
Geiger-Mode Avalanche Photodiodes for Three-Dimensional Imaging

Brian F. Aull, Andrew H. Loomis, Douglas J. Young, Richard M. Heinrichs, Bradley J. Felton, Peter J. Daniels, and Deborah J. Landers

■ We discuss the properties of Geiger-mode avalanche photodiodes (APDs) and their use in developing an imaging laser radar (ladar). This type of photodetector gives a fast electrical pulse in response to the detection of even a single photon, allowing for sub-nsec-precision photon-flight-time measurement. We present ongoing work at Lincoln Laboratory on three-dimensional (3D) imaging with arrays of these diodes, and the integration of the arrays with fast complementary metal-oxide semiconductor (CMOS) digital timing circuits.

LINCOLN LABORATORY HAS LONG been involved with the development of laser radar (ladar) systems, which, like their microwave counterparts, measure the flight time of transmitted pulses of radiation to determine the distance to an object of interest. Real-time acquisition of such *range* information is central to a wide variety of remote-sensing applications. The optical radar, because of its shorter wavelength, can resolve objects subtending a smaller angular field of view. This finer *cross-range* resolution enables us to measure the range to multiple points in a scene, thus acquiring a *ladar image*, or *three-dimensional (3D) image*. Clearly, 3D images complement conventional intensity images in terms of information content.

One approach to building an imaging ladar system is to add scanning optics to a single-point ladar system. The laser sends out multiple light pulses, each directed to a different point in the scene by the scanning mechanism, and each resulting in a range measurement obtained by using a single detector. The scanner, however, limits the speed of image acquisition and adds significantly to the weight, volume, power, and cost of the system.

For real-time imaging of moving targets, it is desir-

able to capture the 3D image with a single laser pulse. For small, lightweight, fieldable systems, the transmitter and receiver components must be compact and must consume modest power. Single-pulse ladar imaging is achieved by flood-illuminating the scene of interest and imaging the returning light onto an array of detectors, each with its own timing circuit to measure the range to the corresponding point in the scene. This process presents the additional challenge of achieving both detector sensitivity and speed. For resolved targets, the fraction of transmitted optical energy returning to the receiver falls off as the inverse square of the range. In a single-pulse ladar imager this energy is further divided among multiple detectors. Therefore, for many scenarios with interesting image sizes and target ranges, a particular detector circuit must time the arrival of a weak optical pulse. If a range precision of even a few centimeters is needed, the timing precision must be less than a nanosecond and the detection circuit bandwidth correspondingly high. High bandwidth, however, means high noise competing with the weak signal.

Lincoln Laboratory is developing 3D imaging systems based on two enabling technologies. The first is the diode-pumped solid-state microchip laser, in-

vented at the Laboratory. Passively *Q*-switched frequency-doubled Nd:YAG (neodymium-doped yttrium-aluminum-garnet) microchip lasers have been developed that produce very short (250 picosec) optical pulses at 532 nm, with pulse energies of 30 μ J. The microchip laser systems, including power supply, are very compact. This microchip laser fulfills the requirements for an imaging lidar transmitter: a small package that delivers many photons in a very short pulse. This technology has been presented in detail in a previous issue of the *Lincoln Laboratory Journal* [1].

The second enabling technology, which is the subject of the present article, is a lidar receiver based on arrays of *Geiger-mode* avalanche photodiodes (APDs) integrated with fast complementary metal-oxide semiconductor (CMOS) time-to-digital converter circuits [2–5]. Geiger mode, discussed in detail later, is a way of operating an APD so that it produces a fast electrical pulse of several volts amplitude in response to the detection of even a single photon. With simple level shifting, this pulse can trigger a digital CMOS circuit incorporated into the pixel. Single-photon sensitivity is achieved along with sub-nanosecond timing precision. Because the timing information is digitized in the pixel circuit, it is read out noiselessly. An integrated APD/CMOS array would also be a compact, low-power, all-solid-state sensor.

An Avalanche Photodiode Primer

An APD is a variation of a *p-n* junction photodiode. When a *p-n* junction photodiode is reversed biased, an electric field exists in the vicinity of the junction that keeps electrons confined to the *n* side and holes confined to the *p* side of the junction. When an incident photon of sufficient energy (>1.1 eV in the case of silicon [Si]) is absorbed in the region where the field exists, an electron-hole pair is generated. Under the influence of the field, the electron drifts to the *n* side and the hole drifts to the *p* side, resulting in the flow of photocurrent in the external circuit. The time integral of the current can be shown to be one electron charge. An electron-hole pair can also be thermally generated, resulting in leakage current, which is also called dark current because it is present even in the absence of incident light.

The physical process of electron-hole generation,

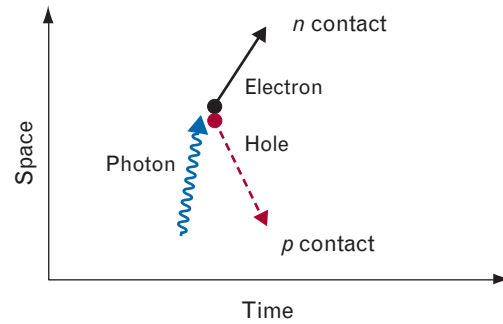


FIGURE 1. Photon detection in a photodiode represented in a simple space-time diagram. The absorption of the photon creates an electron-hole pair, and the two oppositely charged particles drift in opposite directions under the influence of the electric field in the vicinity of the reverse-biased *p-n* junction.

drift, and collection for a *p-n* junction photodiode is shown in Figure 1. The vertical axis represents the spatial dimension along the direction of the electric field (perpendicular to the junction plane); the horizontal axis is time. The electron trajectory is represented by a solid arrow, and the hole trajectory by a dashed arrow. The slope of a trajectory is the drift velocity (which, for high electric fields in Si, is about 10^7 cm/sec for both electrons and holes). When a photodiode is used to detect light, the number of electron-hole pairs generated per incident photon, a metric known as the quantum efficiency, is at best unity. Losses due to reflection or absorption in zero-field regions usually lower the quantum efficiency.

An APD detects light by using the same principle. The difference between an APD and an ordinary *p-n* junction photodiode is that an APD is designed to support high electric fields. When an electron-hole pair is generated by photon absorption, the electron (or the hole) can accelerate and gain sufficient energy from the field to collide with the crystal lattice and generate another electron-hole pair, losing some of its kinetic energy in the process. This process is known as impact ionization. The electron can accelerate again, as can the secondary electron or hole, and create more electron-hole pairs, hence the term “avalanche.” After a few transit times, a competition develops between the rate at which electron-hole pairs are being generated by impact ionization (analogous to a birth rate) and the rate at which they exit the high-field region

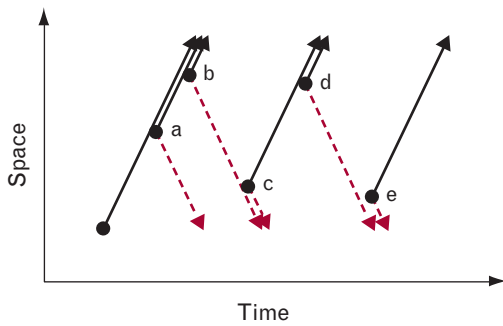


FIGURE 2. Avalanche multiplication illustrated in a space-time diagram. The primary electron (the companion hole is not shown), on the left, starts a chain of impact-ionization events. The solid arrows depict electron trajectories, and the dashed arrows depict hole trajectories. Points a, b, and d represent electron-initiated impact ionizations; points c and e represent hole-initiated impact ionizations.

and are collected (analogous to a death rate). If the magnitude of the reverse-bias voltage is below a value known as the *breakdown voltage*, collection wins the competition, causing the population of electrons and holes to decline. Figure 2 illustrates the avalanche-multiplication process in a space-time diagram.

