

## Single Photon Avalanche Diodes Enable Secure Communications

Mark A. Itzler, Princeton Lightwave Inc.

### Introduction

For centuries, one of the primary challenges faced by cryptographers had been the secure exchange of encryption keys. Once two parties possessed the same key for encrypting and decrypting messages, encoded messages could be readily transmitted. The problem had been finding a means of efficiently, yet securely, delivering a key from one party to the other.

The advent of public network encryption has revolutionized the exchange of secure information. Its essence is an algorithm by which a first party provides a public key with which any other party can encrypt a message; but decryption is only possible with a separate private key held by the first party. The shortcoming of such schemes is that they rely on computational complexity. For instance, to break the predominant encryption scheme in use today, one would require the ability to factor an extremely large number  $N$  (of the order of  $10^{300}$ ) into the two prime numbers that, when multiplied, give  $N$  as their product. Although this task appears exceedingly difficult today, this approach to data encryption could be compromised in the foreseeable future by the development of factoring methods based on either new mathematical algorithms or the emergence of sufficient computational power.

### Quantum Key Distribution

During the past two decades, researchers have devised truly unbreakable digital data encryption by exploiting the quantum mechanical properties of photons. Since it is not possible to measure the values of certain properties of a photon, such as its polarization or phase, without perturbing those values, one can encrypt transmitted data in such a way that any attempt to eavesdrop will fail. The application of this so-called “quantum cryptography” to the problem of securely distributing encryption keys has been dubbed “quantum key distribution” (QKD).

In a QKD system, each bit is represented by a single photon. The value of the bit is assigned using a particular quantum mechanical variable. The use of polarization is simplest to explain; see Figure 1. In choosing a particular value for the polarization of each bit, values are taken from one of two quantum mechanical “bases” represented by horizontal/vertical (H/V) polarizations and  $+45^\circ/-45^\circ$  polarizations. In each basis, one polarization value corresponds to a digital “1” and the other to a “0”. At point A, the sender (commonly referred to as Alice) randomly chooses a basis and prepares a photon with the correct polarization for her first bit value (“1” or “0”). At point B, the receiver (Alice’s compatriot Bob) also randomly chooses a basis and measures the first photon’s polarization. If Bob chooses the same basis used by Alice for preparing this particular bit, he will measure the correct bit value. If he chooses the wrong basis, quantum mechanics dictates that there is a 50% chance he will measure the intended bit value, with a complementary 50% chance that he measures the wrong value.

Alice continues to send her bits by assigning appropriate polarizations, and Bob continues to measure polarization values. Bob then communicates with Alice through an insecure channel to tell her which basis he used to measure each photon. Alice responds by telling Bob for which bits his basis choice was correct. These bits are then used as the encryption key.

QKD is absolutely secure because it is impossible for an eavesdropper (the nefarious Eve to cryptographers everywhere) to intercept the key. She must randomly choose a polarization basis for each photon measurement, just as Bob does. Even if she monitors the public discussion between Alice and Bob concerning their choices of basis, her random basis choices will not have been the same as Bob’s. Moreover, Eve’s situation is further disadvantaged: she can not intercept photons during the key distribution process without being detected. She can attempt to replace intercepted single photons, but without the correct polarization bases, her replacement photons will not be equivalent to Alice’s original photons. Eve’s meddling becomes apparent when Alice and Bob check

the validity of their transmitted key through error checking procedures, and the corrupted key is discarded.

During the past few years, the feasibility of QKD systems has moved beyond research lab demonstrations. Commercial point-to-point secure links employing QKD-based encryption are now available from at least two start-up companies (idQuantique in Switzerland, and MagiQ Technologies in the US), and several large Japanese corporations, including Mitsubishi, NEC, and Toshiba, have product-oriented programs for QKD system development.

### **QKD System Components**

As in traditional optical communications systems, optical sources and detectors are fundamental components of a QKD system. However, QKD requires the transmission and detection of single photons. The transmission of single photons currently limits QKD implementation to single span point-to-point links since there are not yet deployable technologies for single-photon repeaters that faithfully maintain the photon's quantum mechanical properties. Even in point-to-point links, the requirements for single photon transmission present considerable challenges for component design and performance.

