Laser Radar: Progress and Opportunities in Active Electro-Optical Sensing

ISBN
978-0-309-30216-6

310 pages
8.5 x 11
PAPERBACK (2014)

Committee on Review of Advancements in Active Electro-Optical Systems to Avoid Technological Surprise Adverse to U.S. National Security; Division on Engineering and Physical Sciences; National Research Council

Laser Radar

Progress and Opportunities in Active Electro-Optical Sensing

Committee on Review of Advancements in Active Electro-Optical Systems
to Avoid Technological Surprise Adverse to U.S. National Security

Division on Engineering and Physical Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This study was supported by Contract HHM402-10-D-0036-DO#10 between the Department of Defense and the National Academy of Sciences. Any views or observations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-30216-6
International Standard Book Number-10: 0-309-30216-1

Limited copies of this report are available from:

The National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001, (800) 624-6242 or (202) 334-3313, http://www.nap.edu

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON REVIEW OF ADVANCEMENTS IN ACTIVE ELECTRO-OPTICAL SYSTEMS TO AVOID TECHNOLOGICAL SURPRISE ADVERSE TO U.S. NATIONAL SECURITY

PAUL F. McMANAMON, University of Dayton, *Chair*
WALTER F. BUELL, The Aerospace Corporation, *Vice Chair*
MELISSA G. CHOI, Massachusetts Institute of Technology
JOHN W. DEVITT, Raytheon Vision Systems
ELSA GARMIRE, Dartmouth College
GARY W. KAMERMAN, FastMetrix, Inc.
KENNETH A. KRESS, KBK Consulting, Inc.
JEANETTE LURIER, Raytheon
PRADIP MITRA, DRS Technologies
PETER F. MOULTON, Q-Peak, Inc.
JONATHAN M. SMITH, University of Pennsylvania
ABBIE WATNIK, Naval Research Laboratory
ELI YABLONOVITCH, University of California, Berkeley

Staff

TERRY JAGGERS, Board Director
GREGORY EYRING, Study Director
DANIEL E.J. TALMAGE, JR., Program Officer
SARAH CAPOTE, Research Associate (until March 2013)
DIONNA ALI, Research Assistant
CHRIS JONES, Financial Associate

Preface

In today's world, the range of technologies with the potential to threaten the security of U.S. military forces is extremely broad. These include developments in explosive materials, sensors, control systems, robotics, satellite systems, and computing power, to name just a few. Such technologies have not only enhanced the capabilities of U.S. military forces, but also offer enhanced offensive capabilities to potential adversaries—either directly through the development of more sophisticated weapons, or more indirectly through opportunities for interrupting the function of defensive U.S. military systems. Passive and active electro-optical (EO) sensing technologies are prime examples.

In 2010, the National Research Council (NRC) published the report *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*. That report focused on key passive sensor technologies and concluded that detector technology was nearing background-limited infrared photodetection (BLIP) for many tactical scenarios, and that therefore new detectors were unlikely to provide any "surprise" technologies.

This report builds upon and expands the scope of the 2010 report by considering the potential of active electro-optical (EO) technologies to create surprise; i.e., systems that use a source of visible or infrared light (typically but not always a laser) to interrogate a target in combination with sensitive detectors and processors to analyze the returned light. The addition of an interrogating light source to the system adds rich new phenomenologies that enable new capabilities to be explored.

In late 2011, the intelligence community, with the U.S. Army as the lead, approached the NRC to conduct a study to evaluate the potential of active EO systems to generate technological surprise. In response, the NRC formed the ad hoc Committee on Review of Advancements in Active Electro-Optical Systems to Avoid Technological Surprise Adverse to U.S. National Security, and the study contract was signed in September of 2012. The statement of task given to the committee is as follows:

> The NRC ad hoc committee will:
>
> - Evaluate the fundamental, physical limits to active electro-optical (EO) sensor technologies with potential military utility; elucidate tradeoffs among technologies including: direct and heterodyne systems, scanning and flash ladar, Geiger mode, linear mode, and polarization based ladar, synthetic aperture vs. real beam ladar; and parameters including sensitivity, dynamic range, polarization sensitivity, etc. Compare these limits to the near term state-of-the-art, identifying the scaling laws and technical and other impediments currently restricting progress.
> - Identify key technologies that may help overcome the impediments within a 5-10 year timeframe, the implications for future military applications, and any significant indicators of programs to develop such applications. Speculate on technologies and applications of relevance that are high impact wildcards with feasible deployment within 10 years. Discuss available laser illumination technologies, including wall-plug efficiency. Femtosecond pulse width laser sources should be considered. Discuss available detector/receiver approaches and technologies. Discuss laser beam steering approaches. Discuss processing approaches to convert ladar data into useable information.
> - Consider the pros and cons of implementing each existing or emerging technology, such as noise, dynamic range, processing or bandwidth bottlenecks, hardening, power consumption,

weight, etc. Identify which state and non-state actors currently lead worldwide funding, research, and development for the key technologies. Highlight the scale, scope, and particular strengths of these research and development efforts, as well as predicted trends, timescales, and commercial drivers.

- Evaluate the potential uses of active EO sensing technologies, to include 3D mapping and multi-discriminate laser radar technologies. Laser vibration detection, atmospheric compensation, multiple illumination wavelengths, polarization, and speckle considerations should be included as methods of determining object identity and status.

A report will be authored by the committee addressing the foregoing tasks.

This has been a challenging effort because of the breadth of active EO sensing modalities and contributing technologies. A further complication is that discussion of military or dual-use applications of a technology is always limited by classification issues or other restrictions. The main body of this report is unclassified. Where possible we tried to use unclassified, publicly available sources to discuss the areas covered in the statement of task. Some specific areas, however, are more sensitive and therefore are covered in classified appendixes not included in this version of the report. In some cases, information restricted by International Traffic in Arms Regulations is included with the classified material for both context and simplicity.

This report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise in accordance with the procedures approved by the Report Review Committee of the NRC. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We wish to thank the following individuals for their review of the report: Steven R. Brueck, University of New Mexico, Joseph Buck, Fieldcraft Scientific, Ronald G. Driggers, Naval Research Laboratory, James R. Fienup (NAE), University of Rochester, Robert Q. Fugate (NAE), New Mexico Institute of Mining and Technology, William Happer (NAS), Princeton University, Sumanth Kaushik, MIT Lincoln Laboratory, Dennis K. Killinger, University of South Florida, Paul D. Nielsen (NAE), Software Engineering Institute, and Julie J.C.H. Ryan, George Washington University,

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Edwin P. Przybylowicz, Eastman Kodak Company. Appointed by the NRC, he was responsible for making certain that an independent examination of this report was carried out institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

The committee also thanks the NRC staff for its dedicated work, in particular Greg Eyring, the study director, and Dionna Ali, who managed the administrative and logistical aspects with grace and efficiency.

Paul McManamon, *Chair*
Walter Buell, *Vice Chair*
Committee on Review of Advancements in Active Electro-Optical Systems to Avoid Technological Surprise Adverse to U.S. National Security

Contents

Acronyms

1-D	one-dimensional
2-D	two-dimensional
3-D	three-dimensional
AFRL	Air Force Research Laboratory
A-GNR	armchair-edge-boundary graphene nanoribbons
ALIRT	Airborne Ladar Imaging Research Testbed
ALMDS	Airborne Laser Mine Detection System
ALS	airborne laser scanning
APD	avalanche photodiode
ASE	amplified spontaneous emission
BLIP	background-limited infrared photodetector
BTEX	benzene, toluene, ethylbenzene, xylene
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
CCD	charge-coupled device
CLEAR	Center for Lidar Environmental and Atmospheric Research
CMOS	complementary metal-oxide-semiconductor
COD	catastrophic optical destruction
CONOPS	concept of operations
COP	coefficient of performance
CPA	chirped pulse amplification
CPM	critical phase matching
DARPA	Defense Advanced Research Projects Agency
DAS	detector angular sub-tense
DAWN	Doppler aerosol wind lidar
DBR	distributed Bragg reflector
DIAL	differential absorption lidar
DISC	differential scatter lidar
DOP	degree of polarization
DPAL	diode-pumped alkali laser
DTED	digital terrain elevation data
DWELL	quantum dot in the well
EO	electro-optical
ESA	European Space Agency, also excited state absorption
ESPI	electronic speckle-pattern interferometry

FFT	fast Fourier transform
FLIR	forward looking infrared
FM	frequency-modulated
FMCW	frequency-modulated, continuous-wave
FOPEN	foliage penetration
FOV	field of view
FPA	focal plane array
FSM	fast steering mirror
FTIR	Fourier transform infrared
FWHM	full width at half maximum
GM-APD	Geiger-mode avalanche photodiode
GPS	global positioning system
GPU	graphics processor unit
GQD	graphene quantum dot
GSD	ground sample distance
GVD	group velocity dispersion
HALOE	High Altitude Lidar Operational Experiment
HDVIP	high-density vertically integrated photodiode
HME	home-made explosive
ICCD	intensified charge-coupled device
ICL	interband cascade laser
IDCA	integrated detector/cooler assembly
IFF	identify friend or foe
IR	infrared
IR&D	internal research and development
ISR	intelligence, surveillance, and reconnaissance
ITAR	International Traffic in Arms Regulations
JHPSSL	Joint High-Power Solid-State Laser program
Ladar	laser detection and ranging
LASCA	laser speckle contrast analysis
LFM	linear frequency-modulated
LIBS	laser-induced breakdown spectroscopy
Lidar	light detection and ranging
LIF	laser-induced fluorescence
LIMARS	Laser Imaging and Ranging System
LM-APD	linear-mode avalanche photodiode
LO	local oscillator
LOCAAS	Low Cost Autonomous Attack System
LPE	liquid phase epitaxy
LR-BSDS	Long Range Biological Standoff Detection System
LWIR	long-wavelength infrared
MCDS	multiple correlated double sampling
MSM	metal semiconductor metal
MBE	molecular beam epitaxy

MIMO	multiple input, multiple output
MIT	Massachusetts Institute of Technology
MLE	maximum likely estimation
MOVPE	metal organic vapor phase epitaxy
MPE	maximum permissible exposure
MWIR	mid-wavelength infrared
NA	numerical aperture
NASA	National Aeronautics and Space Administration
NCPM	non-critical phase matching
NDVI	normalized difference vegetation index
NEI	noise equivalent input
NEPh	noise equivalent photons
NETD	noise equivalent temperature difference
NIF	National Ignition Facility
NIIRS	national imagery interpretability rating
NIR	near infrared
NIST	National Institute of Standards and Technology
NPL	National Physics Laboratory
NRC	National Research Council
NRI	negative refractive index
OAWL	optical autocovariance wind lidar
OPA	optical parametric amplifier
OPD	optical path difference
OPG	optical parametric generator
OPL	optical phase-locked loop
OPO	optical parametric oscillator
OSA	optical spectrum analyzer
PC	photonic crystal
PED	processing, exploitation, and dissemination
PIN	p-doped-intrinsic (undoped)-n-doped
PMT	photomultiplier tube
PNR	polarization non-reciprocity
PRF	pulse repetition frequency
PSD	power spectral density
QCL	quantum cascade laser
QCW	quasi-continuous wave
QD	quantum dot
QDIP	quantum dot infrared photodetectors
QE	quantum efficiency
QPM	quasi-phase matching
QWIP	quantum well infrared photodetector
R&D	research and development
Radar	radio detection and ranging
RER	relative edge response
RF	radio frequency
RHI	range-height indicator
ROIC	read out integrated circuit

RRDI	range-resolved Doppler image
RULLI	remote ultra-low light level imaging
SAL	synthetic aperture laser ladar
SALTI	Synthetic Aperture Lidar Tactical Imaging
SAR	synthetic aperture radar
SBS	stimulated Brillouin scattering
SC	supercontinuum
SCM	single carrier multiplication
SERS	surface-enhanced Raman spectroscopy
SL	superlattice
SNCR	signal-to-noise and clutter ratio
SNL	Sandia National Laboratories
SNR	signal-to-noise ratio
SPGD	stochastic parallel gradient descent
SPI 3-D	Standoff Precision Identification in Three Dimensions
SPIDAR	spatially processed image detection and ranging
SPM	self-phase modulation
SPS	spectral pattern sampling
SR	spectral reflectance
SRL	scanning Raman lidar
SVI	spectral vegetation index
SVM	support vector machines
SWaP	size, weight, and power
SZ	surf zone
TCPED	tasking, collection, processing, exploitation, and dissemination
TDLAS	tunable diode laser absorption spectroscopy
TEC	thermoelectric cooler
TE-IPD	transferred-electron intensified photodiode
TERCOM	terrain contour matching
ToF	time of flight
UAV	unmanned aerial vehicle
UV	ultraviolet
VAD	velocity azimuth detector
VCSEL	vertical-cavity surface-emitting diode laser
VIPA	virtually imaged phased array
VSW	very shallow water
WANDER	wavelength normalized depolarization ratios
YAG	yttrium aluminum garnet
ZDW	zero dispersion wavelength

Summary

Active electro-optical (EO) sensing systems employ a transmitted optical field (usually from a laser) that interacts with a distant object. The scattered light is then sensed by a receiver system to derive information about the object. Thus, active EO systems are similar to microwave radar systems but operate in the visible or infrared region of the electromagnetic spectrum. As a result of the shorter wavelengths of infrared and visible light, active EO systems have much higher (>20,000 times) angular resolution and range precision than radar systems. They also benefit from controlling the illumination, which provides an advantage over passive EO systems. For example, this allows active EO sensing to be day/night-capable at short wavelengths, and provides greater control over shadowing phenomenology. Thus, active EO systems enjoy resolution advantages over microwave radar and illumination advantages over passive EO systems.