This situation represents the most commonly known mode of operation of APDs: measuring the intensity of an optical signal and taking advantage of the internal gain provided by impact ionization. Each absorbed photon creates on average a finite number M of electron-hole pairs. The internal gain M is typically tens or hundreds. Because the average photocurrent is strictly proportional to the incident optical flux, this mode of operation is known as *linear mode*.

The amplification process adds noise to the signal above and beyond the amplified shot noise that would be projected from scaling up by a factor of M the photocurrent that would flow in a regular photodiode. Impact ionization is a statistical process. While M might be the *average* number of electron-hole pairs per absorbed photon, the actual number varies. This gain fluctuation produces *excess noise*, or *multiplication noise*, which gets progressively worse as the average gain M of a particular diode is increased by raising the reverse bias. Once the point is reached where the multiplication noise dominates over the noise introduced by downstream circuitry, further increases in gain deteriorate the system signal-to-noise ratio.

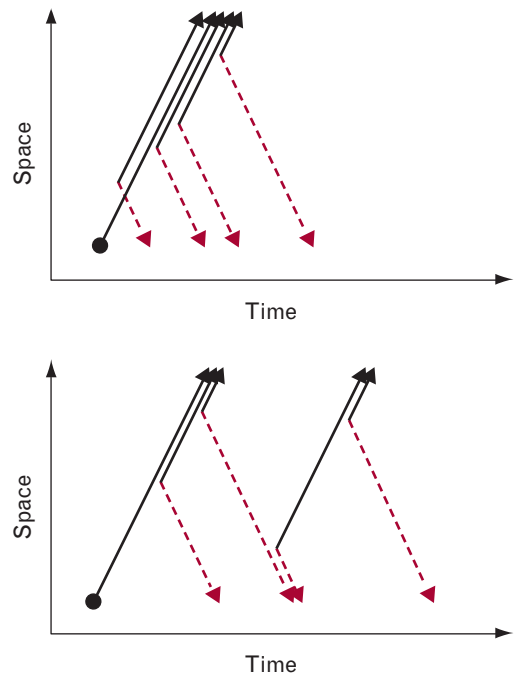


FIGURE 3. Two hypothetical materials. On the top, only electrons can initiate impact ionization. On the bottom, electrons and holes are equally likely to initiate impact ionization.

The severity of multiplication noise also depends on material properties. In general, electrons and holes are not equally likely to initiate impact ionization. In Si, for example, electrons are much more likely to impact ionize than are holes. For linear-mode operation, this property leads to low multiplication noise when the avalanche is electron-initiated. Figure 3 shows space-time diagrams for two *hypothetical* semiconductor materials. The one on the top represents a material in which holes are incapable of initiating impact ionization, and the one on the bottom represents a material in which electrons and holes are equally likely to impact ionize. Both diagrams indicate the same average value of gain M . In the top diagram, the multiplication process occurs in a single pass, with one electron entering the high-field region and M electrons exiting it. In this case, as the avalanche progresses, there are many electrons in the high-field region and therefore many impact-ionization events occurring in parallel. The law of large numbers works to reduce the variance of the gain. If one electron falls short of the average number of ionizations, another electron is

likely to exceed it. In the bottom diagram, on the other hand, a prolonged volley of alternating electron impacts and hole impacts achieves the gain. While the average duration of this process is assumed to lead to the same average gain as in the top diagram, at any one instant in time there may be only one or two carriers in the high-field region. The ionizations tend to occur sequentially rather than in parallel; therefore, the failure of an electron or hole to ionize can lead to the termination of the avalanche. Because termination can occur at any time during the volley, it leads to large statistical variations in gain. The avalanche prolongation also slows the response and limits the gain-bandwidth product of the detector.

What Is Geiger Mode?

Consider what occurs when the APD is reverse biased *above* the breakdown voltage by using a power supply that can source unlimited current. When the reverse bias exceeds the breakdown voltage, the electrons and holes multiply by impact ionization faster, on average, than they can be extracted. (This criterion, in fact, is the best definition of the avalanche breakdown voltage.) The space-time diagram in Figure 4 illustrates the concept. The population of electrons and holes in the high-field region and the associated photocurrent grow exponentially in time. The more above breakdown the APD is biased, the faster the growth time constant. This growth of current continues for as long as the electric fields in the device are negligibly altered by the presence of the growing current and the growing population of electrons and holes. If there is series resistance in the diode, however, more and more voltage is dropped across that resistance as the current grows. This effect reduces the voltage dropped across the high-field region, and therefore slows down the rate of growth of the avalanche. Ultimately, a steady-state condition is reached in which the voltage across the high-field region is reduced to the breakdown voltage, where the generation and extraction rates balance. At this point the current neither grows nor decays, and the series resistance provides negative feedback that tends to stabilize the current level against fluctuations. A downward fluctuation in current, for example, causes a decrease in the voltage drop across the series resis-

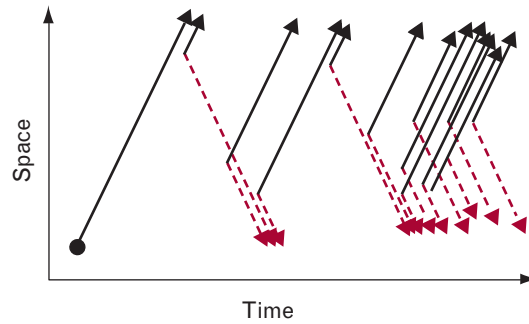


FIGURE 4. Concept of avalanche breakdown voltage. In Geiger mode, in which the avalanche photodiode (APD) is biased above the avalanche breakdown voltage, the growth in the population of electrons and holes due to impact ionization outpaces the rate at which they can be extracted, leading to exponential growth of current.

tance, and an equal increase in the drop across the APD high-field region, which in turn increases the impact-ionization rates and causes the current to go back up [6]. If the level of steady-state current is not too small (less than a few tens of microamps), it continues to flow indefinitely [7]. Therefore, an avalanche initiated by the absorption of a single photon causes the diode current to grow to some resistance-limited value. The turn-on transient of this current is fast, typically lasting tens of picoseconds.

When we measure the current-voltage characteristic of an APD on a curve tracer, the point of avalanche breakdown is evident as a sharp rise in current with a slope equal to the inverse of the device resistance. While the term “breakdown” is applied to this phenomenon, it is not a destructive phenomenon such as the dielectric breakdown that occurs when the field is strong enough to dislocate atoms in the material. Avalanche breakdown is stable against thermal runaway. As the device temperature rises, other kinetic-energy-loss mechanisms, such as lattice vibrations, increasingly compete with impact ionization, causing the breakdown voltage to rise and the steady-state avalanche current at a particular bias voltage to decrease.

Simply connecting an APD to a low-impedance power supply, however, gives no way to either detect the turn-on or shut off the avalanche so that the APD is ready to detect another photon. Shutting off the avalanche current is called *quenching*, and is accom-

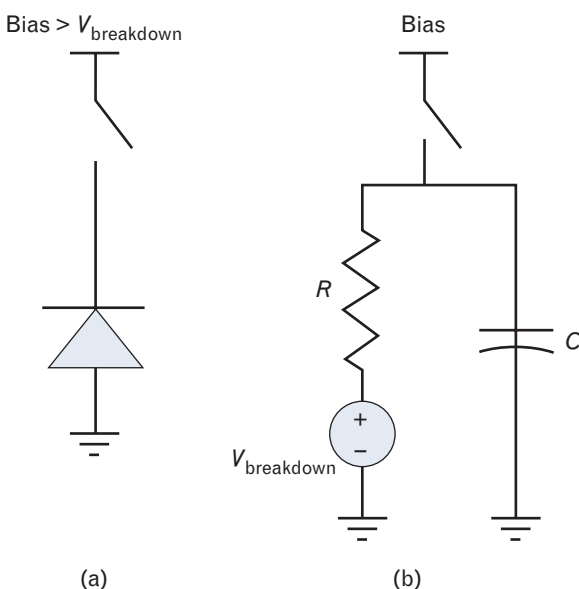


FIGURE 5. Passive-quenching circuits. (a) In Geiger mode, the APD is charged up to some bias above the breakdown voltage V and then left open circuited. (b) Subsequently, once an avalanche has been initiated, the APD behaves according to a simple circuit model.

plished by two types of circuit: *passive quenching* and *active quenching*. In a passive-quenching circuit, the APD is charged up to some bias above breakdown and then left open circuited. Once the APD has turned on, it discharges its own capacitance until it is no longer above the breakdown voltage, at which point the avalanche dies out. An active-quenching circuit senses when the APD starts to self-discharge, and then quickly discharges it to below breakdown with a shunting switch. After sufficient time to quench the avalanche, it then recharges the APD quickly by using a switch.