An ideal single photon transmitter would be capable of deterministically emitting a single photon at a time. Although there are novel device structures with the potential for "on-demand" single photon emission (e.g., diode lasers with quantum dot active regions), a commercially viable source of this type has yet to be demonstrated. Instead, existing QKD systems use highly attenuated pulsed laser diodes with Poisson-distributed photon output. A primary concern is that the presence of two photons representing a single bit can potentially compromise the security of the QKD scheme: theoretically, Eve can strip off one photon without perturbing the second photon. Therefore, the pulsed laser output is attenuated so that the mean photon number per bit slot is on the order of 0.1. Although ~90% of the bit slots contain no photons, this inefficiency is tolerated so that the probability of having two photons per bit slot is well below 1%.

A principle limitation on current QKD link performance is the performance of available single photon detectors. Photomultiplier tubes have long been used to detect single photons, but their detection efficiency is extremely low, of order 1%. Sophisticated photon detectors employing superconducting devices have shown impressive performance, but their requirement for cryogenic operation at liquid helium temperatures makes them impractical for commercial network deployment. At present, the most promising detector technology for QKD systems is the single photon avalanche diode (SPAD) based on InGaAs/InP photodetector technology.

### **InGaAs/InP SPADs**

By virtue of the photoelectric effect, positive-intrinsic-negative (PIN) InGaAs/InP photodetectors can efficiently detect infrared light in the fiber optic communications wavelength range from 1300 – 1600 nm. However, their responsivity is limited to one electron-hole pair generated per input photon. Avalanche photodetectors benefit from internal gain due to a process known as impact ionization that leads to multiple electron-hole pairs per input photon. Applying a larger reverse voltage to the avalanche detector will result in a larger gain, until the breakdown voltage  $V_{br}$  is reached. For bias voltages larger than  $V_{br}$ , the electron-hole generation process can become self-sustaining and result in a run-away avalanche.

Traditional optical communications receivers make use of linear mode avalanche photodiodes (APDs), for which the output photocurrent is linearly proportional to the intensity of the optical input. Internal APD gain can provide significant improvement in high-bandwidth receiver sensitivity [see K. K. Loi and Mark Itzler, *Compound Semiconductor*, April 2000]. In contrast to linear mode operation below  $V_{br}$ , if an avalanche photodetector is biased above  $V_{br}$ , then a single photoexcited carrier can induce a run-away avalanche that gives rise to an easily detectable macroscopic current. In this case, the detector is sensitive to a single photon input and is referred to as a single photon avalanche diode (SPAD).

This mode of operation is often referred to as “Geiger mode” because of its similarity to Geiger-Muller detectors, in which particle emission from radioactive materials causes an avalanche of carriers from ionized gas atoms.

SPADs and linear mode APDs share many of the same device design elements. In Figure 2, we present a schematic cross-section of our design platform. The absorption layer consists of  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ , which has a useful responsivity for wavelengths as long as 1650 nm at room temperature. Since it is not possible to achieve high-field avalanche gain in InGaAs without also generating large leakage currents, avalanche gain is achieved in a separate larger bandgap InP multiplication region. To maintain high field in the multiplication region and low field in the absorption region, a moderately doped InP field control layer is used between them. Since a direct interface of InGaAs and InP tends to trap carriers, a “grading” layer is employed to smooth this interface.

Despite structural similarities, SPAD and linear mode APD design optimizations differ significantly. Thinner multiplication regions are desirable for reducing the noise of linear mode APDs because they result in more deterministic linear mode avalanche processes. However, SPAD performance benefits from thicker multiplication regions due to their larger probability of generating self-sustaining avalanches. In linear mode devices, it is generally required that the p-n junction depletion “punches” through the absorption region at low voltage so that high bandwidth can be achieved at low gains. However, since SPADs are intended to operate solely above  $V_{br}$ , the constraints of early punchthrough are absent. The SPAD design must also account for operation at temperatures of  $-60\text{ }^{\circ}\text{C}$  or lower to reduce the dark count rate. Finally, there is also a crucial distinction between the detection processes for the two types of devices: linear mode APDs are employed as analog devices, while the detection of SPADs avalanche pulses is inherently digital.

