INTRODUCTION TO LASER RADAR

A New Light on Imaging

Maurice J. Halmos

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Names: Halmos, Maurice J., author.
Title: Introduction to laser radar : a new light on imaging / Maurice J. Halmos.
Description: Bellingham, Washington : SPIE Press [2024] | Includes bibliographical references and index.
Identifiers: LCCN 2023029410 | ISBN 9781510661356 (paperback) | ISBN 9781510661349 (pdf)
Subjects: LCSH: Optical radar.
Classification: LCC TK6592.O6 H35 2024 | DDC 621.3848–dc23/eng/20230908
LC record available at https://lccn.loc.gov/2023029410

Published by
SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: http://spie.org

Photograph in cover design by the author.

Printed in the United States of America.
First Printing.
For updates to this book, visit http://spie.org and type "PM360" in the search field.

SPIE.

Contents

Preface

Welcome to *Introduction to Lidar: A New Light on Imaging*. This book began as a set of class notes that I provided to my students here at Raytheon, where I've spent most of my professional life developing, building, and testing prototype lidar sensors for a vast variety of applications. After that, it became an amalgamation of articles written by our team and others in the lidar community. My hope is to have compiled an accessible guide to the fascinating world of light detection and ranging (LiDAR) technology. Over the past few decades, lidar has revolutionized the way we perceive and interact with our environment, allowing us to capture various features and characteristics of objects and landscapes from a distance.

In this fast-paced era of technological advancement, lidar has emerged as one of the most influential tools in various fields, including remote sensing, geospatial mapping, autonomous vehicles, archaeology, forestry, environmental monitoring, and urban planning, among others. Its applications are diverse, its impact profound, and its potential seemingly limitless.

Although a few books have already been written about this technology, my aim is to provide both novices and professionals with a solid foundation in lidar technology, giving emphasis to the nuances of coherent detection. We will explore the fundamental principles that govern lidar systems and use them to explain the advantages and complications in implementing both direct-detection and coherent lidar. The latter is becoming more and more attractive in the community as an expansion of direct-detection lidar and its limitations.

Throughout the chapters, we explore the various technologies that comprise lidar, which include the basic technologies of lasers and detectors as well as radar technology. We'll touch on the quantum mechanical aspect of lidar and point out some of the fundamental similarities between lidar and radar as well as the fundamental differences, the latter of which make lidar a truly unique technology. We will unravel the technical aspects of lidar, explaining complex concepts in a clear and concise manner with the aim of making them accessible to readers from many backgrounds. While the subject can be technical, I promise to present it in an engaging and informative way, ensuring that you, the reader, embark on an enlightening journey into the world of lidar.

Whether you are an aspiring lidar professional seeking to enter this exciting field or a seasoned expert looking to expand your knowledge, my hope is that this book will be your faithful companion. The first half of the book explains the more fundamental aspects of lidar technology, and the second half dives into some of the more intricate issues associated with coherent lidar. Each chapter is carefully crafted to take you through the essential components, applications, and challenges of lidar technology, providing you with practical insights and examples.

I'd like to thank Scott Wright for thoroughly proofreading the original manuscript. And I want acknowledge all my colleagues at Raytheon Technologies in El Segundo, California and McKinney, Texas, who have been working with me in developing this fascinating technology during the last decade. I hope this book will not only broaden your understanding of lidar but also spark curiosity and inspire you to explore its applications in innovative and groundbreaking ways. As you read the pages ahead, I encourage you to embrace the wonders of lidar and envision the possibilities it holds for shaping our future.

Thank you for joining me on this enlightening journey into the realm of lidar. Let's begin this exploration together and unlock the secrets of the world in three dimensions!

Happy reading!

Maurice J. Halmos
December 2023

Chapter 1
Ladar: A New Light on Imaging

1.1 Introduction and Overview

Pulsed optical range measurement devices (rangefinders) using incoherent spark sources existed before 1960, when the laser was discovered. The laser's fast, high-energy pulses and highly collimated monochromatic beams available from Q-switched devices revolutionized their capabilities. Its intense monochromatic output with low divergence makes the laser an ideal source for accurate ranging and high-resolution imaging. It was soon realized that these narrow-linewidth sources would make heterodyne detection possible in the infrared (IR) and optical spectral range.

Laser radar (or ladar—laser detection and ranging) is an extension of conventional microwave radar techniques to much shorter wavelengths (by a factor of 100,000).[1] Like microwave radar, ladar can simultaneously measure range, velocity, reflectivity, and azimuth and elevation angles. Ladar is well suited for precise measurements useful in target classification and recognition but ill-suited for wide-area search because of the time and energy required.

Laser radars, with their optical wavelengths and active sensing, behave like forward-looking infrared (FLIR) sensors in terms of angular resolution, and like microwave radars in terms of range and velocity measuring capability. However, coherent effects and extreme wavelength differences give rise to phenomena not seen in these more traditional sensors.

For example, coherence of the laser transmitter causes speckle in the return from optically rough target surfaces. The apparent brightness of individual scene pixels may fluctuate wildly, giving visually poor intensity imagery unless considerable scene averaging is applied, which requires more time on target and consumed energy. Often, ladar images are better displayed as range images than as intensity images.

The extremely short wavelengths typical of ladars move the noise floors well into the quantum-dominated regime. Where thermal noise is the driving limit in sensitivity in radars, in ladars, the quantized photon energy in the signal and background light drive the noise floor sensitivity. As the wavelengths approach visible light, the signal itself becomes noise-like and the detection threshold becomes roughly one photon. In addition, the short wavelengths give rise to huge Doppler shifts that may require processing bandwidths far greater than needed in conventional radar.

1.2 Lidar or Ladar

Two acronyms are typically used: ladar and lidar. Ladar stands for Laser Radar or Laser Detection and Ranging, which refers to imaging systems (1D, 2D, and 3D), target-oriented sensors, and point sensors (velocimeters, vibration meters, multi-spectral absorption or scattering).

Lidar stands for Light Detection and Ranging, which refers typically to atmospheric measurements of aerosols for global wind sensors (satellite), pollution clouds, and fluorescence phenomena. Frequently, the detected light is not the exact same wavelength as the transmitted light due to florescence or other phenomena. Chemical/biological sensors may fall under this definition, although they are usually included in these categories when ladar is used. Recently, however, the term lidar has become more commonly used in a ladar context when dealing with autonomous vehicles.