This combination of features enables new, unique, and operationally important commercial, military, and intelligence applications. An example of an ideal application is the ability to provide imagery much more similar to what the eye is used to seeing than radar imagery, but at ranges beyond the capability of passive EO sensors. Three dimensional (3-D) mapping is an important commercial and military application of active EO sensing that can provide critical information (e.g., detailed 3-D scene data) not readily available from alternative technologies. It is anticipated that a global, 3-D database derived from active EO sensing measurements will support civil, commercial, and military needs as the foundational layer for organizing geospatial information of all types. One potentially transformational scientific, commercial and military application of active EO sensing is accurate mapping of world wind directions and velocities from space, enabling more accurate weather prediction, more efficient routing of aircraft, and improved accuracy of long-range weapons. Another example is inexpensive 3-D imagery capable of providing the necessary information to guide a driverless car or to control a robot. Active EO sensing is ideal for highly sensitive vibration detection that can tell if power is flowing through a transformer or liquid is flowing through a pipe or that can remotely tell how many cylinders a vehicle has or, possibly, if the vehicle needs a tune up.

The committee's key conclusions and recommendations below are taken from the body of the report and are numbered according to where they appear in the chapters in order to point the reader to a more detailed discussion.

Conclusion 1-1: Active EO sensing enables measurements not possible with passive EO sensors by exploiting the control of the illumination. Exploiting control of the illumination includes its coherence properties.

Conclusion 1-2: Active EO sensing offers many of the controlled illumination advantages of radar and, by operating at much shorter wavelengths, enables greater range precision and angular resolution as well as more literal interpretability.

Conclusion 2-18: Active EO sensing is poised to significantly alter the balance in commercial, military, and intelligence operations, as radar has done over the past seven decades.

The close analogy between conventional radar and active EO sensing can serve as a guide to the trajectory of development for active EO sensing. Since its invention in The World War II as a game-

changing technology, radar has developed a mature technology and exploitation base with novel ways of overcoming limitations and has proliferated to a wide variety of military and commercial applications. Many of the techniques developed for radar have analogs in active EO sensing, such as advanced waveforms and synthetic aperture imaging, and are gradually being implemented in active EO sensing. Our knowledge of the diverse applications of radar technology has enabled a more diverse set of active EO sensing modes in a relatively short time frame.

There are two main types of optical active EO sensing: (1) direct detection, which measures only the intensity of the return signal, and (2) heterodyne—or coherent—detection, which measures the electromagnetic field of the returned signal, including both phase and intensity.[1] Direct detection can produce traditional two-dimensional (2-D) intensity images similar to passive EO images. 2-D intensity images generated by active EO systems can be enhanced with additional degrees of freedom in similar ways to the enhancements that occur with passive EO imagers—namely, with polarization, color, or temporal modulation. What is different in active sensing is that the sensor provides control over the illumination (timing, waveforms, spectral properties, polarization, etc.) One consequence of this controlled illumination is that distance can be measured using the speed of light and a clock to measure the round trip time to an object and back. This makes it easy to create a 3-D image of an area, such as Figure S-1 (Figure 2-11 in the report).

The same clock can be used to turn the detector on and off, so that only backscatter in the range near the object of interest is viewed; that is to say, the ability to range-resolve a target includes the ability to range-gate out intervening scatterers, so long as sufficient illumination reaches the target and returns to the sensor. Eliminating scatter and clutter, in front of and behind/below the object of interest is an important capability of active EO sensing (Figure S-2).

Although the transmitted photons that active EO uses (typically a laser) for sensing are not free and add complexity to the system, the advantage is that one can have as many photons on the object as one is willing to pay for. There are no shadows in monostatic,[2] active EO imaging for any point to which the sensor has line-of-sight access.

Usually, the light used in active EO sensing has a narrow wavelength band. Interfering narrow-band scattered light from different portions of a rough surface creates light and dark areas at the receiver called speckle. Speckle is a noise source in many active EO sensing applications, but speckle also carries information about the target that can be exploited. Both surface scattering and volume scattering objects are of interest. Active EO sensing has a wide range of volume scattering applications for sensing wind, aerosols, and chemical and biological clouds of gas. Both volume and surface scattering can also be applied underwater. The committee found that discussing all of the components that comprise active EO sensing, including the systems architectures and the phenomenologies associated with active EO sensing, has been a substantial undertaking.

Heterodyne (also called coherent) detection uses a sample of the outgoing laser pulse, the local oscillator (LO), to beat against the return signal so that phase can be measured as well as intensity. Coherent detection can even measure the minute Doppler shifts associated with scattering from vibrating objects, a technique called laser vibrometry. Aperture synthesis techniques, such as synthetic aperture ladar (SAL) and multiple input, multiple output (MIMO) techniques also use coherent methods because they need to capture samples of the full pupil plane field, including both phase and intensity across the pupil plane.

[1] As illustrated in Figure 1-3, there are also a variety of lidar techniques that utilize spectroscopic interactions of the laser illuminator with the target volume, such as Raman or Rayleigh scattering, fluorescence, etc.

[2] An active EO sensor can be monostatic, meaning the laser illumination is emitted from the same optical aperture that receives it, or it can be bistatic, meaning separate apertures are used for the transmitter and the receiver.

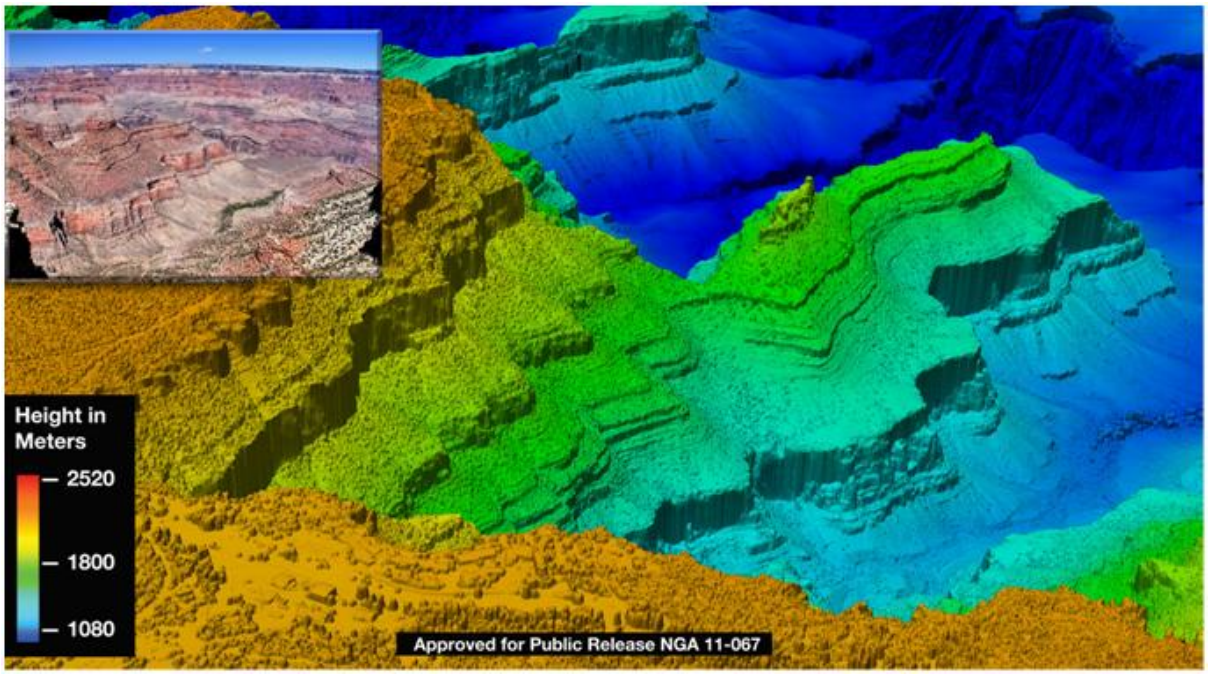

FIGURE S-1 3-D active EO imagery of the Grand Canyon with a photograph of the same region inset. Note the dramatic change in color, therefore range, from light green to dark blue along the ridge. This dramatic range change is not obvious in the passive imagery. SOURCE: Lincoln Laboratory MIT, *Tech Notes: Airborne Ladar Imaging Research Testbed,* http://www.ll.mit.edu/publications/technotes/ TechNote_ALIRT.pdf, 2011. Reprinted with permission of MIT Lincoln Laboratory, Lexington Massachusetts.

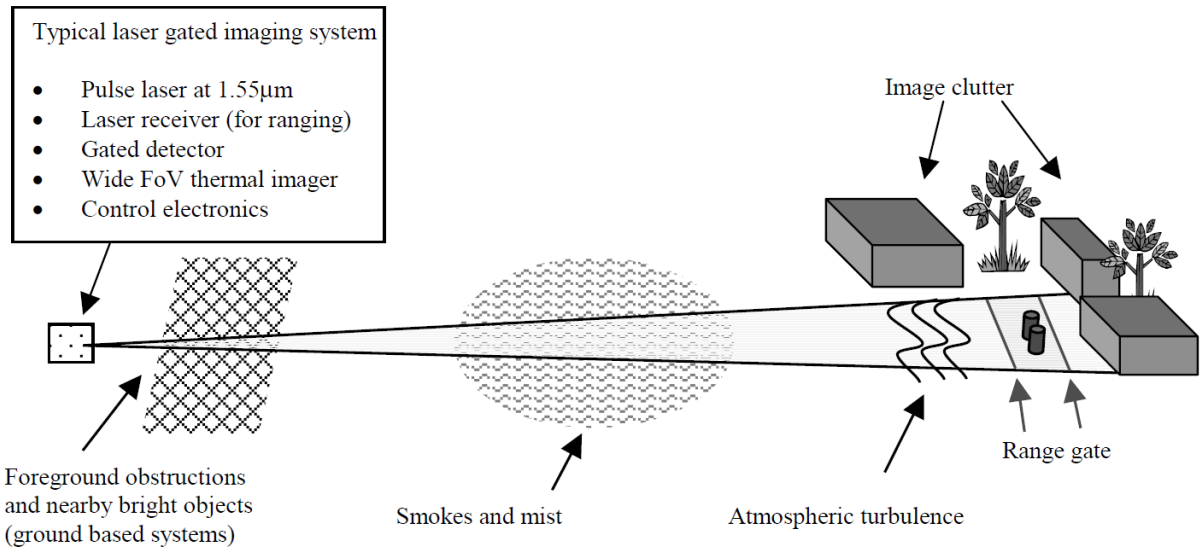

FIGURE S-2 A gated active EO imaging system can eliminate clutter and obscurants that might otherwise degrade the image. Although laser returns from image clutter, smoke, mist and foreground obstructions can be range-gated, effects of atmospheric turbulence remain. SOURCE: I. Baker, S. Duncan, and J. Copley, 2004, "A Low Noise, Laser-Gated Imaging System for Long Range Target Identification," *Proceedings of SPIE*, 5406. Courtesy of Selex ES Ltd.