### **SPAD Performance Parameters**

In linear mode APDs, the probability that a photon will be absorbed in the absorption region and create a photoexcited carrier determines the quantum efficiency. In SPADs, there is the additional requirement that this carrier induce a detectable runaway avalanche. The product of the probabilities for these two processes is central to determining the single photon detection efficiency (SPDE), and a larger applied bias leads to higher SPDE.

Linear mode APD noise is determined by the shot noise associated with leakage current, or dark current, that exists in the absence of input photons. In contrast, SPAD performance is degraded by “dark counts” that arise when carriers are created by processes other than photoexcitation. Both thermal excitation and field-mediated creation of free carriers (i.e., tunneling processes) contribute to the dark count rate (DCR). To improve SPAD performance, InGaAs/InP SPADs are usually operated in gated mode. The detector is biased at a baseline voltage just below  $V_{br}$ , and to “arm” the detector, a gate pulse is applied to bring the detector bias above breakdown for a short period of time, generally between 1 and 100 ns. Once an avalanche is initiated, it must be quenched. Gated quenching allows the avalanche to persist for the fixed duration of the gate. Passive quenching employs a resistor in series with the APD so that the avalanche current induces a voltage across the resistor and drops the APD bias below  $V_{br}$ . Active quenching uses circuitry to rapidly detect an avalanche and actively force the APD bias below  $V_{br}$ .

SPAD timing jitter results from the fluctuations in temporal correspondance between the arrival of a photon and the detection of a resulting avalanche. In particular, fluctuations in the way a runaway avalanche spreads laterally across the active area of a SPAD have been shown to dominate the timing jitter for single photon detection.

The afterpulsing effect, unique to SPADs, occurs when carriers created during an avalanche are trapped by defects in the multiplication region. As these carriers are freed at a later time, generally by thermal emission, they can lead to additional dark counts. To compensate for afterpulsing, one has to wait a sufficient “hold-off” time after the avalanche is quenched so that carriers can detrap before the next gate. Although the intrinsic DCR can be reduced by operating at lower temperature, afterpulsing

effects will be exacerbated by longer detrapping times and require the use of lower gate repetition rate.

DCR and SPDE are usually the two most critical performance parameters, and their optimization requires a trade-off: SPDE improves (i.e., increases) linearly with increasing bias above  $V_{br}$ , while DCR degrades (i.e., increases) exponentially with increasing bias. A plot of DCR vs SPDE is therefore linear on a semi-log plot, as shown in Figure 3, where we present data for two devices with a 25  $\mu\text{m}$  diameter active area that illustrate another performance trade-off. Device A (squares) has 4 to 5 times lower DCR at a given SPDE than Device B (circles) but at the expense of higher timing jitter. For Device A, a lower electric field in the absorption region leads to a lower DCR due to reduced tunneling, it also adversely impacts the carrier dynamics to yield higher timing jitter.

Figure 4 illustrates the effects of afterpulsing. DCR can be reduced by lowering the operating temperature, but the hold-off time required to reach the minimum DCR increases with decreasing temperature. The system-level impact is that if higher repetition rates are desired, they can only be obtained at higher DCR. Nevertheless, repetition rate is not a near-term concern for QKD systems. Many secure point-to-point links in use today have manual key distribution with timeframes of weeks or even months between key updates, while commercial QKD solutions already provide new encryption keys with absolute security in just fractions of a second.

### **SPAD Receiver Integration**

Beyond the need for high performance devices, SPAD-based receivers will require the management of optical, thermal, and electrical interfaces. To provide high coupling efficiency with single mode fiber, we have implemented a scheme that allows for sub-micron fiber alignment to small active area devices housed in a standard 14 pin butterfly-type package. Currently, a single-stage thermoelectric cooler (TEC) is integrated into the hermetically sealed package, and we will migrate to a multi-stage TEC to eliminate the need for cooling outside the package. By mounting a thermistor at close proximity to the SPAD inside the package, accurate temperature monitoring and control can be employed.