1.3 Why Ladar?

Military conflicts at the end of the millennium did not tolerate collateral casualties very well, which forced a major change in the rules of engagement. Mainly, in today's battlefield, an operator needs to identify a potential target before deploying weapons. An aircraft fighter must determine whether a potential target is a military truck, a civilian truck, a bus, a tank, a missile launcher, etc. This is referred to as Combat ID, or CID, and ladar imaging has been identified as one of the few sensors, if not the only sensor, that can effectively do the job, even if the target is somewhat covered by camouflage, foliage, or just clutter.

Other nonmilitary applications also exist. Space missions such NASA's attempt to map possible landing sites for unmanned crafts or space docking guides are also ideal for ladar. A ladar was used to map the ruins of the World Trade Center after the horrific attacks to create a 3D model and determine if the structures were stable enough for rescue and cleanup forces. Ladars are also used to map terrain, the height of trees in forests, and the urban landscape of cities. In recent years, a slow explosion in the field of autonomous vehicles (AV) has brought the use of ladar (or lidar) to the limelight. There are a number of lidar sensor companies as well as software navigation companies that use the output of the lidar (we use the word lidar here because most of the commercial companies refer to their product as such). Ladars are also used to convert complicated existing structures into digital models for computer-aided designs (CADs). On the lighter side, Hollywood has been using ladars to realistically integrate computer animation into real footage, and this has been done so successfully that almost every film has, undetected to the eye, computer constructs intertwined with real scenes. 3D imaging at short and long ranges is becoming a priority.

1.4 Principles of Ladar

Like sonar and radar, ladar is based in the principle of transmitting a laser pulse and listening for the echo, as shown in Fig. 1-1. The difference between the three types of systems shown in the figure is the nature of the wave that is transmitted and received. Sonar works with a sound wave, radar (RAdio Detection And Range) with a radio electromagnetic waves, and ladar with laser light electromagnetic waves. Radar and ladar both operate with the same electromagnetic medium but differ greatly in the wave frequency or wavelength. In all cases, it is important to remember the wave relationships among these parameters:

Wavelength λ is the distance between repeated states.
Period T is the time between repeated states.
Speed v is approximately 3×10^5 m/s for sound waves and 2.998×10^8 m/s for electro-optical waves. The latter is usually denoted as c.
Frequency f is defined as $f = 1/T$ and is also defined in the following relations:
$f = v/\lambda$ for acoustic waves.
$f = c/\lambda$ for electromagnetic waves.

Figure 1-2 shows a visual representation of these relationships.

There are four elementary functions that one can perform with a ladar and that are the basis for all of the more complex functions that are performed or envisioned. The first function is the basic ranging, which measures the time of flight of a laser pulse, as depicted in Fig. 1-3. The distance from the

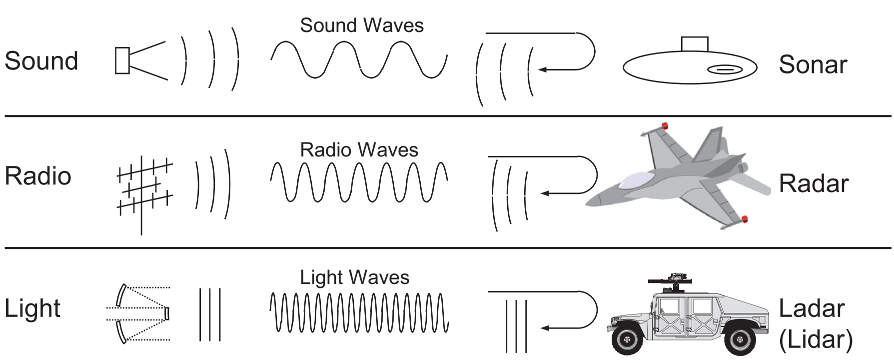


Figure 1-1 All active systems such as sonar, radar, and ladar include a transmitter and a receiver that listens for the echo.

Chapter 3
Receiver Detector

Receiver detectors are the eyes of the ladar. They are used as single elements or arrays (Fig. 3-1). As we will see, there are a few options and characteristics that greatly impact the architecture of the overall sensor. The receiver detector has characteristics such as quantum efficiency, gain, pitch, dark current, bandwidth, and more. These characteristics determine the sensitivity of the ladar, which prescribes the size of the laser transmitter that it needs to carry. They can also help to determine how well we can resolve a target in all three dimensions. In order to understand the choices and what criteria to use to make them, we will review some detector fundamentals.

3.1 Detector Fundamentals

An optical detector is a physical device that converts photons to electrons. The ideal detector will have a one-to-one correspondence such that N photons yield N electrons. Figure 3-2 depicts the function of a typical detector.

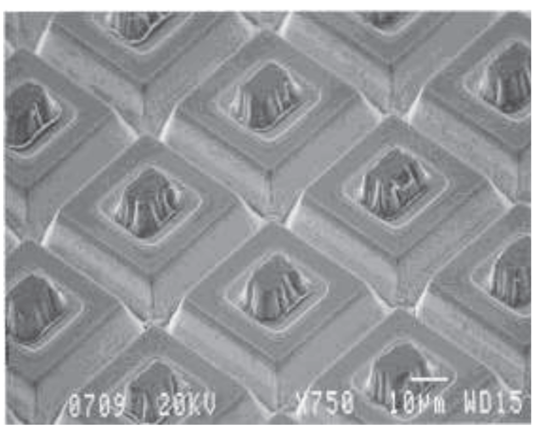


Figure 3-1 Section of a detector array. Detector mesas are shown with indium bumps to connect to a readout integrated circuit (ROIC). (Courtesy of Raytheon Vision Systems, 1998.)

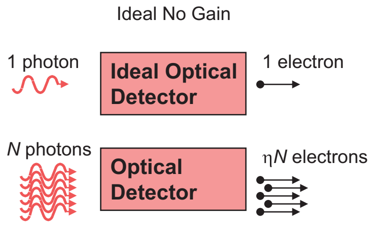


Figure 3-2 Detector function with no gain. *N* photons will yield η*N* electrons, where η is the quantum efficiency between 0 and 1.