Along with the rapid development of sensing modalities, the components for active EO sensing, such as lasers, detectors, optics, and processing, are maturing. The committee expects new and unexpected application areas to develop on a regular basis over the next decade. It anticipates global interest in this technology, driven by many diverse needs. Active EO sensing is a very dynamic area of R&D and has applications in Europe and Asia as well as in the United States (publication data for the past decade are presented in Chapter 1, Figure 1-13). China, for example, has severe air pollution,[3] so it would be natural for it to invest in active EO technologies to measure air pollution parameters.

The United States leads in the development of high-performance active EO systems in the military arena. However, significant EO technologies are now being developed globally. For example, most commercial mapping and surveying systems are produced outside the United States, and all the DARPA Grand Challenge (2004-2005) autonomous vehicles used systems manufactured by the German company SICK because of their low cost and availability. In the literature, Europe has demonstrated a strong capability in a variety of military active EO sensors. Many foreign-based component laser and detector technologies are more advanced than their U.S. counterparts for application in both commercial and high-end active EO sensors.

Complex systems, whether commercial or military, require significant time and resources to develop. Systems developed just for the military require significant resources. The commercial sector may bear the development costs for those active EO sensing technologies with significant commercial applications. Volume manufacturing can dramatically reduce production costs. The potential for inexpensive, high-volume systems makes the robotic application of active EO sensing a potential game changer.

Conclusion 2-16: Commercial applications for active EO sensing will become widespread, dramatically increasing the use of active EO sensing and significantly lowering the cost.

Conclusion 2-17: Robotics—for example, autonomously navigated vehicles—is likely to be a dominant commercial and military application for active EO sensing.

Conclusion 2-19: Most active EO sensors used for military and security applications will fall into one of two categories: (1) adaptation of inexpensive commercial devices to military applications, or (2) development of unique active EO sensing systems for the military. The first category is accessible to nonstate actors and resource-poor countries. The second category is only accessible to countries with significant resources.

High-performance, specialized active EO sensing devices and systems require significant investment to develop, pose difficult engineering challenges to deploy, and require persistence and a long lead time to field practical systems. This capability requires developing and maintaining a large infrastructure of related technologies and is expensive and time consuming.

Conclusion 2-20: Development programs for high-end active EO sensors will involve observable activities or indicators of the direction these programs are taking.

Conclusion 2-21: Widespread, commercially available active EO technology offers a low barrier to entry for asymmetric adversaries with limited resources as well as for those seeking to adapt the technology for large-volume applications.

Recommendation 2-1: Analysts assessing states of technological development and projected timelines for developments of military significance should pay close attention to activity in the commercial sector.

[3] See http://www.ft.com/cms/s/0/eed7c0be-e7ca-11e2-9aad-00144feabdc0.html.

As a final note, the United States imposes certain very specific standards on the use of its sensors. For example, the United States will not deploy a sensor system that has a significant chance of blinding a person. Under the usual offensive rules of engagement, the United States also does not permit a kill decision to be made without the involvement of a responsible human decision maker, as would be possible with robotic attackers. It is not obvious, however, that a potential adversary would employ the same operational constraints that the United States employs. The uses of active EO sensing could be a technological surprise if a state or nonstate actor were to impose fewer ethical or policy constraints on an EO-sensor-equipped weapon system.

Recommendation 1-1: To avoid the possibility of technological surprise, the U.S. intelligence community should consider all potential uses of active electro-optical sensing, including those not allowed by the ethics and policies of the United States.

CONCLUDING THOUGHTS AND OVERARCHING CONCLUSION AND RECOMMENDATION

Recent advances in laser illuminators, sensitive broadband detectors, and rapid data processing are combining to enable active electro-optical (EO) sensing to bring revolutionary advances in areas such as mapping, targeting, autonomous robotics, environmental monitoring, weather prediction, and intelligence gathering. In the context of this committee's task to focus on those areas of active EO sensing that could produce technological surprise adverse to U.S. national security, the committee arrived at the following overarching conclusion and recommendation:

Overarching Conclusion: Active electro-optical sensing is a rapidly emerging technology with many applications across intelligence, military, scientific, and commercial domains. It has the potential to alter the balance in many of these areas. Interest in developing and applying active electro-optical sensing has consequently risen in other nations, putting U.S. leadership at risk, to the degree that in some instances the United States no longer leads.

Overarching Recommendation: Uses and development worldwide of active electro-optical (EO) sensing should be tracked aggressively by the U.S. intelligence community, and active EO sensing should be aggressively developed by the United States.

1

Introduction

Active electro-optical (EO) sensors can measure range, velocity, vibrations, intensity, phase, polarization, and angular position. They can provide literal imagery that can be easily interpreted at ranges beyond what passive EO sensors can provide. They can provide near-photographic quality literal imagery, three-dimensional (3-D) imagery, and other discriminants, even in some cases when the target is obscured by camouflage, foliage, or weather. The richness of the information collected gives rise to a very wide range of applications. Particularly if they are not anticipated, some of these applications could be very damaging to U.S. national interests if pursued by an adversary. Conversely, they can be an advantage if used by the United States.

Active and passive EO sensing can be compared to radar sensing by looking at the electromagnetic spectrum shown in Figure 1-1. Active EO sensing systems have a higher angular resolution than microwave radar systems due to the shorter wavelengths of infrared and visible light and, like radar, also benefit from controlling the illumination. In that respect, active EO systems enjoy resolution advantages over microwave radar and illumination advantages over passive EO sensors.

Some types of active EO sensing go by the name ladar (LAser Detection And Ranging) and sometimes by lidar (LIght Detection And Ranging). Use of these terms is not standardized from one subcommunity to the next (Box 1-1). The National Institutes of Standards and Technology (NIST) has adopted ladar as the standard name for active EO sensing, while the U.S. Department of Defense (DoD) has tended to use lidar. Historically, conventional usage has made a distinction based on the type of target—ladar for "hard" targets (e.g., solids) and lidar for "soft" targets (e.g., gases or aerosols). In this report, the committee followed a similar convention in that active EO sensing of targets with surface scattering is called ladar and active EO sensing of target volume scattering is called lidar. The committee felt that distinguishing between surface or volume scattering was less ambiguous than distinguishing between soft or hard targets.

If one compares a typical 200-THz, 1.5-μm wavelength, active EO system to a typical 10-GHz, 3-cm-wavelength radar system, the EO angular resolution will be nominally 20,000 times finer than that of radar for the same aperture diameter. For a synthetic aperture radar compared to a synthetic aperture ladar, this means flying 20,000 times shorter distance to form the synthetic aperture, but requiring motion compensation that is 20,000 times more accurate. Range resolution can also be higher than microwave radar because the high carrier frequency of active EO sensing allows very high bandwidth modulation to still be a small fractional bandwidth of the carrier. Range resolution depends on the bandwidth of the transmitted signal, as discussed later. Using these same wavelength ratios, the time required for an EO measurement of target velocity is also shorter than for microwave radar by a factor of 20,000. The benefit of active systems as opposed to passive systems is that they provide illumination of the target that does not vary over time unless it is varied intentionally. This allows active EO sensing to be day- and night-capable at short wavelengths and provides greater control over shadowing phenomenology.

In fairness, active EO sensing will not have advantages over passive systems and microwave radar in all cases. Passive EO systems are stealthier and avoid the risk that an illuminator could reveal the location of the system. They do not raise eye-safety issues (see Box 1-4) and may have advantages in terms of lower life cycle costs, lower power requirements, and simplicity. Similarly, radar systems can be

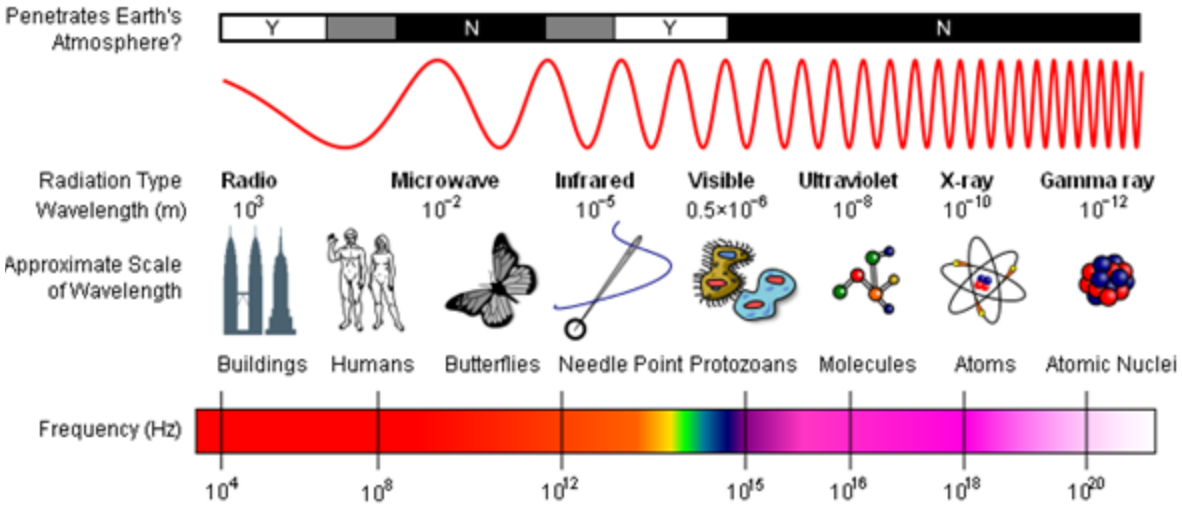

FIGURE 1-1 The electromagnetic spectrum, showing the roughly 20,000-fold difference between the frequencies used in active EO sensing (visible and infrared) and traditional radar (microwave). SOURCE: By Inductiveload, NASA [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/)], via Wikimedia Commons. See http://en.wikipedia.org/wiki/File:EM_Spectrum_Properties_edit.svg.

BOX 1-1
Names Used within the Active EO Sensing Community

- Lidar (Light Detection and Ranging)
  - Has historically been used with atmosphere, or chemical vapor, detection;
  - Is used by the National Geospatial Intelligence Agency for active EO sensing (3-D imaging and mapping); and
  - Is usually used for commercial applications.
  - This report uses "lidar" for active EO sensing of volume scattering targets.
- Ladar (Laser Detection and Ranging)
  - Has historically been used with hard targets and has been
  - Adopted by NIST at the standard term for active EO sensing.
  - This report uses "ladar" for active EO sensing of surface scattering targets.
- Laser radar and laser remote sensing are sometimes used.
- For reference: radar is Radio Detection And Ranging.

very nearly all-weather, and can penetrate clouds better than optical sensors. Radars also have more favorable search detection statistics, and so might be superior for early-warning applications.

Figure 1-2 is a highly simplified diagram of a ladar. It shows a laser generating photons, which are transmitted to an exit aperture to illuminate a target. Light is scattered from the target and returned to the same aperture.[1] The returned light is routed to a detector array and then processed. Some ladars extract a sample of the outgoing signal and interfere it against the return signal in order to obtain phase information.

[1]This is called a monostatic system. In a bistatic system, separate apertures are used to transmit and receive the optical radiation.

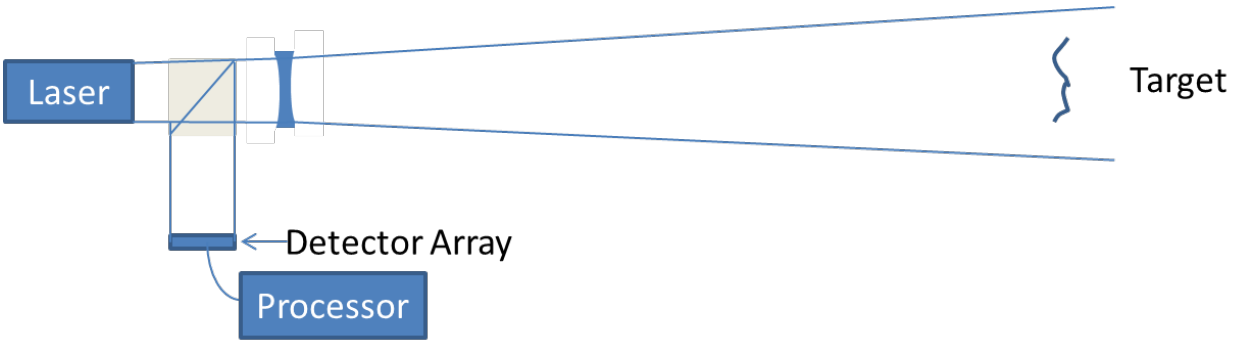

FIGURE 1-2 Highly simplified block diagram of a ladar.