Finally, the electronic circuitry to detect SPAD avalanche pulses is critical to overall receiver performance. The ability to provide flexible high-speed gating and accurate low-threshold sensing of SPAD avalanche pulses is critical for SPAD use with QKD systems. For current testbed requirements, we have integrated our SPAD-based detector component with electronic circuitry to provide a benchtop receiver. As this field evolves, there will be increasing monolithic integration of the necessary gating and quenching circuits into CMOS chips, and the co-packaging of these chips with the SPAD will provide a high-performance, compact component, analogous to fiber optic receivers found in traditional telecommunications systems.

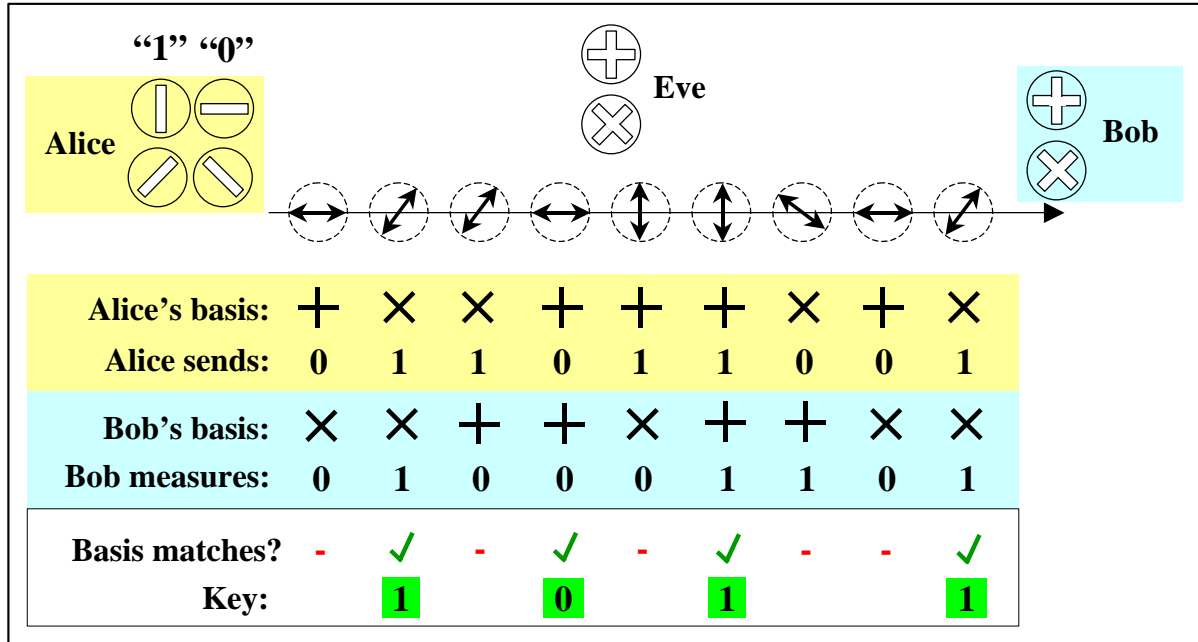
### **Evolution of InGaAs/InP SPADs**

Until very recently, there were no commercially available InGaAs/InP SPADs. Photon counting practitioners were forced to buy commercial linear mode APDs and evaluate their performance as SPADs. Although occasional devices were found to have good SPAD performance, most behaved rather poorly because they were not designed for Geiger mode operation. SPAD performance exhibited tremendous variations, often over several orders of magnitude for a fundamental metric such as a dark count rate.

