W. T. Lynch, Signal and noise response of high speed germanium avalanche photodiodes, *IEEE Trans. Electron. Devices,* **ED-13**: 829–838, 1966.
36. S. D. Personick, Receiver design for digital fiber optic communication systems, I, *Bell Syst. Tech. J.,* **52**: 843, 1973.
37. R. G. Smith and S. D. Personick, Receiver design for optical fiber communication systems, *Topics Appl. Physics,* **39**: 89, 1982.
38. J. N. Hollenhorst, Fundamental limits on optical pulse detection and digital communication, *J. Lightwave Technol.,* **LT-6**: 1135–1145, 1995.
39. R. G. Smith and S. R. Forrest, Sensitivity of avalanche photodetector receivers for long wavelength optical communications, *Bell Syst. Tech. J.,* **61**: 2929, 1982.
40. A. R. Hawkins et al., Silicon heterointerface photodetector, *Appl. Phys. Lett.,* **68**: 3692–3694, 1996.
41. F. Capasso, W.-T. Tsang, and G. F. Williams, Staircase solid-state photomultipliers and avalanche photodiodes with enhanced ionization rates ratio, *IEEE Trans. Electron Devices,* **ED-30**: 381, 1983.
42. J. N. Hollenhorst, Ballistic avalanche photodiodes: Ultralow noise avalanche diodes with nearly equal ionization probabilities, *Appl. Phys. Lett.,* **49**: 516, 1986.
43. R. P. Jindal, A scheme for ultralow noise avalanche multiplication of fiber optics signals, *IEEE Trans. Electron Devices,* **ED-34**: 301, 1987.
44. J. Müller, Photodiodes for optical communication, *Adv. Electron. Electron Physics,* **55**: 189, 1981.

JAMES N. HOLLENHORST
Hewlett-Packard Laboratories

AVALANCHE PHOTODIODE. See AVALANCHE DIODES.

AVIONICS. See AIR TRAFFIC.