Detectors are quantum devices, for which the likelihood of a photon interacting with an electron is given by the quantum efficiency η such that if N photons hit the detector, only ηN electrons will exit. A flow of photons constitutes optical power, which yields a flow of electrons that constitute a current. The conversion factor of power P to current in a detector is called the responsivity $\Re$. Responsivity can be calculated by the following relation:

$$I = P \cdot \Re = P \cdot \eta \cdot \mathrm{e}/h\nu \tag{3-1}$$

such that

$$\Re = \eta \cdot \mathrm{e}/h\nu. \tag{3-2}$$

Equation 3-1 simply says that the amount of electron charge per unit time is the number of photons per unit time multiplied by the quantum efficiency as well as by the electron charge e of 1.602×10^{-19} coulombs. The number of photons per unit time is the power divided by the photon energy, $h\nu$ (Planck's constant times the frequency of the optical wave). The current that is created with the detector can be observed with the aid of an electronic instrument such as an oscilloscope. Figure 3-3 shows how a stream of photons would be converted to a current or a voltage.

The stream of photons can be thought of as a stream of impulse (delta) functions hitting a finite-bandwidth linear system with an impulse response function $h(t)$. The impulse response function has a decay time τ, which approximately equals $1/(2 \cdot \text{bandwidth})$. A stream of photons will not be evenly spaced but will be random in nature, thus, showing a Poisson

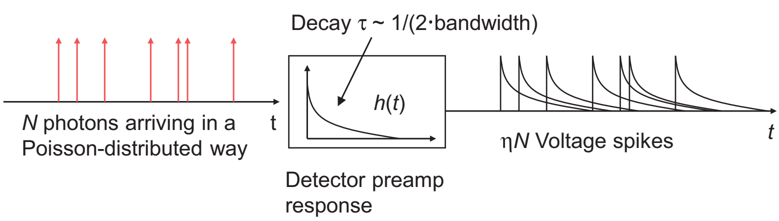


Figure 3-3 Train of Poisson-distributed photons generating an electric current.

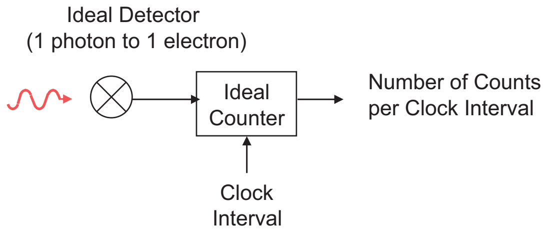


Figure 3-4 Block diagram of a photon-counting receiver.

distribution, as discussed later in this chapter. A discussion of linear systems and noise as shown in Fig. 3-3 is presented by Papoulis.[1] A detector with the impulse response shown above will perform an exponential average, where the impact of a past event diminishes exponentially as time passes. A linear integrator that is reset every period τ would be considered a photon counter. In other words, the number of photon hits per unit time τ would be counted in order to measure the average photon flux. Figure 3-4 shows a block diagram of a photon-counting detector.

A random stream of photons with a constant average flux would be measured by a photon-counting receiver by counting the number of photons per integration time. Figure 3-5 shows the Poisson-distributed stream of impulses (top trace) and the count realization for five measurements. The count integration time is 10 arbitrary time units.

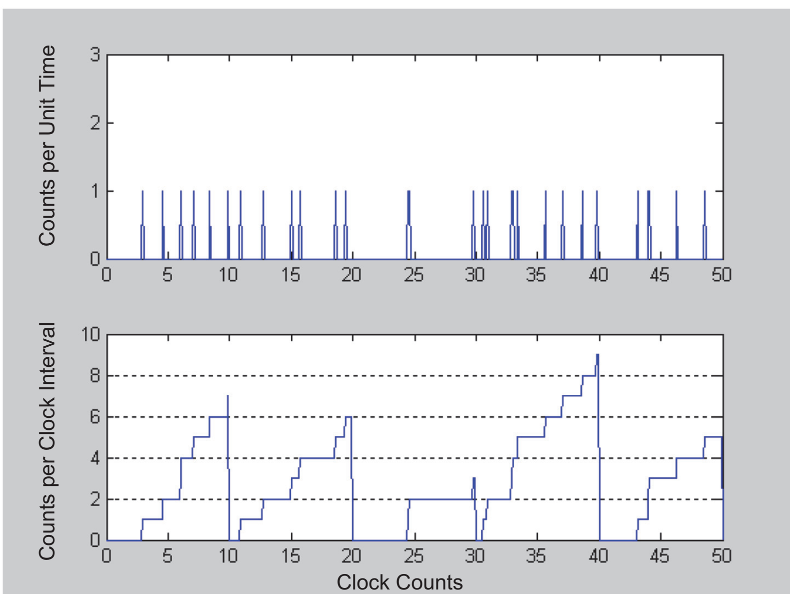


Figure 3-5 Number of counts per 10 units of time of a Poisson average of 0.6 counts per time unit.

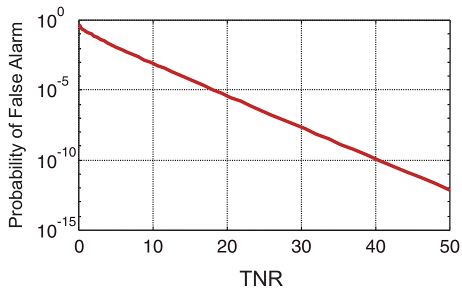


Figure 5-2 Probability of false alarm as a function of TNR.

there will be T/τ measurement opportunities per shot. For example, if the range is at 1.5 km (10 μs round trip) and the pulsewidth is 10 ns, we obtain $T/\tau = 1000$. If we desire the false alarm per laser shot to be ~ 1%, then the desired P_{fa} should be 10^{-5}. Another way that P_{fa} is expressed is in terms of the false alarm rate (FAR). In the example just discussed, the P_{fa} is 10^{-5} every 10 ns, so the FAR would be $P_{fa}/\tau = 1000$ per second, assuming that we are continuously measuring range. However, more likely, a user would be more interested in knowing how many false alarms we get per range determination. So again, from the previous example, if we are measuring ranges at a rate of 10 kHz with a false alarm per laser shot of 1%, then the FAR would be 10 kHz/100 = 100. The P_{fa} described here is what is referred to as the false alarm per range bin ($c\tau/2$).