Figure 5(a) shows the simple passive-quenching circuit and Figure 5(b) shows the same circuit with a first-order circuit model inserted to describe the APD behavior during discharge. The model assumes that once the APD has turned on and reached its resistance-limited current, the ensuing self-discharge is slow enough that the APD will behave quasi-statically, following its DC current-voltage characteristic as it discharges down to breakdown. The corresponding model is a voltage source equal to the breakdown voltage in series with the internal resistance R of the APD. The model predicts exponential decay of the

current to zero and voltage to the breakdown with a time constant RC [8].

Once the avalanche has been quenched, the APD can be recharged through a switch transistor. Another scheme is to connect the APD to a power supply through a large series resistor R_s that functions as a virtual open circuit ($R_s \gg R$) on the time scale of the discharge, and then recharges the APD with a slow time constant $R_s C$. This circuit has the benefit of simplicity, and the APD fires and recharges with no supervision.

In lidar applications, where the APD detects only once per frame, the slow recharge time, typically microseconds, imposes no penalty. There is also interest, however, in using the Geiger-mode APD to count photons to measure optical flux at low light levels. With passive quenching, the count rate will saturate at low optical fluxes because many photons will arrive when the APD is partially or fully discharged, and therefore unresponsive. With a fast active-quenching circuit, the APD can be reset after each detection on a time scale as short as nanoseconds, enabling it to function as a photon-counting device at much higher optical intensities.

Geiger-Mode APD Performance Parameters

In linear mode the multiplication gain of the APD has statistical variation that leads to excess noise. In Geiger mode the concept of multiplication noise does not apply. A Geiger-mode avalanche can, by chance, die out in its earliest stages. If it does, no detectable electrical pulse is observed and the photon that initiated the avalanche goes undetected. If the avalanche progresses to completion, however, the total number of electron-hole pairs produced is fixed by the external circuit, not by the statistics of the impact-ionization process. In the simple passive-quenching case, for example, the avalanche has no further opportunity to die out until the APD has discharged from its initial bias down to the breakdown voltage. This discharge fixes the amplitude of the voltage pulse and, therefore, the total amount of charge collected in the process, typically $>10^7$ electron-hole pairs per detection event.

The user of a Geiger-mode APD is concerned not with multiplication noise, but with *detection probabil-*

ity, the probability that an incident photon will produce a detection event. This probability is the product of the quantum efficiency, which is the probability that the photon will be absorbed in the active region of the device, and the avalanche probability, which is the probability that the photoelectron (or hole) will initiate an avalanche that does not terminate prematurely. Figure 6 is a qualitative plot of the photon detection probability versus bias in Si for both electron-initiated and hole-initiated avalanches. The more the APD is biased above breakdown, the higher the avalanche probability and, therefore, the detection probability.

Another important difference between linear and Geiger modes is that a particular Geiger-mode detection event does not give intensity information. The electrical pulse produced by the detection of a photon is indistinguishable from that produced by the detection of many simultaneously absorbed ones. (In passive imaging applications, intensity is measured by counting multiple events, or by measuring the time that elapses before the first detection event after charging the APD.)

One noise source in a linear-mode APD is the shot noise due to multiplied dark current. In a Geiger-mode APD a single thermally generated electron or hole can initiate an avalanche, leading to an electrical pulse that is indistinguishable from a photon detection. In lidar applications, such an event represents a false alarm whose probability needs to be minimized. For photon counting to measure intensity, such events inflate the count. The average error can be measured and subtracted, but its statistical variance constitutes a noise source. The number of dark-current-induced events per second is the *dark count rate*. The use of high-quality material and proper device fabrication techniques minimizes the number of impurities and defects that contribute to dark current. Geiger-mode APDs have been made in Si with room-temperature dark count rates below 1000 counts per second. In a lidar, this means that the APD can be activated for 10 μ sec with a false-alarm probability less than 1%. A 10- μ sec detector active time corresponds to a range window of about 1.6 km (1 mile).

Since the electrical pulse from the APD is used in a lidar to measure the arrival time of an optical pulse,

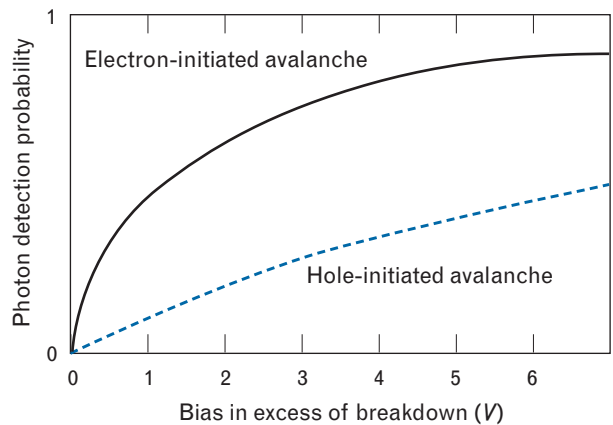


FIGURE 6. Qualitative plot of the photon detection probability versus bias in silicon (Si) for both electron-initiated and hole-initiated avalanches. The probability of photon detection is the probability that the incoming photon creates a photoelectron or photohole that initiates an avalanche that does not prematurely die out. In Si, electrons have a higher impact-ionization probability than holes, so the probability of detection is higher for electron-initiated avalanches than for hole-initiated avalanches.

the user must be concerned with the statistical variation of the time interval between the pulse arrival and the resulting electrical signal from the APD. Such *timing jitter* arises from several sources. First, the photon detection probability is generally less than unity and the transmitted optical pulse has finite time duration. For feeble returns the APD can detect a photon at the leading edge, the middle, or the trailing edge of the pulse. The associated statistical variation contributes to timing jitter [9].

Second, the photoelectron requires finite time to drift from where it is first generated to the high-field layer where the avalanche starts. These regions are often separated in the APD device structure. Depending on how deeply into the detector the photon propagates before it is absorbed, the photoelectron may have a shorter or longer drift delay. Because the spatial probability density for photon absorption is exponential, the depth at which the photon is absorbed has a standard deviation equal to the absorption length. The absorption length in Si monotonically increases with wavelength, so this contribution to the timing jitter also increases with wavelength. For a Si APD operating at 800 nm, for example, the absorption length is roughly 10 μ m. Because the satu-

rated carrier velocity (10^7 cm/sec) is the fastest possible drift velocity, this value translates to a minimum timing-jitter contribution of 100 psec.

The third contribution to the timing jitter arises from statistical variation in the time it takes the avalanche current to grow to its resistance-limited value. The space-time diagrams used previously to describe this current growth are oversimplified, because they neglect the fact that the avalanche starts as a localized filament of current that spreads laterally until it is uniformly distributed over the whole area of the junction. In diodes with a small active volume this spreading occurs predominantly by multiplication-assisted diffusion, and occurs over a time scale of tens of picoseconds. If the initial filament is near the diode periphery, the process takes longer than for a filament in the center because the distance over which the avalanche must spread is greater. In larger-volume diodes, a different mechanism predominates. The hot electrons and holes in the avalanche filament give off photons that have sufficient energy to be reabsorbed elsewhere within the diode, thereby initiating other filamentary avalanches. Avalanche buildup typically occurs over hundreds of picoseconds.