Active EO sensing modalities are very rich and, as previously mentioned, will support a very wide range of measurements. The field of active EO sensing can be subdivided with respect to wavelength of operation, detection technique employed, measurements to be performed, or application. One detailed taxonomy is shown in Figure 1-3, where these modalities are divided into three categories: (1) imaging, (2) spectral sensing, and (3) other, including sensing of vibrational and volume scattering. These modalities are discussed in greater detail in Chapters 2 through 4.

Range resolution, accuracy, and precision (see Box 1-2) of active EO sensors are not dependent on the diffraction limit, so they do not directly benefit from the wavelength difference compared to microwave radar. Instead, range resolution depends on the bandwidth of the transmitted and received electromagnetic signal. This is the same as with microwave radar; however, in ladar the carrier frequency is orders of magnitude larger. Because of this, it is possible to use much higher bandwidth if desired and thus to obtain much higher target range resolution. For example, a recent paper[2] showed that with 3.6-THz transmit bandwidth (about 1.8 percent of the carrier frequency), a 50-μm range resolution could be achieved using ladar. Other aspects of ladar performance are discussed in Box 1-3.

The laser illuminators in active EO sensing systems can raise eye safety issues for humans or animals in the beam path, as discussed in Box 1-4. Lasers operating in the wavelength region from about 400 nm to 1500 nm (blue to short wave infrared) are considered less eye-safe than lasers operating in other regions because the eye efficiently focuses light on the retina in this wavelength range, with the potential for damage to the retina. Lasers that operate outside of this wavelength range are sometimes referred to as eye-safe, even though eye-safe really only means a higher maximum permissible exposure (MPE).

A number of countries have developed commercial ladar systems at shorter wavelengths (830-850 nm) with a novel approach to eye safety. An example is the Elbit system.[3] Eye safety is handled using lower power interrogation pulses such that if no object is determined to be in the beam, the laser power is increased until an object is determined to be in the beam. These systems have a higher resolution than eye-safe systems due to the shorter wavelength. Such approaches may be particularly appropriate for shipboard and aviation platforms where the eye-safe "keepout" zone can be assured with altimeter devices and underway status.

[2] R.R. Reibel et al., 2009, "Ultrabroadband optical chirp linearization for precision length metrology applications," *Optics Letters*, 34: 3692-3694.

[3] O. David, R. Schneider, and R. Israeli, 2009, "Advance in active night vision for filling the gap in remote sensing," *Proc. SPIE* 7482, Electro-Optical Remote Sensing, Photonic Technologies, and Applications III, 748203; doi:10.1117/12.830378.

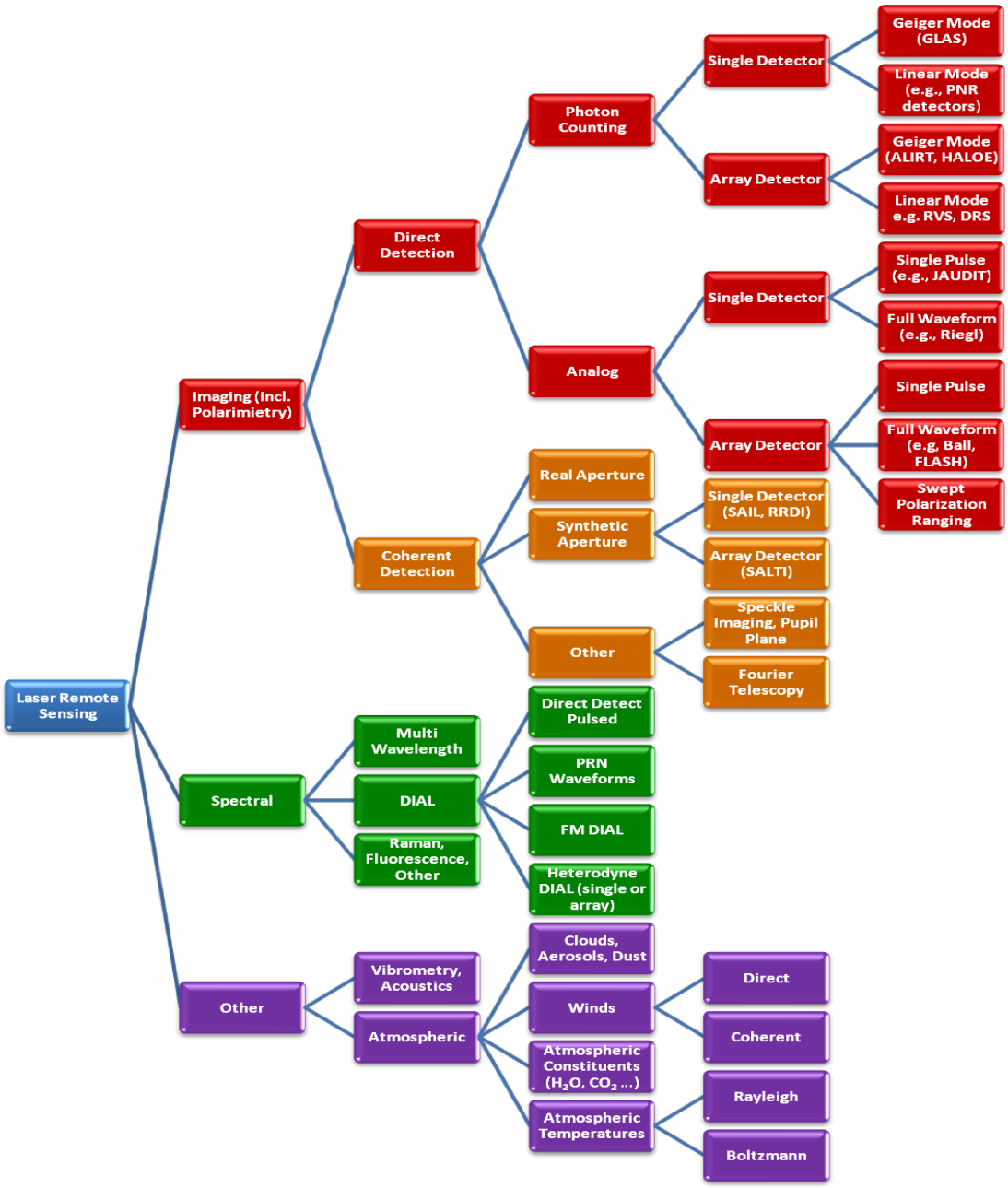

FIGURE 1-3 Schematic showing various modalities of laser remote sensing, divided broadly into imaging, spectral sensing, and other, including sensing of vibrational and volume scattering. "Imaging" here refers to both 2-D and 3-D active imaging, where range is determined by one form or another of time of flight measurement. JAUDIT: Jungle Advanced Under Dense vegetation Imaging Technology, an NGA-developed airborne linear mode lidar. It later became TACOP (tactical operational lidar) as a program of record; Riegl: Commercial company selling scanning linear mode lidars; Ball: Ball Aerospace; FLASH: A term used in the community generically denoting modest arrays of linear-mode (conventional large signal) APDs for lidar; SAIL: The Aerospace Corporation's demonstration of Synthetic Aperture Imaging Ladar using FMCW (Frequency Modulated-Continuous Wave) waveforms. The same basic technology as used in the Raytheon implementation of the DARPA SALTI program; SALTI: Synthetic Aperture Lidar Tactical Imaging, a DARPA program to demonstrate synthetic aperture ladar from an airborne platform; RRDI: Range Resolved Doppler Imaging, a coherent ladar technique; PRN: Pseudo Random Number, a coded approach to correlation-based ranging; DIAL: Differential Absorption Lidar; PNR Photon Number Resolving (i.e., can measure 1, 2, 3, 4, ... photons, as opposed to Geiger mode that can count one at a time); GLAS: NASA's Geoscience Laser Altimeter, a ranging ladar flown on the first IceSat mission.

BOX 1-2
Explanation of Terms: Range Resolution, Precision, and Accuracy

Range Resolution

- Quantifies the ability to detect two objects separated in range along a single line of sight.
- Limited to c/(2 × *B*) where c is the speed of light and *B* is the system bandwidth.

Range Precision

- Quantifies the relative uncertainty of a range measurement to an object.
- Limited by the range resolution and the signal-to-noise ratio of the measurement. It can be significantly better than the range resolution.

Range Accuracy

- Quantifies the degree to which a range measurement yields the "true value" of the absolute range.
- Depends on range precision as well as systematic errors (e.g., clock rate, drift, and timing offsets)

BOX 1-3
The Laser Radar Range Equation

The performance of an active EO system can be predicted by use of the laser radar range equation. The ladar equation calculates the laser power collected by the receiver.[1] There are multiple forms of the ladar equation, depending on definitions and assumptions. One form is shown in equation 1.

$$P_R \approx P_T * \frac{\sigma}{A_{illum}} * \frac{A_{rec}}{\pi R^2} * \eta_{atm}^2 * \eta_{sys} . \quad (1)$$

where P_R = power received, P_T = power transmitted, σ = cross section in square meters, A_{illum} = area illuminated, A_{rec} = area of the receiver, R = range, η_{atm} = transmission efficiency through the atmosphere, η_{sys} = receiver system optical systems efficiency. It is assumed the object to be viewed is within the illumination area. "The power received is the power transmitted × two ratios of areas × appropriate efficiency terms."[1] "The first ratio of areas is the cross section divided by the illuminated area at the object plane."[1] The cross section considered is the reflection within a given range bin. "The second ratio of areas is the receiver aperture area divided by the effective average area illuminated by Lambertian reflection."[1] Radar practitioners, as well as some of the people working active EO sensing, use a different definition of cross section based on reflection from a gold ball. In that case, instead of π in the denominator, representing Lambertian scattering on return over an effective angle of π steradians, the committee uses 4π, representing specular scattering from a round ball over 4π steradians. The EO handbook uses this definition.[2] It is very important to be aware of the definitions used regardless of which form of the ladar equation is used.

[1] P.F. McManamon, 2012, "Review of ladar: a historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6), 060901.
[2] C.S. Fox, ed., *The Infrared and Electro-Optical Systems Handbook*, Volume 6: Active Electro-Optical Systems.

BOX 1-4
Eye Safety in Active EO Sensing Systems

Active EO sensing systems have the potential to cause eye damage to humans or animals exposed to the illuminating laser beam. The potential for damage depends on the light irradiance at the eye (for continuous wave systems) or pulse fluence, pulse duration, and pulse rate (for pulsed systems) as well as the wavelength. Such damage may include surface heating of the cornea if the laser light has a wavelength that is strongly absorbed, or a burn on the retina if sufficient light passes through the lens and is focused there.

Maximum permissible exposure (MPE) levels for laser light in the workplace are set by the U.S. Occupational Safety and Health Administration.[1] The American National Standards Institute also publishes laser safety standards (ANSI Z136.1). In evaluating the eye safety characteristics of procured active EO systems, at least one U.S. government agency uses a software code that takes system inputs and evaluates eye exposure against the OSHA standards.[2] Of course, observers in the beam path using binoculars or telescopes will experience a higher fluence than would the unaided eye, so eye safety for a person looking with binoculars would require a lower MPE.

Figure 1-4-1 shows the MPE levels for laser pulses of various durations as a function of wavelength. Note the MPE is much lower in the wavelength region from 400 nm to 1500 nm. This is because the eye efficiently focuses light on the retina in that wavelength range. Lasers that operate outside of that wavelength region are sometimes referred to as eye-safe, even though eye-safe really only means a higher MPE.

As can be seen in the Figure 1-4-1, lasers operating at wavelengths longer than about 1500 nm or shorter than about 400 nm are safer because molecular absorption occurring in the volume of the eye attenuates the beam and reduces the intensity of the light that would be focused on the retina.[3] There is a slight decrease in the MPE at wavelengths longer than 1500 nm because light is absorbed more strongly at longer wavelengths, and so is absorbed in a smaller portion of the eye volume. At 10,000 nm (approximately the wavelength of a CO_2 laser), the eye has strong absorption, so the light is absorbed near the surface in the cornea, and burns may result.