Using the graph in Fig. 5-2, we see that a P_{fa} of 10^{-5} required a TNR of approximately 19. We observe that the TNR dependence is relatively low; if, for instance, we reduce the P_{fa} by a factor of 10 to 10^{-6} (increase the range to 15 km), the TNR increases only by ~ 0.5% to 20. It is also worth noticing that when the TNR is zero, i.e., the threshold is set to zero, the P_{fa} is 0.5. This occurs because the noise is both positively and negatively distributed with equal magnitude; therefore, there is equal chance that a noise-only signal is found to be positive (above threshold) or negative (below threshold).

The threshold level that determines the TNR is set usually with only the consideration of the resultant P_{fa} as a function of the noise. The expected target signal level is not considered. Part of the reason for this is that target detection and identification is more sensitive to false pixels than to missing image pixels, so the detection threshold is usually set completely independent of the expected signal.

5.1.2 Probability of detection P_d

We now assume that the Gaussian noise receiver described by Fig. 5-1 has a signal of magnitude i_S. The probability density function of the resultant current is the same Gaussian distribution that was centered around zero, now shifted to be centered around i_S. This is analogous to the situation where we

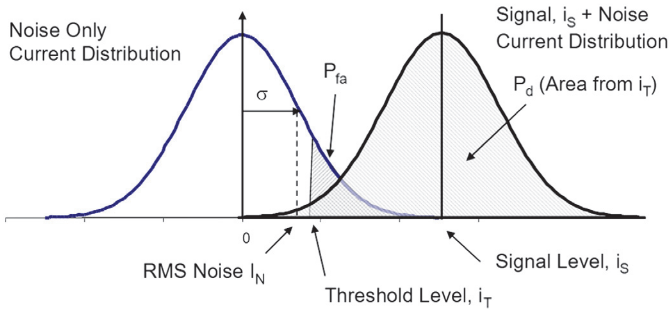


Figure 5-3 Signal-plus-noise current distribution. The area under the curve above the threshold current is the probability of detection P_d.

observe a noisy receiver with an oscilloscope; with no signal, the noise appears as a fuzzy vertical spread around the zero line. If now a signal is injected of say one vertical division, we would observe the same fuzzy spread across the screen of the oscilloscope but shifted up by one division. The shifted PDF distribution is shown in Fig. 5-3.

A signal will shift the curve to the positive direction, making the area under the curve that is above the threshold ideally closer to unity, which is much larger than the previous P_{fa}. Typical P_d values sought are from 80% to 99.99%, depending on the sensor mission.

We can derive an expression for P_d by integrating, as before, the received current PDF from the threshold i_T to infinity. By denoting the received current as i_R, the PDF can be expressed as

$$\begin{aligned} p(i_R) &= \frac{1}{\sqrt{2\pi}\cdot I_N}\cdot \exp\left(-\frac{(i_R - i_S)^2}{2I_N^2}\right) \\ E(i_R) &= i_S \\ \sigma_R &= \sqrt{E(i_R^2) - E(i_R)^2} = I_N \equiv \text{rms noise}, \end{aligned} \tag{5-9}$$

where i_S is the signal strength. We observe that the standard deviation term σ_R is still the rms noise current I_N. In order to make the integral from i_T to infinity, we perform the following substitutions:

$$z \equiv \frac{1}{\sqrt{2}}\frac{(i_R - i_S)}{I_N},\ \mathrm{TNR} \equiv \left(\frac{I_T}{I_N}\right)^2,\ \mathrm{SNR} \equiv \left(\frac{I_S}{I_N}\right)^2, \tag{5-10}$$

where, again, z is the integration variable, and TNR and SNR, as before, are the threshold-to-noise ratio and the signal-to-noise ratio, respectively. These substitutions not only allow the integration to be done, but also reduce the absolute terms of noise, signal, and threshold to figure-of-merit relative terms of TNR and SNR. The integration of the probability density function from the threshold current to infinity in Eq. (5-9) becomes

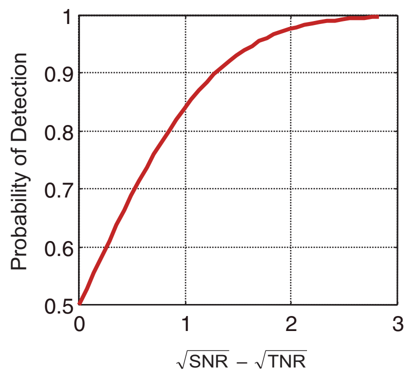


Figure 5-4 P_d vs SNR and TNR.

$$P_d = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{(\sqrt{\text{SNR}} - \sqrt{\text{TNR}})/\sqrt{2}} e^{-z^2} dz$$
$$= \frac{1}{2}\left[1 + \text{erf}\left\{(\sqrt{\text{SNR}} - \sqrt{\text{TNR}})/\sqrt{2}\right\}\right], \tag{5-11}$$

where the error function erf(x) is defined by Eq. (5-8). The probability of detection as a function of the TNR and SNR is plotted in Fig. 5-4.

Notice that when the threshold is at the signal level (i.e., SNR = TNR), P_d is 50%. In this case, we have the same situation as we had with the P_{fa} when the threshold was zero. If the threshold is at the signal level, the noise will spread the measured current evenly above and below the threshold, yielding a 50% probability that the current level is correctly identified as a signal present. As the SNR becomes much larger than the TNR ($i_S \gg i_T$), the P_d approaches unity.

5.1.2.1 Example

Use Figs. 5-2 and 5-4 and determine what TNR and SNR values are needed to obtain a P_d of 90% and a P_{fa} of 10^{-6}.

5.1.2.2 Answer

From Fig. 5-2, we determine that a TNR of ~20 will yield the required P_{fa} of 10^{-6}. From Fig. 5-4, we determine that, in order to have a P_d of 90%, we need $\sqrt{\text{SNR}} - \sqrt{\text{TNR}}$ to be 1.3. Solving for the SNR, we obtain

$$\text{SNR} = (1.3 + \sqrt{20})2 = 33.$$

If we were to desire $P_d \sim 99\%$, the required SNR would increase to ~46. Typical SNR requirements of ~50 are common, which in SNR_V ($\text{SNR}_V = \sqrt{\text{SNR}}$) denomination, is about 7.