The fourth contribution to timing jitter comes from thermal noise due to APD resistance. This noise produces jitter because the APD is followed by circuitry that uses a thresholding operation to detect when the APD is discharging, and amplitude fluctuations result in fluctuations in threshold-crossing time.

In monolithic arrays of Geiger-mode APDs, the emission of photons by discharging devices introduces *crosstalk*; the photons from one APD can trigger avalanches in others. Measurements of the radiation from Si devices indicate that about 3×10^{-5} photons are emitted with energies greater than the bandgap for every carrier that crosses the junction. Above 1.7 eV the spectral distribution of this emission is a Maxwellian distribution with an effective temperature of 4000 K [10]. Because the number of photons emitted is proportional to the amount of charge traversing the junction during the discharge, the crosstalk can be minimized by minimizing the parasitic capacitance that the APD must discharge. Active quenching circuitry is also helpful, because it shunts most of the charge through a switch.

The most important performance parameters for a Geiger-mode APD used for photon timing are photon detection probability and timing jitter. For a photon-counting passive imager, dark count rate and spatial uniformity of detection probability are important, and jitter is not particularly important. For both applications, we want low crosstalk.

APD Fabrication and Characterization

Lincoln Laboratory has fabricated several lots of APDs on 4-in and 6-in Si wafers. The diodes are fabricated in 4×4 and 32×32 arrays, and there are also a number of test structures. Figure 7 shows a cross-sectional diagram of the basic APD device structure. The Si substrate is heavily *p*-doped (*p*+) with 10^{18} boron atoms/cm³ and overgrown by a lightly *p*-doped (10^{14} boron atoms/cm³) epitaxial layer. The diode is fabricated by ion implantation of *n*-type (arsenic and phosphorus) and *p*-type (boron) dopants. The structure that results, rather than being a simple *p-n* diode, is *p-π-p-π-n* (*π* denotes very lightly *p*-doped). The lower *π* layer is where the photons are absorbed when the device is illuminated from the substrate side, which is the intended mode of operation. Reverse-biasing at the proper operating voltage establishes a modest electric field (10^4 V/cm) in this photon absorption layer that causes the photoelectron to drift up into the upper *π* layer. The field in the upper *π* layer is much stronger (several 10^5 V/cm), sufficient to cause impact ionization that initiates an avalanche.

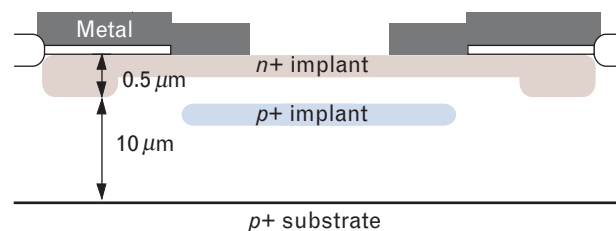


FIGURE 7. Cross-section of the basic APD device structure. The APD is fabricated in a lightly *p*-doped epitaxial layer grown on a heavily *p*-doped (*p*+) substrate. The heavily *n*-doped (*n*+) implant forms the *n*-side contact of the diode and the substrate form the *p*-side contact. The *p*+ implant creates a sheet of charge that segments the structure into regions of low electric field (below the implant), where photon absorption takes place, and high electric field (above the implant), where impact ionization occurs.

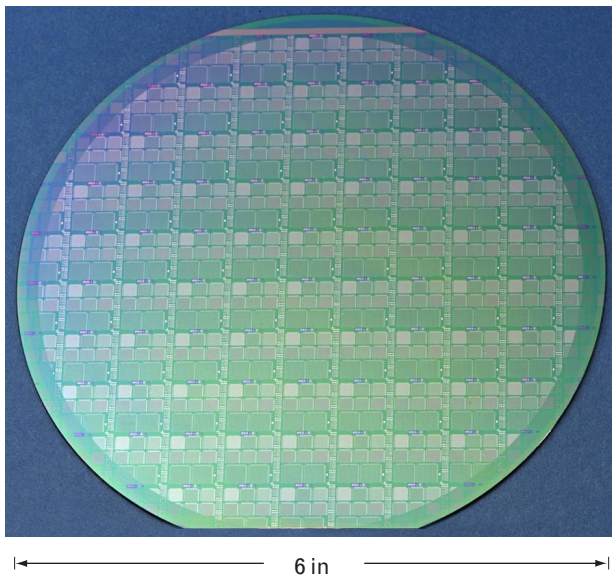


FIGURE 8. Avalanche photodiode arrays fabricated on a 6-in-diameter Si wafer in the Lincoln Laboratory Microelectronics Laboratory. Each die includes 32×32 APD arrays, 4×4 arrays, and test devices.

The photoelectron and the secondary electrons are collected at the top n layer, and the photohole and secondary holes are collected at the substrate.

Note that the $p\text{-}\pi\text{-}p\text{-}\pi\text{-}n$ structure exists only in the central portion of the APD. Because of the absence of a p implant in the diode's peripheral portion, this por-

tion is a simple $p\text{-}\pi\text{-}n$ structure, in which the field is intermediate in strength between the photon-absorption region and the avalanche region. This peripheral diode serves as a "guard ring" that performs two functions. First, it tailors the electric-field profile so that avalanche breakdown occurs in the central portion of the diode, not at the periphery. Second, it ensures that electrons generated outside the region directly under the p implant do not drift to the avalanche region, but are collected in the peripheral portion of the n layer without being able to start an avalanche. This collection minimizes the volume from which dark current is multiplied, and therefore minimizes the dark count rate. The price paid is that the fraction of the chip area sensitive to light is limited. In imaging applications that cannot waste photons, the efficiency needed can be reclaimed by using a microlens array to concentrate the incident light in each pixel's region so that a high percentage of it is detected.

Figure 8 shows a 6-in-diameter Si wafer on which APD arrays have been fabricated. There are 4×4 and 32×32 arrays with pixel-to-pixel spacings of $100 \mu\text{m}$ and $150 \mu\text{m}$. APD active area diameters also vary from $30 \mu\text{m}$ to $50 \mu\text{m}$. Figure 9 shows measurements of detection probability and timing jitter. Room temperature dark count rates of less than 1000 counts per second have been achieved in recent lots.

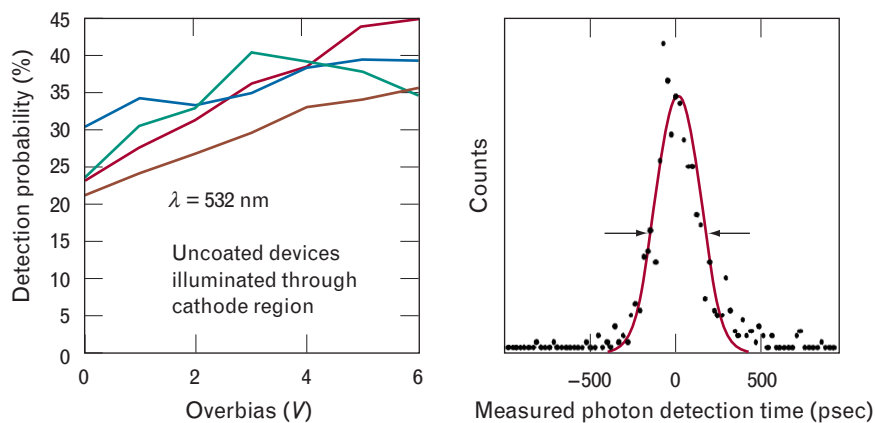


FIGURE 9. APD performance measurements. On the left is a plot of measured detection probabilities on uncoated, front-illuminated APDs from several different wafers. On the right is a histogram of measured photon detection times obtained by repeatedly illuminating an APD with a laser pulses of 250-psec duration, attenuated so that the detection probability is low (10%). The width of the histogram (290 psec) is due to a combination of the duration of the laser pulse and the timing jitter of the APD. The latter is inferred to be less than 150 psec.