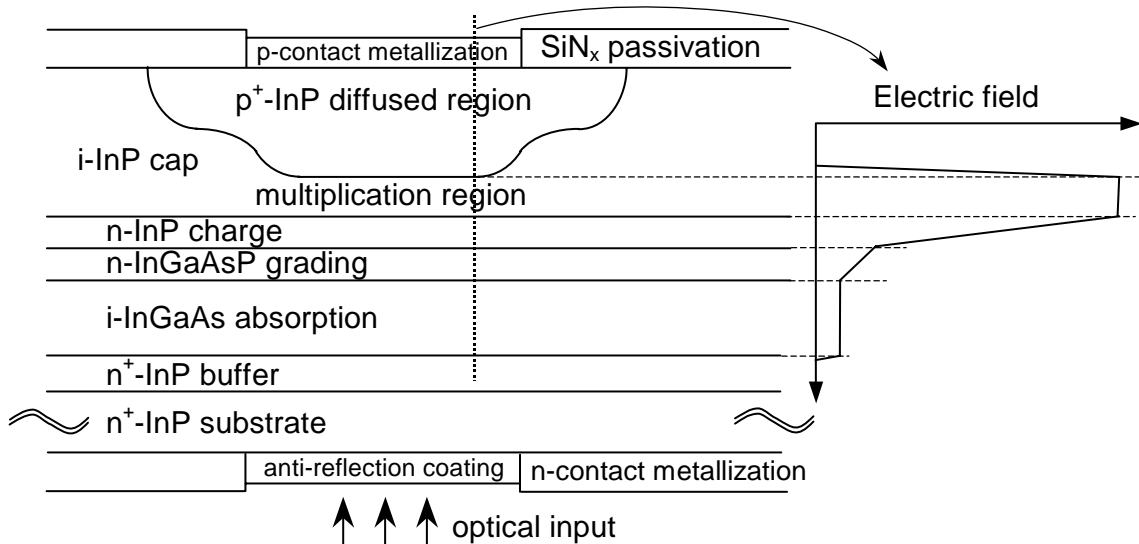
To address emerging single photon applications such as QKD, we have recently developed state-of-the-art InGaAs/InP SPADs. We expect to see progress in QKD deployment as the relevant component technologies are advanced, and SPADs will be one of the key technologies to enable improved system performance. We intend to focus on SPAD performance improvements that could be leveraged for higher SPDE or lower DCR. QKD link reach is currently limited by SPAD DCR performance, whereas SPDE is more important in determining single photon transmission rate. As described above, transmission rate is deemed to be sufficient in the first commercial systems available

today, and we expect that further SPAD performance improvements will be leveraged to further increased QKD link reach.

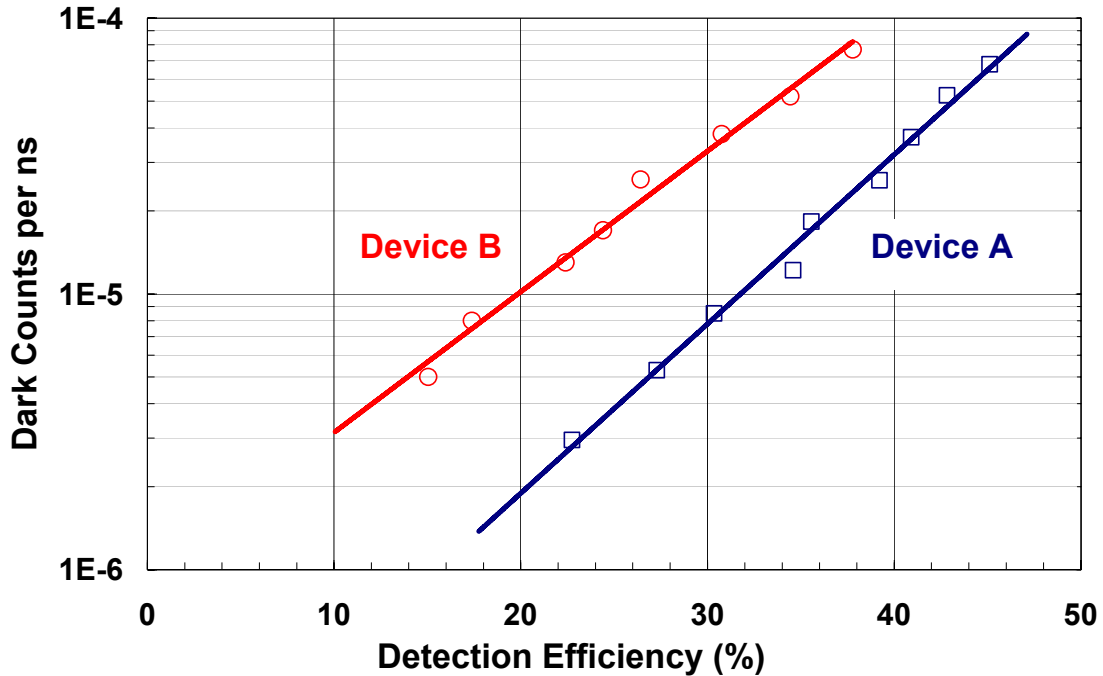
We wish to thank Dr. Radu Ispasoiu and Prof. Sergio Cova for device characterization and insightful discussions. Mark Itzler is Chief Technical Officer at Princeton Lightwave, Inc.