5.5.2 Speckle distributed signal averaging in Gaussian noise

So far, we have assumed that the signal averaging is done on a deterministic signal; however, one usually has one or more speckle averages in the single detection event before performing the temporal average. As mentioned, averaging multiple speckle lobes as well as averaging multiple pulse events will improve the statistics of the signal, possibly yielding a greater improvement than is described in the previous section. In order to calculate the SNR required per pulse when N speckles are captured and M independent pulses are averaged, one calculates from Eq. (5-23) the combined probability density function, which is obtained through the convolution of the Gaussian noise with the $N \times M$ auto-convolutions of the Rayleigh-distributed signal. The PDF is then integrated from a threshold to infinity to obtain the detection probability. Osche[11] developed an analytical expression, given by

$$P_\mathrm{d} = \frac{1}{2}\mathrm{erf}\left(\sqrt{\mathrm{TNR}\frac{\mathrm{TNR}}{2}}\right) + \frac{\left(\frac{n\cdot M}{\sqrt{\mathrm{SNR}}}\right)^{n\cdot M-1}}{\sqrt{2\pi}} \times \exp\left[\frac{(n\cdot M)^2}{4\mathrm{SNR}} - \frac{n\cdot M}{2}\sqrt{\frac{\mathrm{TNR}}{\mathrm{SNR}}} - \frac{\mathrm{TNR}}{4}\right]\left[D_{-n\cdot M}\left(\frac{n\cdot M}{\sqrt{\mathrm{SNR}}} - \sqrt{\mathrm{TNR}}\right) + \sum_{j=1}^{n\cdot M-1}\left(\frac{n\cdot M}{\sqrt{\mathrm{SNR}}}\right)^{-j} D_{-n\cdot M+j}\left(\frac{n\cdot M}{\sqrt{\mathrm{SNR}}} - \sqrt{\mathrm{TNR}}\right)\right]. \tag{5-35}$$

$D_\nu(z)$ is the parabolic cylinder function (where ν is the order of the function), defined as[4]

$$D_\nu(z) \equiv 2^{\nu/2}\exp\left(\frac{-z^2}{4}\right)\left[\frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1-\nu}{2}\right)}\cdot {}_1F_1\left(\frac{-\nu}{2},\frac{1}{2},\frac{z^2}{2}\right) + \frac{z}{\sqrt{2}}\frac{\Gamma\left(-\frac{1}{2}\right)}{\Gamma\left(\frac{-\nu}{2}\right)}\cdot {}_1F_1\left(\frac{1-\nu}{2},\frac{3}{2},\frac{z^2}{2}\right)\right], \tag{5-36}$$

where $\Gamma(z)$ is the gamma function and ${}_1F_1(x,y,z)$ is the confluent hypergeometric function.

The expressions in Eqs. (5-35) and (5-36) can be cumbersome to evaluate; however, a simpler approximation may be possible. Let's assume that the speckle intensity distribution is Gaussian rather than negative exponential. The sum of the speckle-limited signal and the background thermal Gaussian noise would yield a Gaussian distribution as well. Using the derivation for Eq. (5-11), we write

$$P_d = \frac{1}{2}\left[1 + \mathrm{erf}\left\{\frac{i_S - I_T}{\sqrt{2\mathrm{var}(i_S)}}\right\}\right]. \tag{5-37}$$

The variance of the current i is now composed of the noise I_n as well as the variance of the signal itself due to the speckle statistics. We can write the variance as the rms of the two quantities:

$$\mathrm{var}(i) = I_N^2 + i_S^2/N, \tag{5-38}$$

where the second term, i_S^2/N, is the variance of the speckle-limited signal averaged N times (temporally or spatially). Replacing Eq. (5-38) in (5-37), we obtain

$$P_d = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\sqrt{\mathrm{SNR}} - \sqrt{\mathrm{TNR}}}{\sqrt{2}\cdot\sqrt{1 + \mathrm{SNR}/N}}\right)\right]. \tag{5-39}$$

From the law of large numbers, we know that most distributions approach Gaussian when multiple trials are averaged, so we expect Eq. (5-38) to closely track the more exact expression in Eq. (5-35) when N is large. For small values of N (less than 5), a correction factor can be used to bring the evaluated expressions of Eq. (5-39) closer to Eq. (5-35). The empirical correction function is used:

$$\eta(M) \approx \frac{1}{1 + 0.2\exp(-N/2)}. \tag{5-40}$$

Notice that the correction factor approaches unity for $N > 5$. Figure 5-16 shows a plot of the probability of detection as a function of M, the number of averages. The values used are SNR = 30, TNR = 14.

Although Eqs. (5-35) and (5-36) can be evaluated, they are complex and difficult to use, particularly for finding roots. The expression in Eq. (5-39) and the empirical correction in Eq. (5-40) are much more practical to use and yield fairly accurate results. This would be especially true if we consider that the noise and signal distributions are not always exactly Gaussian and negative exponential due to nonideal implementations of the receiver.

We observe in Fig. 5-16 that the probability of detection rises rapidly as the number of integrations goes from 1 to 10, and then the improvement slows down. The main reason is that the intensity variations of the speckle require a relatively high SNR for a single pulse to maintain a low probability of dropping

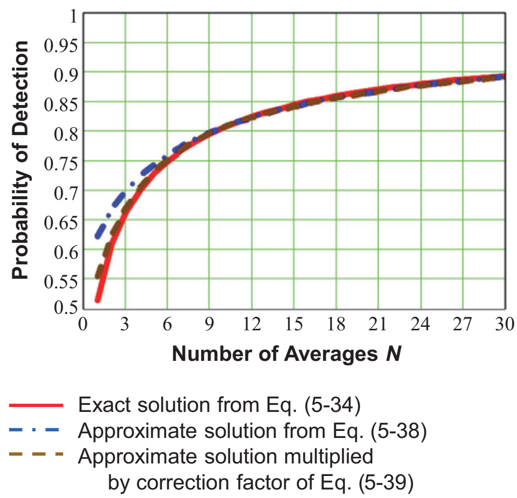


Figure 5-16 Calculations of P_d for fixed SNR and TNR as function of the number N of speckle or temporal averages.

below threshold (i.e., keeping the probability of detection high). However, as we average signals, the speckle-induced excursions are reduced, thus requiring a lower average signal-to-threshold ratio. This effect can be observed by plotting the integration loss L, as shown in Fig. 5-17.