Ladar Imaging

The development of APD-based ladar focal planes has occurred in three stages. In the first stage, 4×4 APD arrays have been packaged as discrete devices and incorporated into a printed circuit board with circuitry to amplify the pulses from the APDs. The amplified pulses are then conveyed by coaxial cabling to commercial rack-mounted timing modules. A ladar system, known as the Gen I brassboard, was developed by using this scheme. In the second stage, 4×4 APD arrays were hybrid integrated with CMOS chips that have sixteen timing circuits. In the third stage, 32×32 APD arrays were fully integrated with 32×32 arrays of CMOS timing circuits. The companion article in this issue describes the brassboard and the imagery obtained from it. Subsequent sections in this article describe the development of the CMOS timing circuitry and hybrid-integrated and fully integrated APD/CMOS imagers.

CMOS Timing Circuits

Using the MOSIS foundry service, we have prototyped arrays of time-to-digital converter pixel circuits. Figure 10 is a block diagram of the pixel circuit of the timing chip. It consists of a pseudorandom counter clocked by a 500-MHz master clock that is broadcast to all pixels. The pseudorandom counter is a shift register with a feedback path that has an exclusive OR gate; if the shift register is n bits long, it cycles through a sequence of $2^n - 1$ distinct states. The clock is fed to the counter through a transparent latch whose output is frozen when the photon is detected. The state of the counter thus encodes the number of clock cycles that elapsed from the start of counting to the photon detection time. The state of the transparent latch, which is also read out, indicates whether the photon was detected in the high-clock or low-clock portion of the clock period. Thus 1-nsec photon-flight-time resolution is achieved. An additional vernier bit is created by generating a secondary clock delayed by 90° with respect to the master clock and storing its state in a second transparent latch, and 0.5-nsec resolution is achieved. In ladar applications this approach yields a range resolution of 3 in (7.6 cm).

The pseudorandom count value must be decoded

by table lookup or reverse encoding after it is read out. The benefit of this implementation is that it gives a compact pixel circuit; a conventional binary counter would require much more chip real estate. The vernier bit scheme facilitates achievement of sub-nsec precision without requiring 2-GHz clocking. Therefore, we can use an established commercial foundry process with $0.5\text{-}\mu\text{m}$ or $0.35\text{-}\mu\text{m}$ gate geometries. A 16-pixel array that uses a 17-bit version of the counter circuit has been demonstrated by using the MOSIS HP $0.5\text{-}\mu\text{m}$ CMOS process, and 32×32 arrays of 15-bit and 10-bit counter circuits have been demonstrated by using the MOSIS TSMC (Taiwan Semiconductor Manufacturing Company) $0.35\text{-}\mu\text{m}$ process. Bench tests of the 32×32 CMOS chips confirm functionality at clock rates over 650 MHz. Figure 11 shows a portion of a 32×32 timing chip.

Figure 12 shows the device that was constructed to validate the operation of the CMOS/APD combination. A 4×4 APD array was bonded to a large pad in the middle of the 16-pixel MOSIS chip, and wire bonds were used to make an electrical connection from each APD to a corresponding timing circuit. The 16 timing circuits were daisy chained and read out serially. Timing values were observed both in response to dark counts and in response to

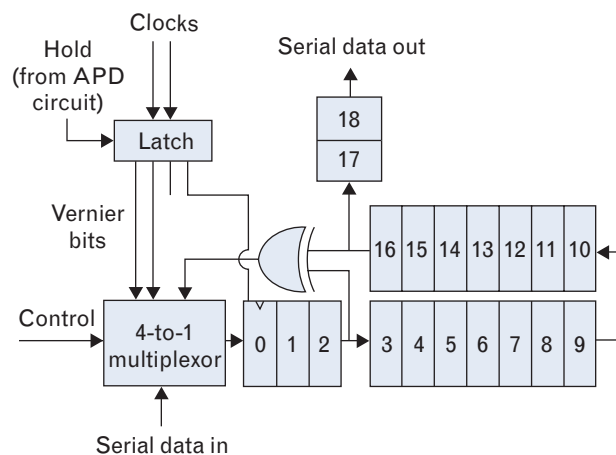


FIGURE 10. Block diagram of the pixel circuit of the timing chip, which functions as a “stopwatch” and times the detection of the photon by the APD. A 17-bit shift register with an exclusive OR (XOR) feedback loop is a pseudorandom counter. Photon detection causes the clocks that drive the counter to freeze. The counter and the stored states of the clocks encode the time of detection.

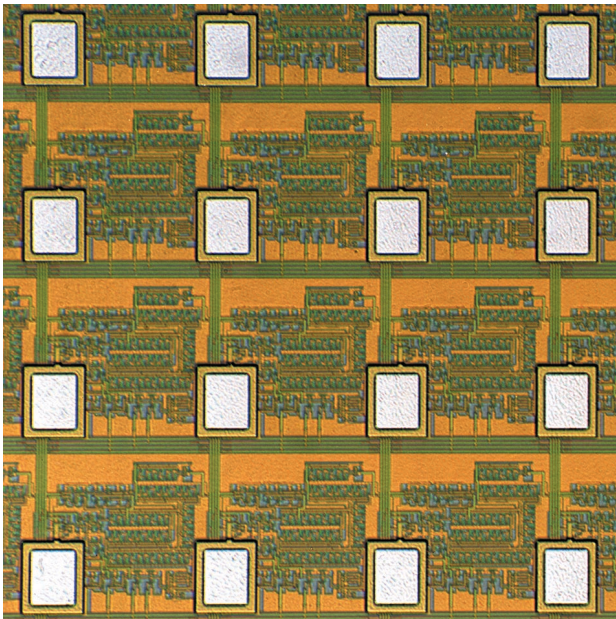


FIGURE 11. Portion of a 32×32 timing chip. This chip was fabricated through Mosis by using a complementary metal-oxide semiconductor (CMOS) process with $0.35\text{-}\mu\text{m}$ gate lengths. The white rectangle in each pixel is the bonding pad to which the corresponding APD connects.

optical pulses from a frequency-doubled passively Q-switched Nd:YAG microchip laser. The correctness of the timing values was validated. This chip was clocked through an off-chip source and worked up to a clock frequency of about 350 MHz. (A design flaw in the clock distribution circuitry limited the speed.) The level of pixel-to-pixel crosstalk was also checked by illuminating one APD with the laser pulses and monitoring the timing values from the pixel circuit attached to the neighboring APD. The timing values observed were random and showed no correlation with those from the illuminated pixel. Thus the crosstalk probability is low ($<1\%$).