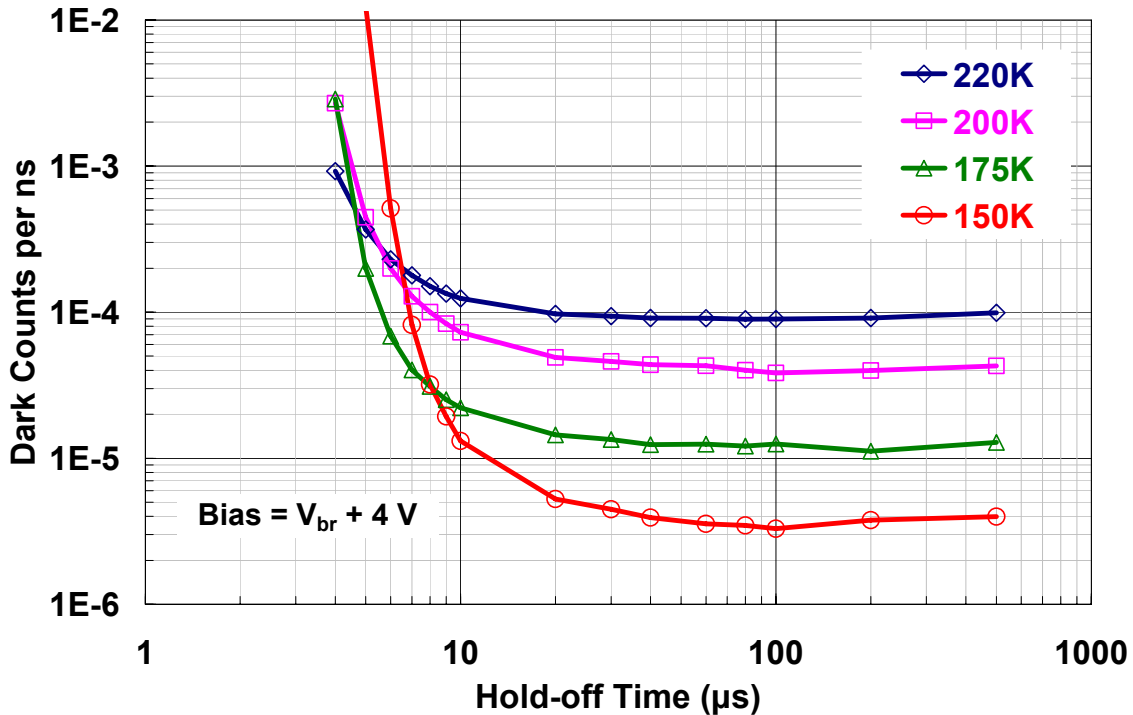
**Figure 1.** Representation of a quantum key distribution link employing photon polarization states. Alice and Bob randomly choose the horizontal/vertical basis or +45°/-45° basis for the encoding and measurement of “1” and “0” bits. Basis choice is compared, and when there is agreement, bits are retained for the encryption key.



**Figure 2.** Schematic cross-section of SPAD device structure. As in linear mode APDs, a doped charge layer maintains high field to generate avalanche gain in the multiplication region while keeping sufficiently low field in the absorption region to minimize field-induced leakage currents.



**Figure 3.** Measured SPAD performance for dark counts per nanosecond vs. single photon detection efficiency (SPDE). Device A achieves lower DCR for a given SPDE, but at the expense of increased timing jitter. Devices were operated at 200K with active quenching and a 10 kHz gate repetition rate.



**Figure 4.** Dark counts per nanosecond vs. hold-off time at four different temperatures for a 40 µm diameter SPAD illustrates worse DCR at short hold-off times due to afterpulsing effects. Devices were operated with 20 ns gates, gated quenching, and an operating bias at 4 V beyond the breakdown voltage.