Figure 5-17 shows that total energy transmitted from a train of incoherently averaged pulses can be lower than the energy required from a single pulse for the same P_d. The higher the required P_d the higher the excess energy of a single pulse must be to overcome the speckle-induced Rayleigh fading. By averaging a few pulses, the energy required to maintain the low probability of a fade drops rapidly, making the total energy required less than

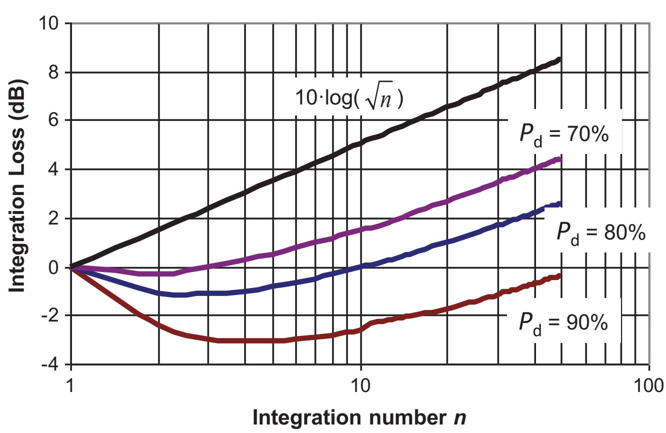


Figure 5-17 Numerical calculations of the integration loss for three P_d values, 90%, 80%, and 70%, while keeping $P_{fa} \sim 5 \times 10^{-4}$.

5 pA/$\sqrt{}$Hz. Even more challenging would be packaging the circuitry into the small unit cell of a receiver FPA of 50- to 100-μm pitch. In addition to the TIA preamp, threshold setting and logic must be implemented per detector element. All of this circuitry creates an additional challenge when a large FPA is desired (64×64, 128×128, or larger). If all of the circuitry does not fit in a single unit cell, the rest must be fanned out to the periphery of the array, requiring signal wires to cross over the entire array. This will limit the size of the FPA that would be achievable, especially when very low noise must be maintained.

As mentioned earlier, as difficult as it is to meet the TIA requirements in Table 5-1, those requirements are expected to be optimistic because, among other things, we assumed that the APD gain would be perfect (not distributed around an average). This has recently been discovered to be possible with HgCdTe material when operating in an electron-only carrier.[7,8] Other materials will have a gain distribution that follows the McIntyre model.[9]

5.8 Crosstalk Noise

So far, all of the analysis has been with respect to a single detector. One effect that becomes apparent when the detectors are in an array configuration is the crosstalk effect. When a detector triggers, a large number of electrons are generated in the avalanche region of the detector. The effect is so strong that even photons are generated as well. Not all of the electrons and new photons can be fully confined to a single detector; therefore, the possibility of triggering neighboring detectors occurs. The probability of triggering varies from a fraction of a percent to a few percent and may follow a pattern that skips the adjacent detectors. In the following numerical example, we will use one possible pattern, as shown in Fig. 5-32.

5.9 Example: Numerical Simulation of Geiger-Mode Detection

In order to understand the effect of the dark count noise and the ability to remove it using coincidence filtering, we have implemented the above statistics in a numerical model. In addition, we have built into the model the effect of crosstalk (detector triggers generated from other valid or noise triggers).

A test scene was created consisting of a pentagon, a cube, and undulating ground with a few mounts. The no-noise point cloud from the scene is shown in Fig. 5-29. We run our detection simulation with a dark count level of approximately 30 kHz, using 40 pulses and an expected return of 0.2 detected photons per pulse. Figure 5-30 shows the expected point cloud plus the additional false detections due to the dark counts.

We then apply a coincidence threshold of 3, eliminating most of the dark triggers, but occasionally eliminating a target return as well, as seen in

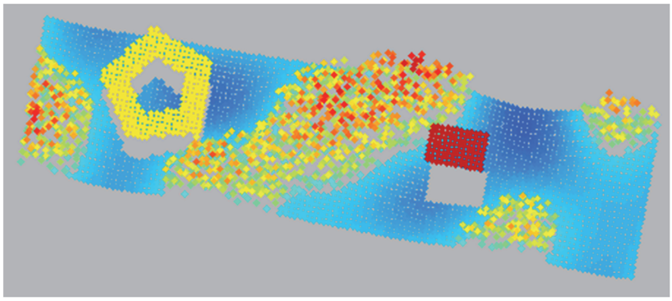


Figure 5-29 Point cloud of the test scene with no dark counts or crosstalk triggers.

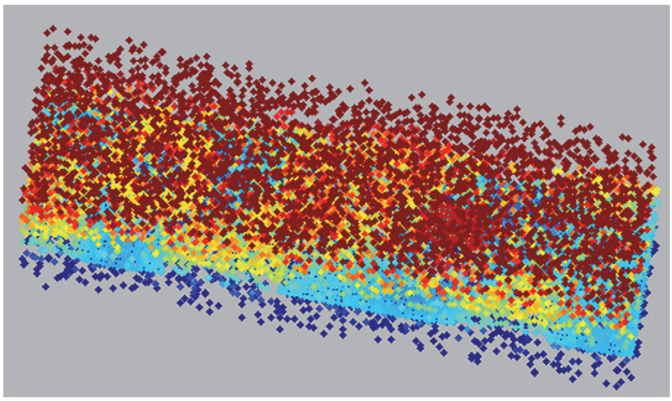


Figure 5-30 Point cloud of the test scene with 30-kHz dark counts but no crosstalk triggers.

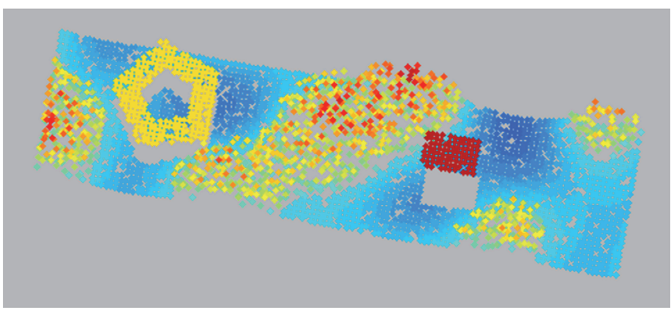


Figure 5-31 Point cloud of the test scene with 30-kHz dark counts and no crosstalk triggers after a coincidence filter of magnitude 3.