APD/CMOS Integration

Clearly, as the array size is scaled up, the wire-bonding approach used for the 4×4 array becomes impractical. We can envision monolithic integration of APDs and timing circuits. This integration, however, presents two problems. The voltage requirements and fabrication process requirements of a Geiger-mode APD are quite different from those of a CMOS logic process. A standard CMOS process cannot be used,

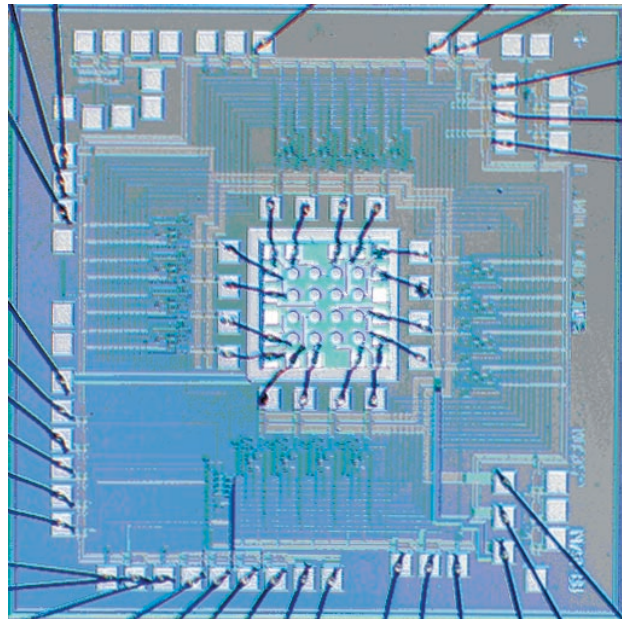


FIGURE 12. 4×4 APD/CMOS array, which is a "chip within a chip." The larger chip is a Mosis-fabricated timing chip with circuitry for 16 pixels, fabricated in a CMOS process with $0.5\text{-}\mu\text{m}$ gate lengths. The smaller chip epoxied in the center of the Mosis chip is a 4×4 APD array. The connections from the APDs to the timing circuits are made with wire bonds.

forcing costly development of a specialized process that might necessitate compromises between circuit performance and detector performance. Second, a monolithic process most likely involves a partition of the pixel real estate into detector and circuit sections, thus limiting the fill factor of the detector array. In the long run, silicon-on-insulator technology and three-dimensional integration techniques are likely to mature and yield monolithic implementations without these drawbacks. In the short run, our approach has been to use inexpensive foundry services to prototype the timing circuits, which are received as single unpackaged chips, and to pursue in-house fabrication of full wafers of APDs. The integration of the two is then achieved by bonding the chips face to face with the APD arrays, and illuminating the APD array from the substrate side (so-called back illumination).

A common technique, used for focal planes based on III-V or II-VI semiconductors, is to bump-bond the detectors with readout circuits by using arrays of indium (In) or solder bumps. For example, such "flip-chip" methodologies have been used for indium-gal-

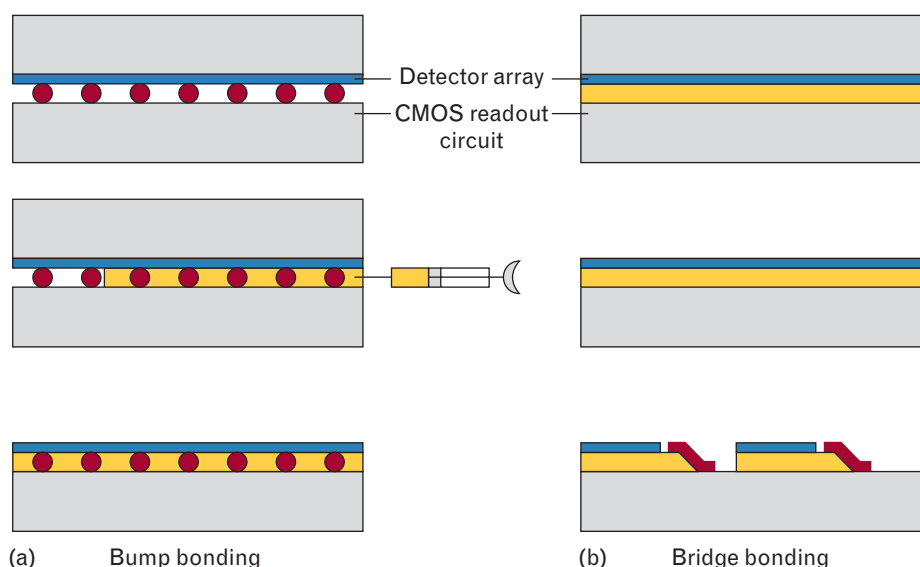


FIGURE 13. APD/CMOS integration. (a) In bump bonding (left), indium or solder bumps are formed on one or both chips and they are then pressed together. For Si detectors the air spaces between the bumps must be filled, so that when the substrate is removed to allow light to reach the detectors, the device is mechanically solid and sturdy. (b) In bridge bonding (right), the two chips are epoxied together, and the APD substrate removal is carried out. Electrical connections are made last by etching vias between the APDs and patterning metal connections in the vias.

lium-arsenide (InGaAs) photodiodes grown epitaxially on indium-phosphide (InP) substrates. In this case, nature is kind, providing a substrate material that is relatively transparent at the wavelength where the epitaxial detector is used. Flip-chip bonded devices can be constructed and used in back-illuminated mode without removing the substrate.

The Si APDs, however, are homoepitaxial devices. The Si substrate is optically opaque at the wavelengths where the detector is useful. Photoelectrons created in the substrate are almost immediately annihilated by holes and do not reach the high-field region in the epitaxial layer. A viable flip-chip bonding technique must therefore include removal of the substrate, leaving a detector structure that may be only a few microns thick. In such a bump-bonding process, mechanical sturdiness might be ensured by filling the spaces between the bumps before detector substrate removal, which is illustrated in Figure 13(a).

Questions and concerns about the scalability of this type of process to large array sizes lead us to develop an alternative process known as bridge bonding, which is illustrated in Figure 13(b) [11]. First,

the CMOS chip is epoxied face to face with the corresponding APD array without making any electrical connection whatsoever. The APD wafer, with several CMOS chips bonded to it, is epoxied, substrate up, to a handle wafer. The resulting APD/CMOS/handle “sandwich” can then undergo the same types of clean-room fabrication processes that are applied to normal full wafers. The APD substrate is removed by using an electrochemically selective etch developed at Lincoln Laboratory for fabrication of back-illuminated CCD focal planes. A shallow p^+ implant is added to the exposed backside of the APD wafer, followed by a laser anneal, in order to replace the p -side electrical contact layer formerly provided by the substrate. Vias are then etched in between the APDs and metal “bridges” patterned within the vias to connect each APD with the corresponding timing circuit. The backside of the APDs is also metalized to provide a common electrical contact to the p sides. In operation, this backside contact is held at a negative voltage slightly smaller in magnitude than the breakdown voltage, and the n side of each APD is charged up to a positive voltage (typically 4 or 5 V) by the pixel circuit.

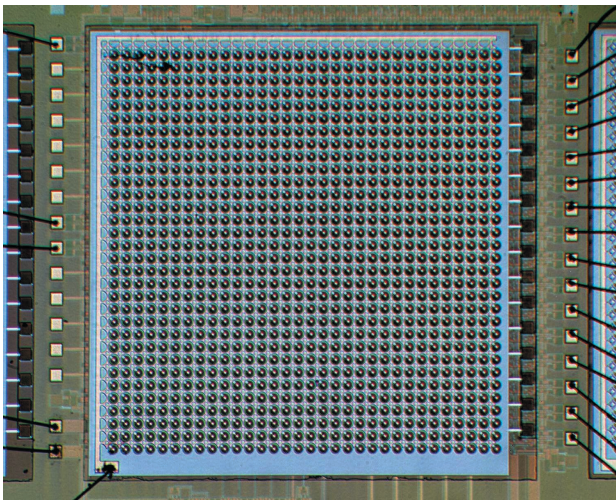


FIGURE 14. Photomicrograph of a bridge-bonded APD/CMOS device showing the 32×32 array.

Successful development of the bridge-bonding process required overcoming a number of technological hurdles. Thinning must be uniform. Curing of the epoxies used must not lead to destructive mechanical stresses due to the large thermal-expansion coefficient mismatch between semiconductors and epoxies. The vias through the epoxy must have sloped sidewalls to allow good step coverage of the bridge metal. Because of the vias, most of the required photolithographic steps are done on a nonplanar surface, which presents

challenges of nonuniform photoresist thicknesses and exposure depth of focus issues. Finally, the handling of the APD must not result in excess increases in leakage current or dark count rate.