A second factor to consider in relation to eye safety is laser pulse width. Short pulses have higher peak power for the same energy per pulse, and so may cause more damage to the eye. As shown in Figure 1-4-1, at all wavelengths the MPE decreases with decreasing pulse length. Of course at a high enough power, a laser operating at any wavelength and pulse duration can cause damage, but the permissible exposure levels are lower in certain wavelength bands and for shorter pulses.

The wavelength chosen for any laser remote sensing system depends on: (1) the phenomenology being sensed (reflectivity/absorption, atmospheric transmission, spatial resolution, etc.), and (2) technology constraints, including compatibility with sensors, maturity, and eye safety. Whether or not the laser is eye-safe, the system should be designed to be eye-safe at the range of operation, possibly with hardware or software constraints to preclude operations in non-eyesafe conditions (say before the platform reaches operational altitude.) Some applications require operational wavelengths not

There are two main types of optical active EO sensing: (1) direct detection, which measures only the intensity of the return signal, and (2) heterodyne, or coherent, active EO sensing, which measures the field, including both phase and intensity.[4] Owing to its relative simplicity, the direct detection technology base is mature and proliferating rapidly. Direct detection ladars are widely used today to provide excellent 3-D images, maps, and infrastructure models for commercial, civil, and military applications. This has fostered the emergence of a sophisticated global enterprise to obtain and exploit a high-

[4] As illustrated in Figure 1-3, there are also a variety of lidar techniques that utilize spectroscopic interactions of the laser illuminator with the target volume, such as Raman or Rayleigh scattering, fluorescence, etc.

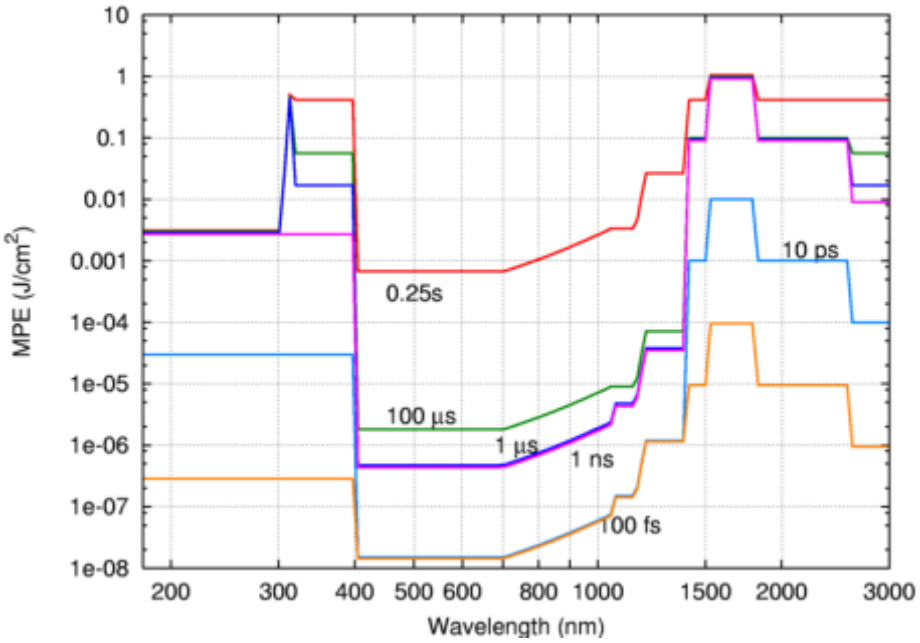

FIGURE 1-4-1 Maximum permissible exposure (MPE) levels for laser pulses of various durations as a function of wavelength. SOURCE: Hankwang at en.wikipedia [CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/), GFDL (http://www.gnu.org/copyleft/fdl.html), CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0/) or CC-BY-2.5 (http://creativecommons.org/licenses/by/2.5)], from Wikimedia Commons. See http://en.wikipedia.org/wiki/Laser_safety.

considered eye-safe (very near infrared spectroscopy, bathymetry, etc.). Nonetheless, it is often possible to design such systems to be eye-safe by choice of illumination, dwell time, receiver parameters, and concept of operations.

The committee notes that the active EO systems developed by potential U.S. adversaries may not operate under the same eye safety restrictions as U.S. systems. The implications of this possibility are discussed in the text.

[1] See https://www.osha.gov/dts/osta/otm/otm_iii/otm_iii_6.html#2.
[2] Dr. Richard Heinrichs, Program Manager, DARPA/STO, personal communication to the committee on October 17, 2013.
[3] See http://en.wikipedia.org/wiki/Laser_safety, downloaded October 15, 2013.

resolution database of the Earth.[5] The National Enhanced Elevation Assessment (NEEA), sponsored by the United States Geological Survey (USGS), has documented the enormous economic value to the United States of obtaining and maintaining this information. With over 600 applications documented in the NEEA,[6] it is clear that the exploitation of 3-D data is just beginning.

[5] The 2013 International Lidar Mapping Forum (ILMF) hosted attendees from over 30 countries.
[6] Dewberry, 2011, Final Report of the National Enhanced Elevation Assessment (revised 2012): Fairfax, Va., Dewberry, 84 pp. plus appendixes. Available at http://www.dewberry.com/Consultants/GeospatialMapping/FinalReport-NationalEnhanced ElevationAssessment.

Coherent laser radar was originally developed to mitigate the fact that detectors at CO_2 laser wavelengths were rather poor, and heterodyne detection was employed to overcome detector noise and reach the shot noise limit by using a powerful local oscillator. Later, the ability to measure small frequency shifts in coherent ladar was used to measure the velocity of moving targets, using both CO_2 and solid-state ladars. Since those early days, coherent ladar has blossomed to include ultraprecise range measurement, laser vibrometry, synthetic aperture ladar, spatial heterodyne imaging, ladar using multiple transmit and receive apertures, and to encompass a wide array of laser and detector technologies (see Chapter 3). Coherent ladar is now part of the engineering trade space for many active EO needs, and enables the use of inexpensive detectors and commercial off-the-shelf (COTS) laser components.

The reasons for selecting coherent ladar as a solution to achieve desired sensing goals have evolved as new sensing concepts emerge, and as a result the technology has evolved in new directions. One of the main reasons to use coherent ladar is to be able to capture the field (phase and intensity), rather than just the intensity, which is captured by a direct detection ladar. Given the dramatic expansion of coherent active EO in the past decade or so, the committee expects both applications and technological solutions to expand and proliferate.

Fiber lasers are emerging as powerful and flexible sources for active EO systems. As discussed in Chapter 4, fiber lasers are best suited to low peak power waveforms (e.g., high pulse-repetition-frequency pulse trains, modulated continuous-wave waveforms) The advances in high-sensitivity, low-noise detectors (Chapter 4) are thus synergistic with the advances in fiber laser technology as applied to active EO systems. The frequency agility and spectral quality of seed sources for fiber lasers enable a wide range of high bandwidth waveforms for both direct detection and coherent systems. Fiber lasers can also be combined, using a variety of coherent and incoherent beam combining techniques to produce high-power beams with the waveform quality of the modest-power individual fiber lasers. Finally, all of these properties are available with COTS or near-COTS components with heritage in the optical communications industrial base. These properties combined, in concert with either high-sensitivity, low-noise detector arrays or coherent detection, lead the committee to expect fiber laser solutions to become ever more prevalent in advanced active EO systems. This will also lower the barrier for entry into advanced active EO systems to groups with access to the COTS market.

Active EO sensing systems are sophisticated electro-opto-mechanical systems and are the product of a complex set of interdependent system engineering trades. The ultimate engineering solution for a given active remote-sensing problem will depend strongly on both the sensing problem at hand (parameters to be measured, operational range, propagation path, and so on) and any externally imposed constraints (system size, weight and power, operational environment, concept of operations, etc.). Given that the range of sensing possibilities and phenomenologies for active EO sensing is so much broader and richer than either passive EO sensing or radar, the committee expects the range of technological solutions available to the system designer to further broaden in the foreseeable future. One corollary to this situation is that observations of specific technology development choices and paths can provide significant insight into the interests and intentions of the technology developers.

An early example of a range-resolved image using a CO_2 laser is shown in Figure 1-4, which illustrates many of the important features of laser radar systems. As a result of the availability of long wave infrared detectors and optics, CO_2 lasers were the coherent illumination sources of choice for active EO systems from the early 1970s until the 1990s. Most CO_2 laser radars were at a wavelength of 10.6 μm, although some used different isotopes of CO_2 to avoid atmospheric absorption at that wavelength.[7]

The system that produced the image shown in Figure 1-4 achieved the angular resolution of an optical system and the range resolution and precision of a very advanced radar. Specifically, Figure 1-4 demonstrates milliradian angular resolution and better than 30 cm range precision through only a 2.5 cm clear aperture.

[7] P.F. McManamon, M. Huffaker, and G. Kamerman, 2010, "A history of laser radar in the United States," Proceedings of SPIE, 7684: 76840T.

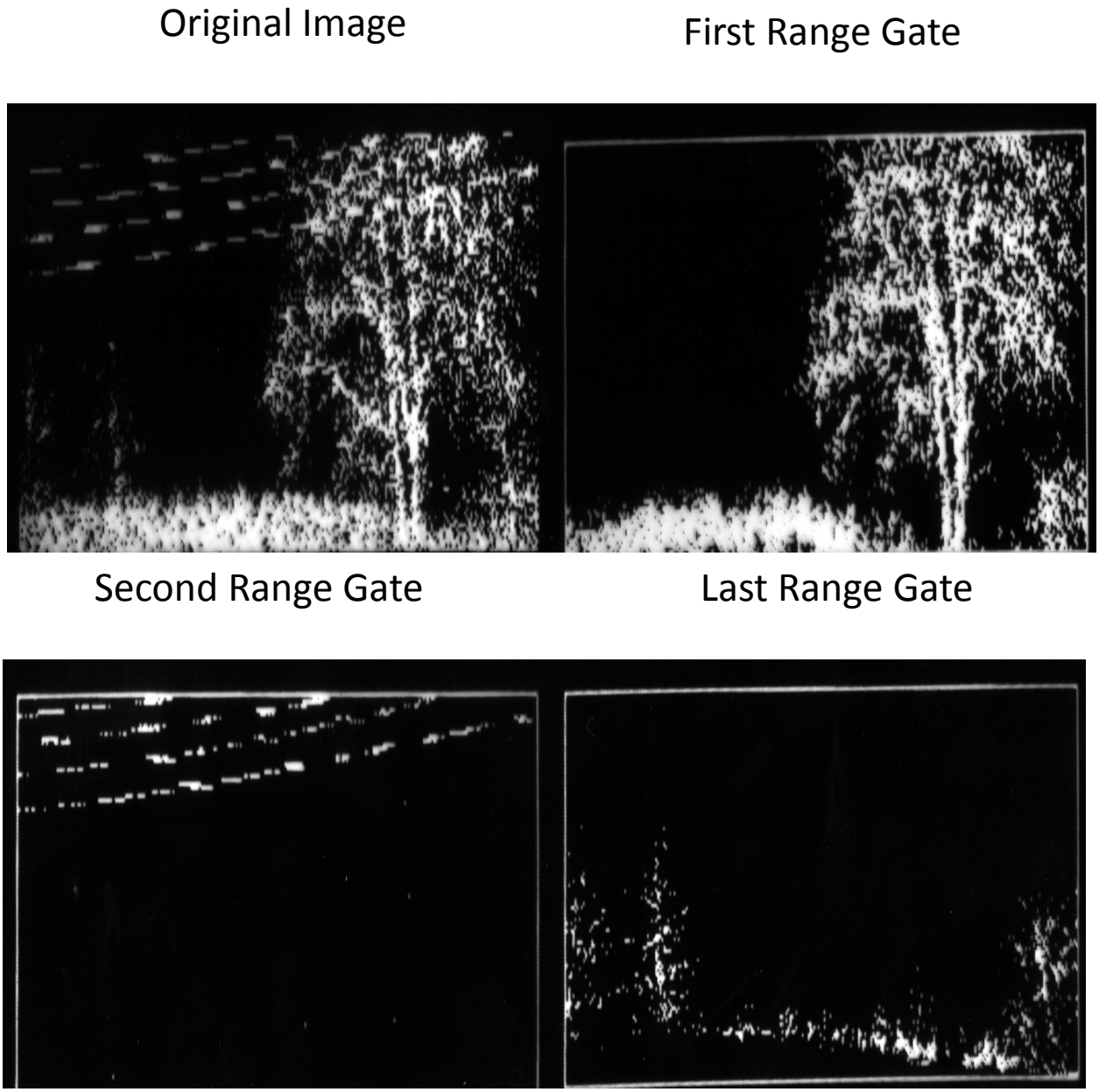

FIGURE 1-4 Early CO_2 ladar images. The original gray-scale encoded image is shown at the top left. The subsequent range images were produced by selecting a range window and displaying all returns within that window in white and all other ranges in black. The second and last range windows are more distant than the first. By gating out scattered light returning at different times, different features in the foreground and background of the scene can be isolated. SOURCE: P.F. McManamon, M. Huffaker, and G. Kamerman, 2010, "A history of laser radar in the United States," *Proceedings of SPIE*, 7684: 76840T.