Fig. 5-31. In order to observe the effect of crosstalk, we turn the dark counts off and apply 2% crosstalk in the pattern shown in Fig. 5-32.

The 2% probability of crosstalk trigger refers to the bright, highest probability of trigger, while the non-black cells represent lower non-zero values. One can imagine that the crosstalk triggers will blur the image somewhat, and

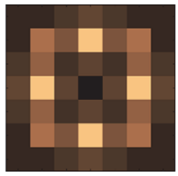


Figure 5-32 Crosstalk pattern. The center cell is triggered by a photon, the brighter neighbors represent a higher probability of a crosstalk trigger, and the darker ones represent a lower probability of a crosstalk trigger.

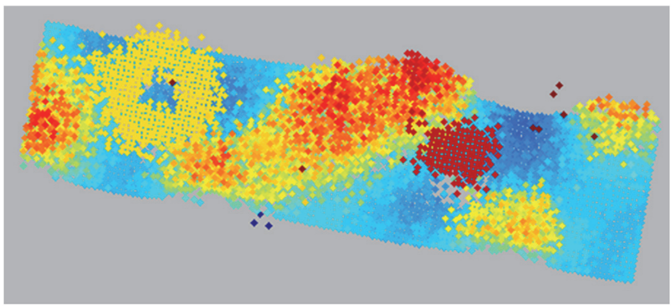


Figure 5-33 Point cloud image with no noise and 2% crosstalk in the pattern shown in Fig. 5-32.

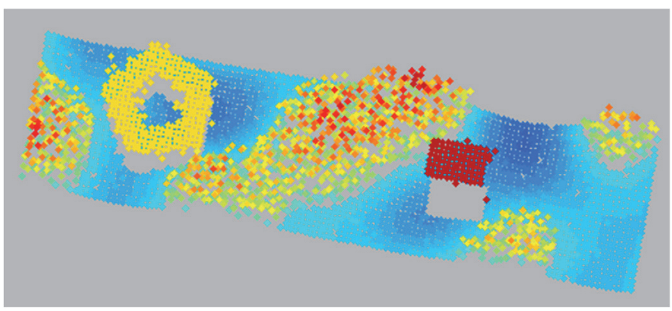


Figure 5-34 Point cloud of the test scene with no dark counts and 2% crosstalk triggers after a coincidence filter of magnitude 3.

this effect is indeed seen in Fig. 5-33, where both the pentagon and the cube appear to be very blurred.

However, applying coincidence filtering to the image will eliminate a fair amount of crosstalk because those false triggers don't occur every time, but only a small percentage of the time. Figure 5-34 shows the point cloud image after coincidence filtering of three detections.

We observe that the coincidence filter is quite effective in reducing both dark-count noise and crosstalk noise in the point cloud. Despite the filter, we observe plenty of false points at the edges of the shapes. The crosstalk level of the detector must be at a sufficiently low level not to corrupt the image by blurring it.

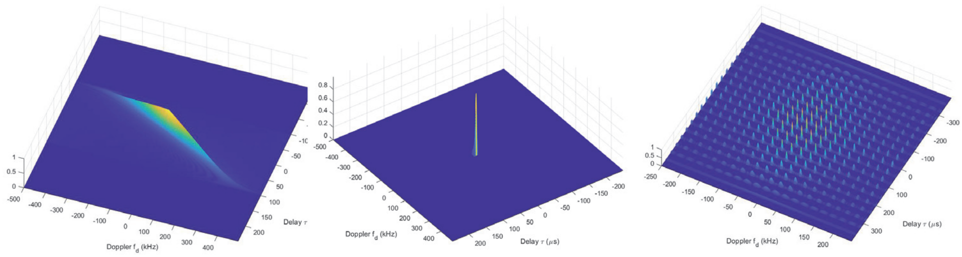


Figure 7-3 Three groupings of ambiguity function types according to Skolnik:[1] (left) knife edge, (center) thumbtack, and (right) bed of spikes.

obtained when the time evolution of the pulse is sharp, i.e., when we have a short pulsewidth. In the same vein, if we sample at the time of the return—the peak of the signal across all frequencies—we obtain the best frequency resolution when the filter bank shows a sharp peak across the frequencies. In other words, the ideal ambiguity function for high resolution in time and frequency is one that has a narrow peak in both time and frequency.

Skolnik[1] pointed out that the ambiguity function shapes can be grouped into three types: knife edge (or ridge), thumbtack, and bed of spikes, as depicted in Fig. 7-3. Clearly, for maximum resolution, the thumbtack shape would be desired; however, due to the fact that time and frequency coordinates are not always independent, one finds that the knife edge or bed of spikes is more likely to be achieved.

7.2.2 Example: Short squared pulse

Let's assume that the waveform is a square pulse of duration τ_p, represented by

$$\begin{aligned} s(t) &= \frac{1}{\sqrt{\tau_p}}\,\mathrm{rect}\left[\frac{t}{\tau_p}\right] = \frac{1}{\sqrt{\tau_p}} \quad \text{for } |t| \le \tau_p/2 \\ &= 0 \qquad\qquad\qquad\qquad\quad\;\; \text{elsewhere.} \end{aligned} \tag{7-7}$$

The ambiguity function for this waveform can be calculated as

$$\chi(t, f) = \mathrm{rect}\left[\frac{t}{2\tau_p}\right] e^{i\pi f t} \cdot \frac{(\tau_p - |t|)}{\tau_p} \cdot \frac{\sin \pi f(\tau_p - |t|)}{\pi f(\tau_p - |t|)}. \tag{7-8}$$

A 3D graph of Eq. (7-8) is shown in Fig. 7-4.