After several iterations of the process, which initially used dummy CMOS chips and then functional ones, the first two bridge-bonded 32×32 APD/CMOS arrays were fabricated, and initial functionality tests performed on one of these arrays. Figure 14 is a photomicrograph of a 32×32 array and Figure 15 is a magnified view showing the features of individual pixels.

Initially, the device was tested by using spatially varying patterns of steady light, either room light or coronal light from a flashlight, projected onto the device. Illuminated APDs detected a photon very early during the timing interval, and the corresponding pixels reported out very low detection-time values. APDs that were masked from the light, in most cases, did not fire at all during the timing interval, and these pixels reported out timing values corresponding to the duration of the timing interval (250 nsec). The timing values were displayed as gray scale (black corresponding to immediate detection, light gray corresponding to 250 nsec), and therefore the displayed images are “negative tone” replicas of the illumination patterns. Figure 16 shows the images obtained in the

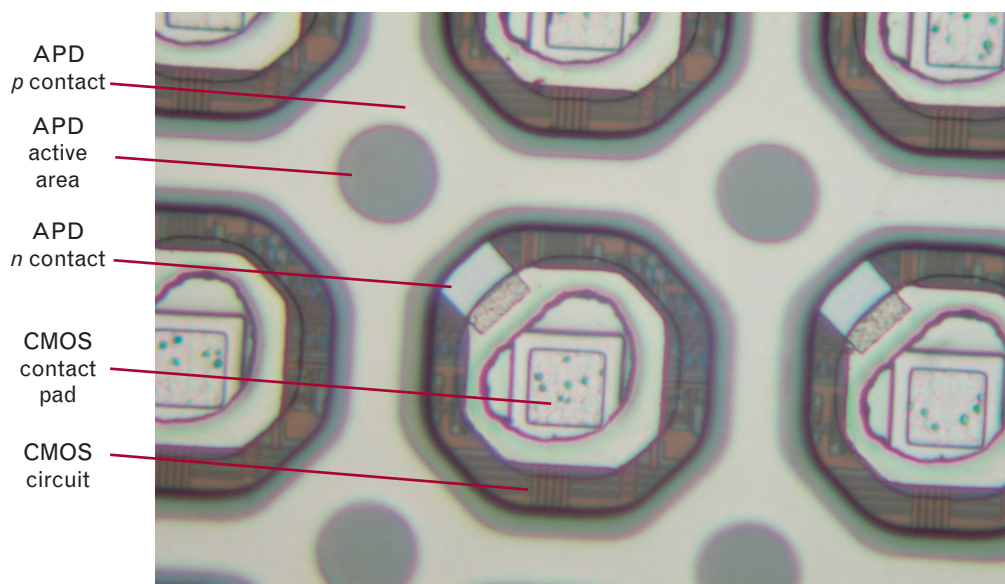


FIGURE 15. Photomicrograph of the same bridge-bonded APD/CMOS device shown in Figure 14, at much higher magnification, showing the features of individual pixels.

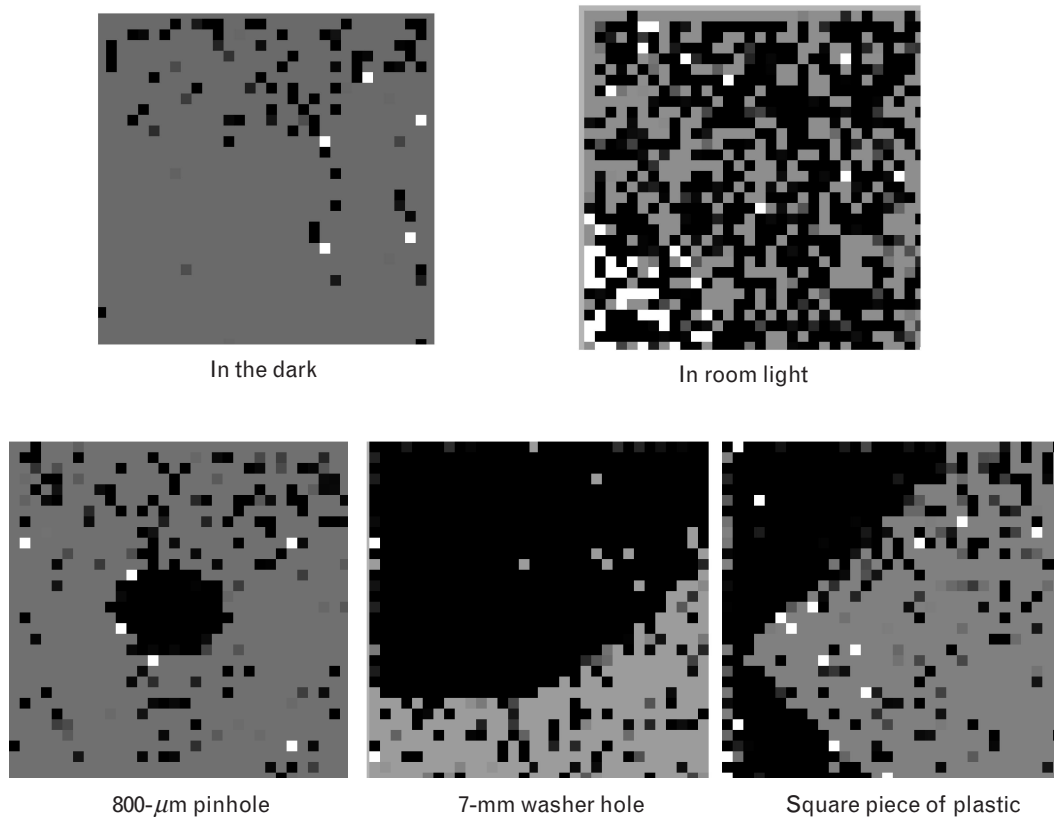


FIGURE 16. “Negative tone” intensity images from bridge-bonded APD/CMOS array obtained by illumination with different patterns of steady light. The black pixels represent small time values resulting from early detection events in the illuminated portions of the array.

dark, in room light, and with coronal flashlight illumination through a pinhole, through a washer hole, and around the edges of a square piece of black plastic. The white pixels represent values in excess of 250 nsec, which indicates that a bit error was made.

In this experiment the p side of the APDs was biased at about -23.5 V, a few tenths of a volt in magni-

tude below the avalanche breakdown voltage, and the CMOS circuit turned the APDs on by charging the n side to $+4$ V. The magnitude of the p -side bias on the APDs was decreased to more than a volt below the avalanche breakdown to confirm that the response of the APDs to the light was indeed Geiger-mode pulses. This decrease resulted in the disappearance of

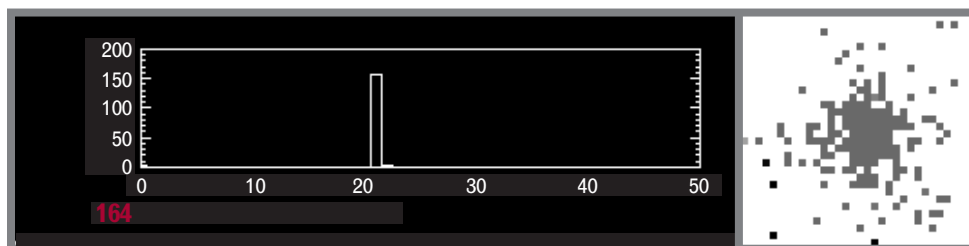


FIGURE 17. The range image on the right was obtained by illuminating the central portion of an APD/CMOS array with a sub-nsec pulse from an 800-nm-wavelength diode laser. The histogram on the left shows that 164 pixels reported detection events during the first 50 clock periods, and that nearly all of these were during the 21st period, which corresponds to the arrival time of the laser pulse.

the black pixels, indicating that the images are not due to linear-mode photocurrents.