Another very interesting feature of active EO sensing is the ability to gate out the light backscattered from obscurants. This results from the fact that it utilizes a high-bandwidth sensor with fine range resolution. A ladar can look at returns from a given range and ignore those that come back sooner or later. This is illustrated in Figure 1-4, where the power lines (bottom left) are clearly visible

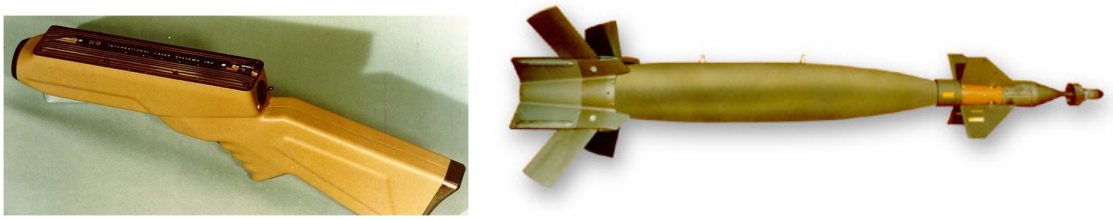

FIGURE 1-5 The first laser designator is shown on the left. The Pave Way laser-guided bomb, shown on the right, has revolutionized tactical air-to-ground warfare since its introduction in 1968. SOURCE: P.F. McManamon, M. Huffaker, and G. Kamerman, 2010, "A history of laser radar in the United States," *Proceedings of SPIE*, 7684: 76840T.

FIGURE 1-6 The Thanh Hoa (Dragon's Jaw) bridge in Vietnam off its abutment after attack. SOURCE: See http://en.wikipedia.org/wiki/File:ThanhHoaBridge1.jpg.

through and behind the tree foliage (top right). So long as there is sufficient return signal penetrating the obscurant to be detected, backscatter can be eliminated. This can be a significant advantage in military engagements in which smoke or fog may be encountered.

Active EO sensors have long played an important role in tactical military systems. While the technology has the potential for a wide range of applications, implementation has been limited by cost. Precision-guided munitions based upon semiactive EO technologies (Figure 1-5) have revolutionized modern warfare since laser designators first demonstrated their effectiveness by destroying the Thanh Hoa Bridge in Vietnam (Figure 1-6) over 40 years ago.[8] The introduction of precision-guided munitions, enabled by semi-active laser designators and weapon seekers, was a revolution in military affairs. Laser designators and seekers are a bistatic form of active EO sensing—the illuminator and sensor are located in different places.

"Wind sensing, navigation, long-range target tracking, terrain following, two-dimensional (2-D) and three dimensional (3-D) imaging, and velocity detection are other laser radar uses that have been pursued."[9] CO_2 laser-based ladar navigation systems have been deployed but have gone out of common use with the advent of global positioning system (GPS) navigation.[10] Modern ladar systems do, however, offer an alternative for navigation in GPS-denied environments.

[8] J.T Correll, 2010, "The emergence of smart bombs," *Air Force Magazine*, March: 60.

[9] P.F. McManamon, M. Huffaker, and G. Kamerman, 2010, "A history of laser radar in the United States," Proceedings of SPIE, 7684: 76840T.

[10] Ibid.

APPLICATIONS

Numerous other applications of active EO technologies are discussed in this report. The specific applications mentioned below give a flavor of some of the major opportunities. More discussion of these applications can be found in Chapters 2 and 3.

3-D Mapping

Ladar systems are rapidly emerging as the global standard for commercial 3-D mapping and are increasingly used to support mapping for U.S. military operations. Commercial 3-D mapping ladars are proliferating around the world and operate effectively in secure airspace.[11] According to the NEEA, lidar data had been collected over 28 percent of the conterminous United States and Hawaii as of 2011.[12]

The availability of ladar-generated 3-D maps has enabled immediate responses to humanitarian and emergency situations. After the earthquake in Haiti, the affected area was 3-D mapped by the Airborne Ladar Imaging Research Testbed (ALIRT) system to assure that victims were not migrating into flood-prone areas, to monitor the movement of displaced persons, and to evaluate the ability of candidate routes for disaster relief supplies to carry traffic.[13] After the attacks on the World Trade Center on September 11, 2001, 3-D ladars were used to map the rubble at ground zero to determine if it was shifting or sinking. Movement of the rubble could have posed a hazard to emergency response personnel.

In a recent Defense Advanced Research Projects Agency (DARPA) presentation[14] the director highlighted the 3-D mapping High Altitude Ladar Operational Experiment (HALOE) system. HALOE can map large areas in 3-D in a relatively short time. During Operation Enduring Freedom, over half of Afghanistan has been mapped with a fleet of airborne ladars.[15]

High-resolution mapping of the bottom of water bodies has become an important tool for littoral and riverine navigation. This optical bathymetry typically operates in the blue-green wavelength region for improved water penetration.

The committee expects that ladar will eventually replace most standard surveying techniques to provide native 3-D maps with significant economic benefits.

Wind Sensing

Global wind sensing can significantly improve weather prediction. Lidars are ideal for mapping global winds if one or more lidars can be placed in orbit. These can be either coherent Doppler lidars or direct-detection lidars that measure velocity via Doppler-induced spectral shifts. Wind sensing has been used around large wind turbine farms to assist in planning how to best harvest electricity from the winds; it can be used by the military for improved cargo or bomb drop accuracy, or for improved aim point accuracy on unguided weapons. Lidars can also be used on a sailboat to find the best method of using the winds or on an aircraft to detect clear air turbulence. At airports, a lidar can help detect microbursts and wind shear and can assist in planning aircraft separation. Global wind sensing from orbit could significantly improve weather prediction accuracy. Wind sensing is expected to continue to be an important application of future lidar systems.

---

[11] Michael S. Renslow editor, 2012, Manual of Topographic Lidar, American Society for Photogrammetry and Remote Sensing, 34.

[12] Available at http://nationalmap.gov/3DEP/neea.html.

[13] Available at http://www.ll.mit.edu/news/pathfindercover.html.

[14] Regina Dugan, former director of DARPA, Plenary Presentation, 2011, Defense Security, and Sensing Symposium, International Society for Optical Engineering (SPIE), Orlando, Fla., April 26, 2011, available at http://spie.org/x48217.xml.

[15] Personal communication from Walter Buell, committee member, August 26, 2013.

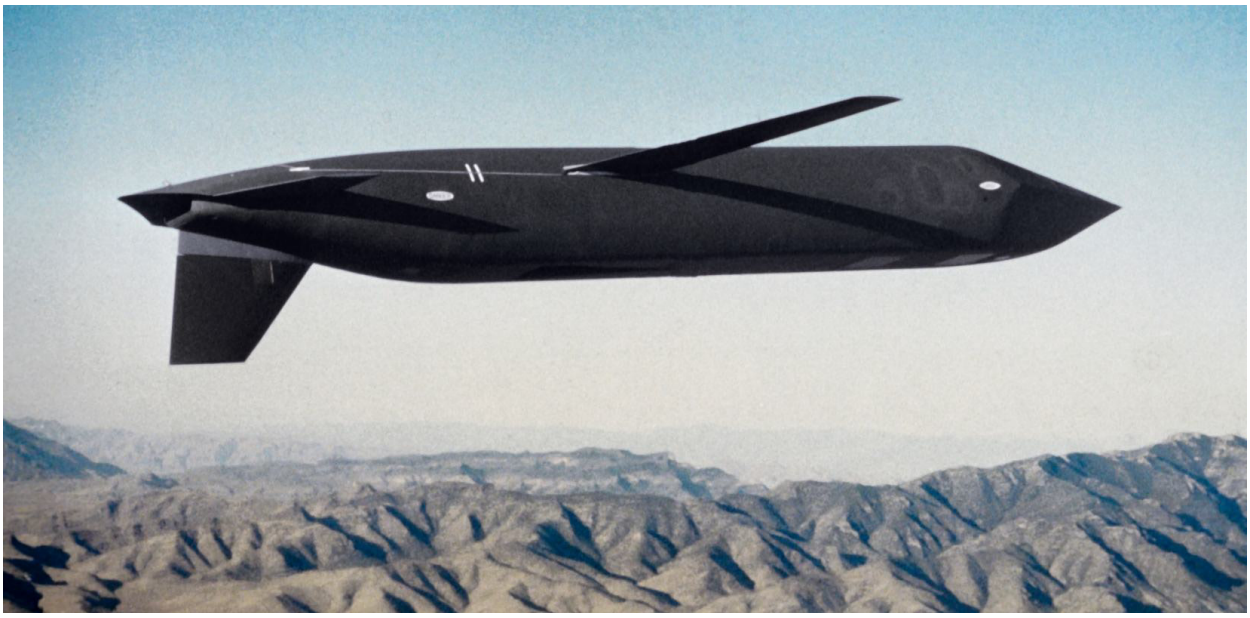

FIGURE 1-7 The AGM-129A is a subsonic, turbofan-powered, air-launched cruise missile. To reduce electronic emissions from the missile, the radar was replaced with a combination of inertial navigation and terrain contour matching (TERCOM) enhanced with highly accurate speed updates provided by a laser Doppler velocimeter.[16] SOURCE: See http://commons.wikimedia.org/wiki/File:Agm-129_acm.jpg.

Robotic Vehicle Control

The concept of robotic, or at least remotely controlled, combat vehicles is not new, but could be significantly enabled by active EO sensing. The German V-1 "buzz bomb" of the Second World War, the modern cruise missile, and the naval torpedo are operational examples of robotic vehicles. The AGM-129A, shown in Figure 1-7, even relies upon active EO for its guidance.

Previous robotic guidance systems required manual selection of the target, or a course to the target. The emergence of high-resolution 3-D imaging, near-real-time processing, and automatic target recognition algorithms could result in the deployment of autonomous, robotic combat vehicles. The high precision of ladar ranging makes it an attractive technology solution for fine-grained self-navigation of autonomous vehicles. Ladar has additional advantages for vehicular applications, such as coping with real-world obscurants such as fog, dust, and rain. As an example, all of the vehicles in the DARPA Grand Challenges (2004 and 2005)[17,18,19] and 2007 DARPA Urban Challenge[20] competitions were equipped with ladar sensors. The essential role of ladar technology for this application was local navigation. Stanley, Stanford University's winning entry in the 2005 race (now in the Smithsonian), is shown in Figure 1-8.

---

[16] P. McManamon, M. Huffaker, and G.W. Kamerman, 2010, "A brief history of laser radar," Military Sensing Symposia Specialty Group on Active-EO Systems, Orlando, Fla., September 28-30.

[17] Umit Ozguner, Keith A. Redmill and Alberto Broggi, 2004, "Team Terramax and the DARPA (1) Grand Challenge: A General Overview," *IEEE Intelligent Vehicles Symposium:* 232.

[18] R. Behringer, S. Sundareswaran, B. Gregory, R. Eisley, B. Addison, W. Guthmiller, R. Daily and D. Bevly, 2004, "The DARPA grand challenge: development of an autonomous vehicle," *IEEE Intelligent Vehicles Symposium*: 226.