The peak of the ambiguity function falls at $t = 0$ and $f = 0$ because the delay and frequency shift of the signal are assumed to be zero. In order to get a closer look at the function, we can slice the 3D plot in frequency at $t = 0$ and slice it in time at $f = 0$. From the definition of the ambiguity function in Eq. (7-6), we saw that when we let $f = 0$, we get the autocorrelation of the waveform, which in this case, as seen from Eq. (7-8), would be a triangular pulse with a base width of $2\tau_p$:

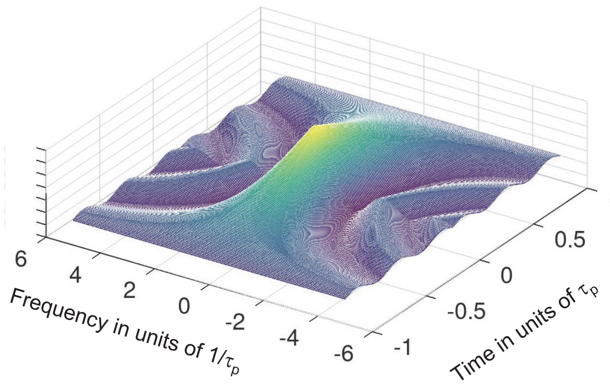


Figure 7-4 Plot of the ambiguity function of a square pulse.

$$\chi(t, 0) = \text{rect}\left[\frac{t}{2\tau_p}\right] \cdot \frac{(\tau_p - |t|)}{\tau_p}. \tag{7-9}$$

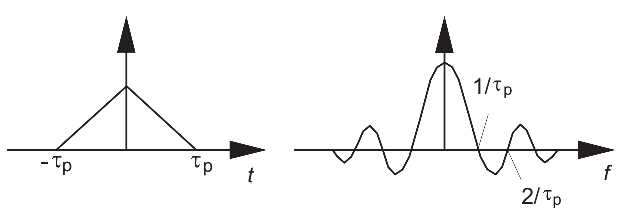
On the other hand, when we let $t = 0$, we get the Fourier transform, which again from Eq. (7-8), we see is a sinc function with the first nulls at frequency $f = 1/\tau_p$ and $-1/\tau_p$:

$$\chi(0, f) = \frac{\sin \pi f \tau_p}{\pi f \tau_p}. \tag{7-10}$$

Figure 7-5 shows plots of the ambiguity function at $t = 0$ and $f = 0$ of Eqs. (7-9) and (7-10).

If we require high time (range) resolution, we would make τ_p very small, but that would make the frequency resolution poor. On the other hand, if we require high resolution in frequency, we would need to make τ_p as large as possible, which would mean poor range resolution. This is illustrated in Fig. 7-6, where the full-width at half-maximum (FWHM) of the ambiguity function is depicted.

We can think of the FWHM view of this ambiguity function as a balloon; if we squeeze the time domain for better resolution, the balloon bulges in the



Drawings of ambiguity function cuts

Figure 7-5 Autocorrelation and Fourier transform of a square pulse waveform.

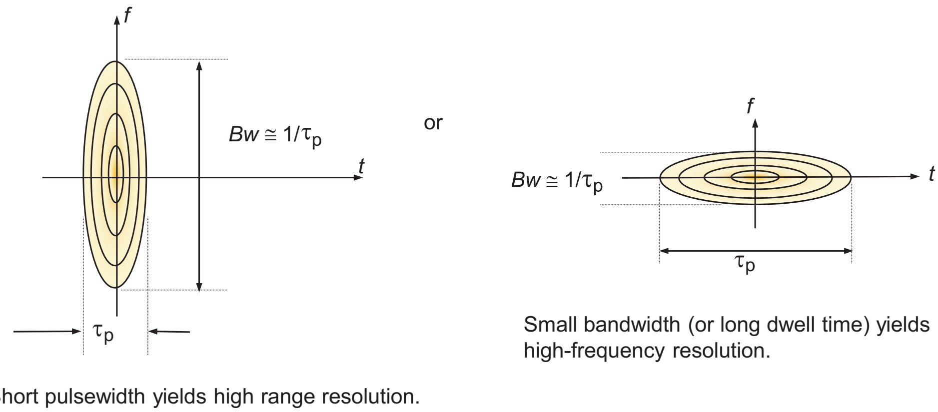


Figure 7-6 Frequency- and time-domain illustration of the width of the ambiguity function of a square pulse.

frequency direction, yielding poor frequency resolution. If we squeeze the frequency axis, the balloon bulges in the time axis, yielding poor resolution there. When using a simple pulse waveform, either range or velocity (time or frequency) resolution fidelity is possible but not both simultaneously. We see that this ambiguity function, which can be very narrow in one direction, would fall under the knife-edge type. This trade is a fundamental feature of Fourier theory and can also be derived from the Heisenberg uncertainty principle.

7.2.3 Example: Linear frequency-modulated (LFM) chirp

During the early development of radar, scientists quickly realized the time-frequency limit of a simple pulse waveform. A slightly more complex waveform was soon discovered, the linear frequency-modulated (LFM) chirp pulse, which seemed to overcome this limitation. The previous example used a pulse that was at zero carrier frequency; however, the same analysis could easily be carried out at some offset frequency or carrier tone, which would shift the ambiguity function (shown in the example in Fig. 7-4) in the frequency axis by the carrier magnitude. However, rather than using a simple tone for the duration of the pulse, the frequency can be swept, as shown in Fig. 7-7, creating a frequency-modulated (FM) chirp (the term "chirp" comes from the audio frequency sweep that birds make).

A LFM chirp is characterized by the pulsewidth T_p and the total frequency chirped B. The chirp is called "linear" because if we plotted the instantaneous frequency (IF) against time, we would see a linear ramp. To understand the benefit of this waveform, I will first show how it was used in both early radar and some early ladar sensors. The LFM chirp was detected with a matched filter, which was a dispersive delay line. In the early implementations, we used an acoustic delay line with reflective electrodes that

Maurice J. Halmos is a Principal Engineering Fellow working for more than 30 years in laser & ladar technology and computational active imaging at Raytheon. He recently acted as Raytheon's Principal Investigator for the U.S. Air Force Research Laboratories' synthetic aperture ladar (SAL) programs and was a past Technical Director for DARPA's laser vibration sensor demonstration of airborne SAL. He has also served as Raytheon's Technical Lead of other closed programs.

Dr. Halmos is a Fellow of the Military Sensing Symposia (MSS), past Chair of the MSS Active EO Group, twice recipient of Raytheon's Excellence in Technology award, and a member of the former technical advisory council of the U.S. Air Force's Ladar and Optical Communication Institute. He obtained a BSEE degree from Massachusetts Institute of Technology, an MSEE degree from Technion—Israel Institute of Technology, and a Ph.D. in EE, Quantum Electronics from University of California, Los Angeles.