A fast near-infrared diode laser ($\lambda = 800$ nm, pulse duration < 1 nsec) was used to illuminate the central portion of the device to verify its timing functionality. The pulse was timed to occur 60 nsec after the start of a 500-nsec timing interval. Figure 17 shows the resulting image, and a timing-value histogram. The dark gray area corresponding to the laser illumination is clearly visible in the center of the image. The horizontal axis of the histogram is elapsed clock periods, and the clock period used was about 3 nsec. The peak in the histogram corresponds to the time of the laser pulse. Work is in progress to characterize and optimize the detection probability and timing precision of these devices.

Conclusion

The work reported here represents only the initial stages in the development of a new and very powerful focal-plane-array technology. The images presented in the companion article illustrate the efficacy of 3D imaging for target recognition and image segmentation. APD/CMOS technology is clearly an enabling technology for 3D imaging. The next phase of the technology development effort will be to refine the performance of the imagers and achieve high detection probability, low dark count rate, and high spatial uniformity. If this phase is successful, we hope to scale up the number and density of the pixels and the time-measurement precision of the CMOS timing circuits.

Acknowledgments

This work was sponsored by the Department of the Air Force and the Defense Advanced Research Projects Agency.

REFERENCES

1. J.J. Zayhowski, "Microchip Lasers," *Linc. Lab. J.* 3 (3), 1990, pp. 427–46.
2. B.F. Aull, A.H. Loomis, J.A. Gregory, and D.J. Young, "Geiger-Mode Avalanche Photodiode Arrays Integrated with CMOS Timing Circuits," Mtg. of the Boston Chapter of the IEEE Electron Devices Society, Lexington, Mass., Nov. 1998.
3. B.F. Aull, "Geiger-Mode Avalanche Photodiode Arrays Integrated with CMOS Timing Circuits," *56th Annual Device Research Conf. Dig., Charlottesville, Va., 22–24 June 1998*, pp. 58–59.
4. B.F. Aull, "Geiger-Mode Avalanche Photodiode Arrays for Imaging Laser Radar," Solid State Research Report, Lincoln Laboratory (Mar. 1997), pp. 31–33.
5. B.F. Aull, "Geiger-Mode Avalanche Photodiode Arrays for Imaging Laser Radar," Solid State Research Report, Lincoln Laboratory (Mar. 1996), pp. 39–41.
6. There is another mechanism that acts like an additional series resistance. The electrons and holes generated by impact ionization require finite time to transit the device to their respective extraction electrodes. As the avalanche grows, electrons tend to "pile up" at the n side of the high-field region and holes at the p side. The electrostatic screening due to the associated space charge reduces the field in the high-field region. Since this reduction in field grows in proportion to the electron and hole population (and therefore in proportion to the current), the effect can be modeled as a series resistance.
7. At lower levels of steady-state current, the population of electrons and holes is sufficiently small that there is a significant probability that they will all transit out of the high-field region at some point in time without ionizing. The steady-state current level below which this phenomenon is observed is known as the *latching current*.
8. Strictly speaking, the voltage source needs to be slightly below breakdown so that the avalanche current decays at a rate that matches the RC time constant of the circuit. In practice, however, such circuit time constants are sufficiently long that this is a very small correction.
9. In addition to this random timing variation, there is a systematic dependence of the average detection time on pulse intensity. The more intense the returning pulse, the more likely detection will occur at its leading edge. This effect, known as *range walk*, or *range bias*, is not unique to Geiger-mode avalanche photodiodes, but occurs with any detection scheme in which thresholding is used.
10. A.L. Lacaita, F. Zappa, S. Bigliardi, and M. Manfredi, "On the Bremsstrahlung Origin of Hot-Carrier-Induced Photons in Silicon Devices," *IEEE Trans. Electron Devices* 40 (3), 1993, pp. 577–582.
11. A.H. Loomis and B.F. Aull, "Bridge Bonding of Geiger-Mode Avalanche Photodiode Arrays to CMOS Timing Circuits," Solid State Research Report, Lincoln Laboratory (Mar. 1999), pp. 33–37.



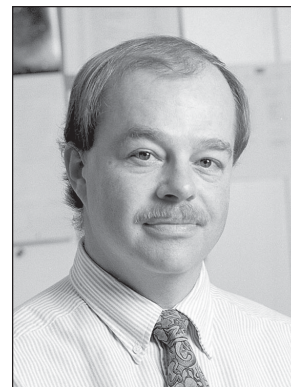
BRIAN F. AULL develops 3D imaging and photon-counting focal planes by using Geiger-mode avalanche photodiodes as a staff member in the Advanced Imaging Technology group. He joined the Laboratory in 1985, after earning a Ph.D. degree in electrical engineering from MIT. During his first decade at the Laboratory he led teams that demonstrated multiple-quantum-well spatial light modulators with gallium-arsenide/aluminum-gallium-arsenide (GaAs/AlGaAs) charge-coupled device (CCD) arrays for electrical addressing and developed optoelectronic switching and neural processing devices by using quantum-well modulators and double-barrier resonant tunneling diodes. During the past two years Brian has developed arrays of Geiger-mode avalanche photodiodes integrated with high-speed CMOS timing circuits that are being integrated into a 3D ladar system. He also holds a B.S. degree in electrical engineering from Purdue University.



ANDREW H. LOOMIS is an associate staff member with the Advanced Imaging Technology group. He supervises the production of specialized back-illuminated CCDs and Si sensors and develops advanced wafer-processing techniques for various Si imagers. His current research interests are in wafer bonding and specialized back-end CCD fabrication techniques. Before joining Lincoln Laboratory in 1986, he worked two years for at GCA corporation in Bedford, Massachusetts, developing laser scanning systems for defect identification. In 1984, Andy received a B.S. degree in electronic engineering from Wentworth Institute of Technology, Boston, Massachusetts.



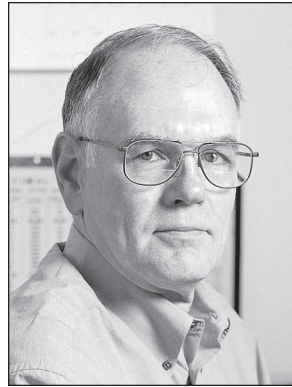
DOUGLAS J. YOUNG is an associate staff member of the Advanced Imaging Technology group and responsible for the fabrication of CCD imagers and photodiode arrays. He is also area engineer for i-line photolithography and etch areas of the Microelectronics Laboratory. Before joining the Laboratory in 1989, he was a technical instructor for MIT's Microelectronics Educational Facility. Doug holds a B.S. degree in electrical engineering from the University of Vermont.



RICHARD M. HEINRICHS leads the Laser and Sensor Applications group, which studies direct detection and coherent laser remote sensing. Rick joined Lincoln Laboratory in 1986, after completing a postdoctoral position in nonlinear fluid dynamics at the University of California, Santa Barbara. He holds an S.B. degree in electrical engineering from MIT and a Ph.D. degree in physics from the University of Massachusetts.



BRADLEY J. FELTON provides computer programming and CAD support for projects in the Advanced Imaging Technology group. Currently, Brad is working with the APD lidar development team to provide data acquisition and display routines. Brad joined the Laboratory in 1977. He has a B.S. degree in mathematics from the University of Lowell (formerly Lowell Technological Institute). He has also pursued graduate-level courses in mathematics and computer science at University of Lowell. Brad's other interests include circuit design verification and image processing.



PETER J. DANIELS supervises a microelectronic packaging facility in the Advanced Imaging Technology group. Pete holds a B.E.T degree in electrical engineering from Northeastern University. He is a member of the Sigma Epsilon Rho Honor Society. He joined the Laboratory in 1964.



DEBORAH J. LANDERS is a project technician for the Advanced Imaging Technology group. She packages Laboratory-processed semiconductor devices used in lidar ranging systems and in CCDs used in ground and space based astronomy. Before joining Lincoln Laboratory in 1984, Debbie packaged semiconductors for missile systems at Microwave Associates in Burlington, Massachusetts.