[19] Martin Buehler, Karl Iagnemma, and Sanjiv Singh, eds., 2007, *The 2005 DARPA Grand Challenge: The Great Robot Race*, Springer.

[20] Martin Buehler, Karl Iagnemma, and Sanjiv Singh, eds., 2009, *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic*, Springer, 2009.

FIGURE 1-8 Stanford University's 2005 DARPA Grand Challenge winner, Stanley. SOURCE: Linda A. Cicero, Stanford University News Service, copyright 2005 The Board of Trustees of the Leland Stanford Junior University.

Autonomous vehicles are typically equipped with a suite of sensors including radars, cameras, and GPS. GPS is ideal for course-following—for a set of waypoints for a multiple-mile course through desert terrain—but it cannot cope with the obstacles and hazards encountered on a more local scale, such as boulders and ditches or, in the case of the Urban Challenge, other movers.

These autonomous vehicle competitions can be viewed as technology trials for self-driving automobiles, which are now being evaluated experimentally with the intent to commercialize.[21] This is potentially a substantial market for ladar technology. Experience with solutions for self-driving automobiles is likely to provide insights into related domains requiring motion coupled to considerable autonomy. Examples of domains include extraterrestrial rovers as well as possible future military applications such as battlefield robotics that may exploit other ladar technology features such as object identification for targeting or identify-friend-or-foe (IFF). The market consideration is important, since it will be a spur for reducing cost and advancing performance as well as maintaining an industrial base for ladar technology. Battlefield robotics is a particularly compelling national security consideration, as it is a priority for U.S. armed services[22] and is clearly a battle space game changer.

While primarily focused on the logistics task, Boston Dynamics' ladar-equipped "BigDog"[23] (Figure 1-9) is one precursor of battlefield robots that can traverse rough terrain, with considerable implications for national security, since BigDog enables fighting with fewer soldiers carrying less equipment with them. The possibilities for a weaponized version of BigDog can also alter battlefield tactics for both offense and defense.

[21] Available at http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works.

[22] National Research Council. 2002, *Technology Development for Army Unmanned Ground Vehicles*, Washington, D.C.: The National Academies Press.

[23] Marc Raibert, Kevin Blankespoor, Gabriel Nelson, Rob Playter, and the BigDog Team, 2008, "BigDog, the rough-terrain quadraped robot," *Proc. International Federation of Automatic Control*.

FIGURE 1-9 Boston Dynamics' BigDog quadraped robot. SOURCE: BigDog Robot image courtesy of Boston Dynamics.

Object/Material Identification

Object identification is a primary use for active EO sensing because of the power of the various ladar modes for object discrimination, discussed below. For example, a range-gated 2-D active EO image is far superior to a passive EO image at night. With active EO, quality illumination is assured, and one can use 1-μm or 1.5-μm illumination rather than the mid-wavelength infrared (MWIR, 2.7-6.2 μm) or long-wavelength infrared (LWIR, 6.2-15 μm) cameras required at night for passive sensors. This means that with active EO, the diffraction-limited spatial resolution can be at least 2-3 times better than with a passive camera,[24] and there is no crossover issue[25] where available signal becomes limited. 3-D imagery provides even more information, including the possibility of very high-resolution range information in each pixel.

Laser vibrometry (see below) can provide high-quality object identification for most objects with moving parts, such as engines. Synthetic aperture ladar can provide 2-D or possibly 3-D images at long ranges. Use of multiple illuminating laser polarizations and wavelengths simultaneously can provide additional discriminants. Speckle characteristics caused by interference in the light scattered by the target can also be used to help identify an object, as can surface material identification, using surface fluorescence or other methods.

[24] Passive short-wave infrared (SWIR) imaging is possible at night using airglow (Meinel bands in the 1.3-1.9 μm range). This would offer spatial resolution comparable to that of the active EO system discussed here.

[25] At crossover the contrast of the object compared to the background goes to zero, making it difficult to see. Crossover occurs twice a day for passive IR sensing, once as an object is heating up, and once as it is cooling off.

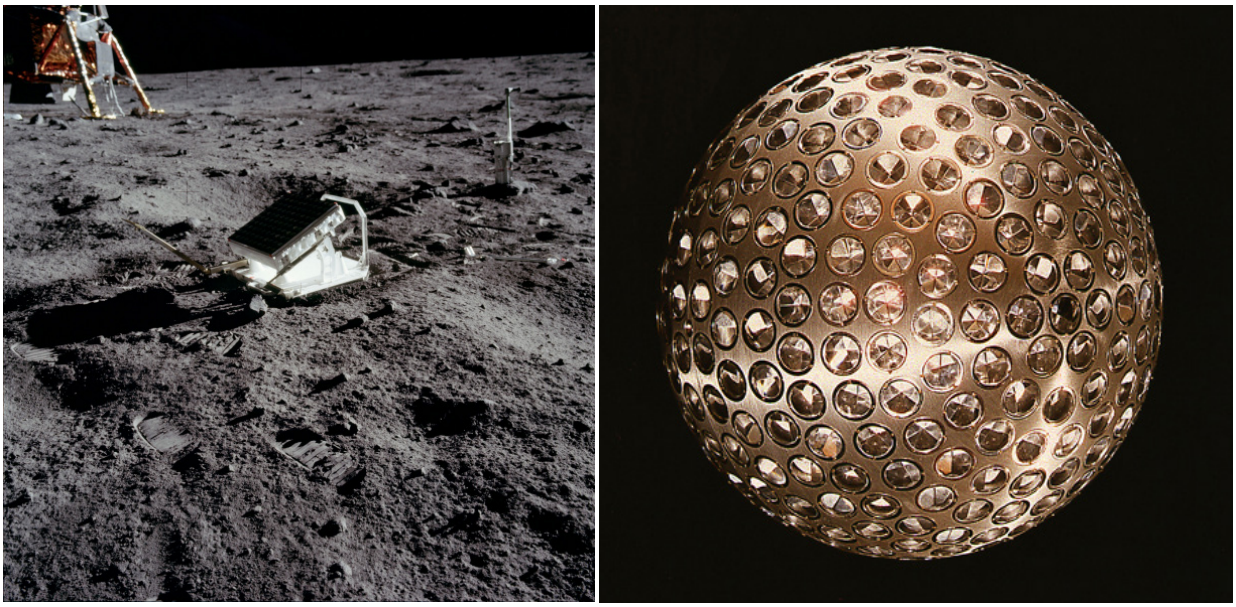

FIGURE 1-10 Retroreflectors on the Moon (*left*) and a satellite (*right*) aid in precise laser ranging. SOURCE: See http://library01.gsfc.nasa.gov/gdprojs/images/LAGEOS.jpg (*left*) and see http://www.hq.nasa.gov/office/pao/History/alsj/a11/AS11-40-5952.jpg (*right*).

Vibration Measurement/Object Characterization

The ability to measure very fine vibrations using active EO sensors enables status determination in many situations and is a method of characterizing objects.[26] Vibration measurements can potentially determine whether power is flowing through a transformer or whether liquid is flowing through a pipe, based on the acoustic or vibrational signature produced by such activities. It may be possible to determine not only how many cylinders an engine has but even if it needs a tune-up. Laser vibrometers have been used to determine based on harmonic variations, whether a bridge is degrading. Companies concerned about noise from machinery use laser vibrometers to find noise sources in the machinery, and they can be used to make quieter cars and airplanes. In the future, oil companies may be able to explore for oil using laser vibrometers. Anytime there is a need to characterize object motion, a laser vibrometer will be a useful sensor.

Satellite Laser Ranging

Precise long-distance ranging can be facilitated by attaching retroreflectors on the target. Retroreflector arrays have been deployed on the moon for precise ranging. In Figure 1-10 the array left by Apollo 11 is shown on the left. On the right of Figure 1-10 is shown the LAGEOS-1 satellite, which was launched in 1976 and is covered with 426 corner-cube retroreflectors to enhance laser ranging.

[26] An example of one commercial system can be found at: http://www.polytec.com/us/.

FIGURE 1-11 The Airborne Laser Mine Detection System (ALMDS). SOURCE: Photo credit: Northrop Grumman Corporation *(left)* and http://www.navy.mil/management/photodb/photos/100608-N-0001S-002.jpg *(right)*.

Underwater Mine Detection

Water is very opaque to microwave radiation because of its conductivity. While blue-green lasers may not penetrate to large depths, they do penetrate to depths of tens of meters, depending on water turbidity. Just as blue-green lasers can be used for short-range underwater communication, they can be used to find objects in the water. 3-D mapping of coastal underwater areas can locate mines, a very valuable service for the military. Figure 1-11 shows a mine detection system.

Weapon Seekers

A 3-D ladar can provide range as well as angle/angle information. This will make closing on a target easier. Also, coherent ladar can provide target velocity, and direct-detection ladars can use multiple range measurements to obtain velocity. Systems such as the Air Force Research Laboratory's (AFRL's) Low Cost Autonomous Attack System (LOCAAS) have used ladar to make a 3-D image of objects and automatically recognize them. This could provide an autonomous seeker capability.

Under the usual offensive rules of engagement, the United States does not permit a kill decision to be made without the involvement of a responsible human decision maker, as would be possible with robotic attackers. It is not obvious, however, that a potential adversary will employ the same operational constraints that the United States employs. The same observation could be made about U.S. adherence to eye-safety standards (see Box 1-4). The uses of active EO sensing could be a technological surprise if a state or nonstate actor were to impose fewer ethical or policy constraints on an EO-sensor-equipped weapon system.

Police Speed Measurement

Police forces around the world now use ladar to identify vehicles that are speeding. These systems are relatively inexpensive, on the order of thousands of dollars. For a police force, the ladar systems pay for themselves very rapidly. Most, if not all, current police speed measurement ladars are direct-detection ladars, using multiple range measurements to estimate speed.

Entertainment

The use of active EO sensing is only limited by the imaginations of people using it; for example, the Microsoft Kinect game system shows an innovative use of active EO sensing. The committee was also briefed on research aimed at developing a handheld video camera that would generate "real" 3-D movies.[27] Such future commercial entertainment applications could provide profitable opportunities for developing the technology.

REPORT SCOPE AND COMMITTEE APPROACH

By comparison with passive EO sensing systems reviewed in an earlier NRC report,[28] the added complexity of laser illumination in active EO systems means that the scope of this report is considerably broader, as reflected in the committee's statement of task (see the Preface). In order to focus the report as much as possible on the sponsors' priorities, the committee chose to bound the discussion in several ways.

First, it distinguished between what can be called the "engineering limits" of key technologies (those that may be exceeded by the invention of new materials or methods) and the fundamental limits (those imposed by the laws of physics) and tended to focus on the latter.

Second, the committee focused on active EO technologies that might be developed by potential adversaries in the next 10-15 years to generate technological surprise. The adversaries considered included peer countries with highly developed technological bases as well as nonstate actors whose capabilities might derive only from access to commercial off-the-shelf systems.

Third, the committee attempted to identify areas of foreign strength in active EO technologies (as well as areas in which the United States is falling behind) based on the open literature and its own expertise, as required in the statement of task (see the Preface). It did not attempt to describe and characterize active EO R&D programs in terms of either content or level of funding. Such an undertaking would not only have been time- and resource-intensive, it would likely have been futile because many such government programs are classified or not discussed in the open literature. Thus, the committee chose to focus on technologies and key application areas rather than programs.

Fourth, a comprehensive examination of active EO capabilities would include those for tasking EO assets, collecting data, processing data into useful information, exploiting the information, and disseminating the products in a timely way to the warfighter or other ultimate user in a user-friendly form. The acronym TCPED is sometimes used to describe this comprehensive chain of events. Communication methods, including considerations of speed, data fidelity, and data security, are critical at each stage of TCPED. The committee recognizes that each of these aspects of active EO sensing is important, but the statement of task places emphasis on the front-end technologies of TCPED such as lasers, optics, and detectors, and this is where the committee put its emphasis, although it did consider some aspects of processing sensor data into useful information.

Finally, in discussion with the study sponsors, it was decided that this report would focus on active EO sensing systems and not on technologies for damaging them or other kinds of countermeasures that might be employed to reduce their effectiveness. However, active EO technologies that could be used for battle damage assessment or to assess the remaining capabilities of EO assets that had been attacked did fall into the sensing category and therefore were included in the committee's scope.

[27] Paul Banks, Tetravue, presentation to the committee on March 6, 2013.

[28] National Research Council, 2010, *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*, Washington D.C.: National Academies Press.

STRUCTURE OF THIS REPORT

This chapter has introduced active EO technologies and their applications. The remaining chapters roughly follow the tasks identified in the statement of task (see the Preface). Chapter 2 provides an overview of existing active EO approaches, including applications, methods, and what the data products look like. In Chapter 3, emerging active EO systems are explored, particularly those that might be realized in the next 10-15 years. The individual components of active EO systems, including lasers, detectors, associated optics, and signal processing, are discussed in Chapter 4. Finally, Chapter 5 explores the fundamental limits of these technologies and their implications for the future.

Appendix A provides the biographies of members of the committee. Appendix B lists the committee's public meeting sessions and the presentations made in them. Appendix C discusses various types of laser sources and their fundamental and engineering limits in a series of tables. Subsequent appendixes are classified or otherwise restricted and are available in the full classified version of this report.

CONCLUDING THOUGHTS

Active EO sensing provides a very rich phenomenology to explore. The laser, detector, and aperture technologies to support active EO sensing are maturing and becoming cheaper. The telecommunications revolution over the last few decades has dramatically improved the quality and lowered the cost of many components used in active EO sensing. Whether it is referred to as ladar, lidar, or laser radar, the application space is large and growing, and is being pursued diligently by the United States, its allies, and its potential adversaries around the world. Understanding the historical and current development trajectories of passive EO and radar systems will serve as a useful guide for projecting the future course of active EO sensing technologies.

The committee performed a Web search of the annual number of publications in different regions of the world over the past 10 years that use the terms "lidar" or "ladar" in their titles or as their "topic;" the results are shown in Figure 1-12. The data show that there has been a steady increase in the number of such publications in each region. However, the annual number of publications in China has been increasing rapidly over the past several years. Active EO sensing is a game-changing sensing technology.

Conclusion 1-1: Active EO sensing enables measurements not possible with passive EO sensors by exploiting the control of the illumination. Exploiting control of the illumination includes its coherence properties.

Conclusion 1-2: Active EO sensing offers many of the controlled illumination advantages of radar, and by operating at much shorter wavelengths enables greater range precision and angular resolution as well as more literal interpretability.

Recommendation 1-1: To avoid the possibility of technological surprise, the U.S. intelligence community should consider all potential uses of active electro-optical sensing, including those not allowed by the ethics and policies of the United States.

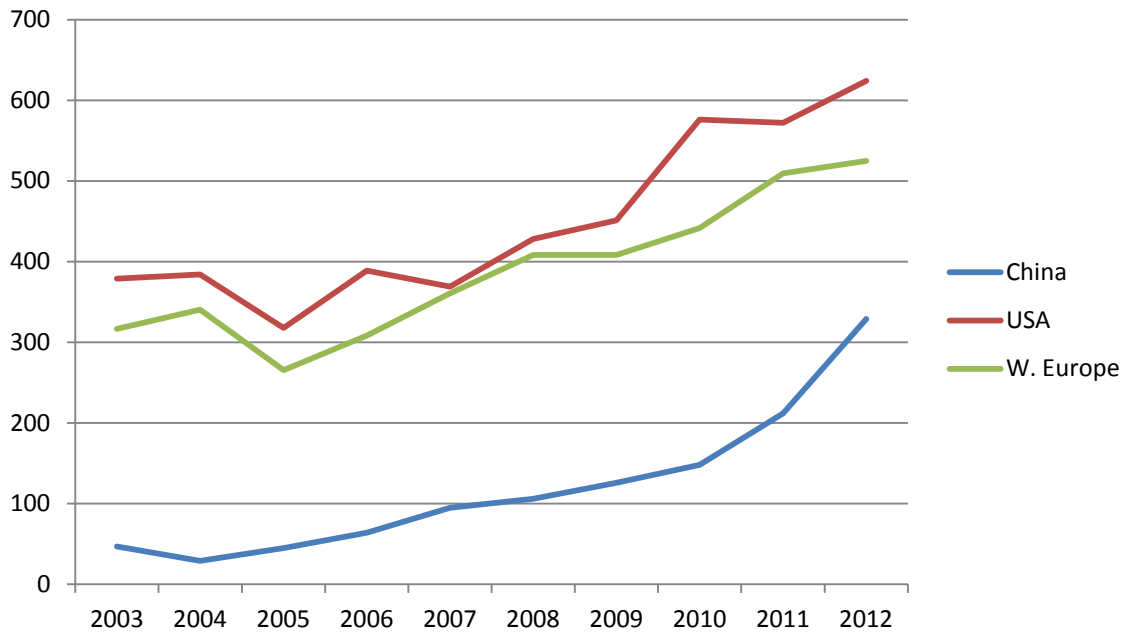

FIGURE 1-12 The number of publications on lidar or ladar per year is growing in the labeled countries/regions. It appears that the growth in the United States is comparable to that in Western Europe, while the recent rapid increase in the research in China is remarkable. The number of publications in Western Europe was determined by summing those from Germany, France, Great Britain, and Italy, and then increasing this number by 19 percent, which is the fractional increase in the number of publications due to the other Western European countries during this 10-year period.

2

Active Electro-Optical Sensing Approaches

In this chapter the committee discusses mature and fielded active electro-optical (EO) sensing technologies. Emerging or future EO technologies are discussed in Chapter 3.

RANGE MEASUREMENT TECHNIQUES

One of the most widely exploited features of active EO sensing is the ability to measure range very accurately. While some lidars and ladars (such as 2-D imagers) do not make use of accurate range measurement capability, most of them do. Range can be measured using active EO sensors in many fashions. The earliest, and in many ways simplest, method is to emit a short laser pulse and measure the time when it returns. This method is simplest, but it has limitations. This is the method of range measurement used with direct detection ladars.

With a single pulse, there is a trade-off between the precision of a range measurement and the precision of a velocity measurement, since a high-resolution range measurement requires a short pulse, and a high-resolution Doppler velocity measurement requires precise measurement of frequency, which takes time. Range measurement precision for a simple pulsed ladar is based on a convolution of the detector response and the return pulse shape. To have both precise measurement of range and velocity at the same time, a series of laser pulses can be used. The range precision comes from the rise time and width of a single pulse. The velocity precision comes from the pulse trains, assuming phase coherence across the pulses in the measurement.

High range precision does not necessarily require a short pulse. Any waveform that has wide bandwidth can provide high range resolution. In addition to short pulses and an array of short pulses, range can be measured with great precision by using a linear frequency modulated (LFM) chirp or a pseudo random coded waveform with sufficient bandwidth. The range resolution will be approximately the inverse of the bandwidth, multiplied by the speed of light, and then divided by a factor of 2 because the light has to go both ways. A good rule of thumb is that 1 nsec provides one-half foot—or 0.15 m—of range resolution, due to the two-way path.

To measure frequency modulation requires heterodyne detection. One major benefit of LFM as a method of range measure is it enables chirping of the local oscillator (LO) along with the emitted signal. Chirping the LO time delay compared to the linear chirp of the master oscillator is sometimes called stretch processing. The time delay of the linear chirp limits the difference in frequency between the LO and the return signal without limiting the ladar bandwidth associated with range resolution. For example, Bridger Photonics recently published results[1] with 3.6 THz LFM, resulting in 50 μm range resolution, but there are no detectors capable of bandwidths even close to that level. Of course, if one only needs range resolution consistent with available detector bandwidth then one does not need this advantage. Even if one has available detectors with sufficient bandwidth, use of stretch processing may save money by allowing use of cheaper, lower bandwidth detectors.

[1] R.R. Reibel et al., 2009, "Ultrabroadband optical chirp linearization for precision length metrology applications," *Optics Letters*, 34: 3692.

Digital holography, or spatial heterodyne (see Chapter 3), can measure range by using more than one wavelength. As the two or more wavelengths walk off from each other in phase, range can be determined. Usually the two wavelengths are separated by large bandwidths, resulting in very fine range resolution, but also very fine unambiguous range.

LASER RANGE FINDERS

Laser range finders are very simple forms of laser radar. A range finder determines the range to a target object based on the round-trip time-of-flight of a laser pulse to and from the object. Most laser range finders are direct-detection ladars, meaning the return pulse hits the detector and triggers a range measurement. A combination of the return pulse shape and the detector response determines how accurately one can measure the round trip time.

An early application of a laser range finder was the determination of the precise distance from the Earth to the Moon.[2] "The first laser ranging retroreflector was positioned on the Moon in 1969 by the Apollo 11 astronauts (see Figure 1-10)."[3] By beaming a 60-Joule ruby laser pulse at the reflector from Earth, scientists were able to determine the round-trip travel time, which gave the distance between the two bodies at any time to a precision of about 3 cm.[4]

While the U.S. military pioneered laser range finders,[5] they are now available for construction, surveying, golf, hunting, or many other applications where one wants to know the range to an object. Laser range finders, also called laser rangers, are widely available commercially and cost from $100 up to tens of thousands of dollars. Usually, commercial range finders are direct detection pulsed ladars.[6]

The main limitation on range finders is the atmosphere. If the laser beam cannot be transmitted to the target, or received from the target, then one cannot use a laser ranger. For long-range measurements, slight variations in the index of refraction of the atmosphere can slow the speed of light, introducing slight inaccuracies in the range measurement.

ONE-DIMENSIONAL RANGE PROFILE IMAGING LADAR

A pulsed laser and a single detector can provide a range profile from an object with a range resolution that depends on the transmitter pulse rise time, pulse width, or in the case of coherent ladars, the width of the frequency chirp (e.g., how short the laser pulses are). Of course one must match the transmitter pulse with how fast the detector can resolve the pulse, and a wide frequency chirp must be matched by a detector with sufficient bandwidth. The range profile does not depend on the size of the receiving optical apertures, except for signal-to-noise considerations. This can be very useful in environments that are not very cluttered and for objects that are far away, where the transverse dimensions of the object are not resolved by the optical system and the only spatial information about the object comes from the range profile.

A high-range-resolution profile is most useful when the orientation of the object is known; for example, one generally assumes that an airplane is oriented so the nose is pointed in the direction that the plane is traveling. Figure 2-1 shows how an airplane's scattering signal will vary with range, depending on what scatterers are in a given range bin. Starting from the time when the detector is first triggered by

[2] P.F. McManamon, M. Huffaker, and G. Kamerman, 2010, "A history of laser radar in the United States," *Proceedings of SPIE*, 7684: 76840T.

[3] Lunar Planetary Institute, 1994, "Apollo Laser Ranging Experiments Yield Results," LPI Bulletin 72. http://eclipse.gsfc.nasa.gov/SEhelp/ApolloLaser.html.

[4] Ibid.

[5] J.T. Correll, 2010, "The emergence of smart bombs," *Air Force Magazine*, March: 60.

[6] See, for example, patent publication number US5359404 A.

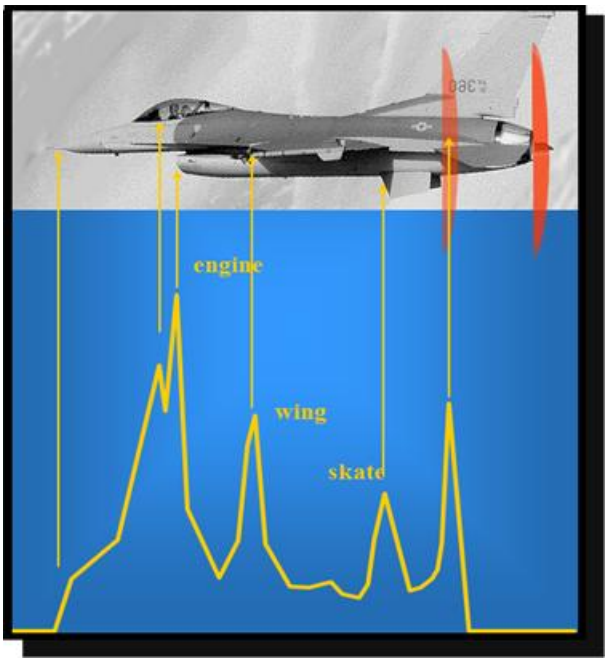

FIGURE 2-1 Range profile of an airplane viewed in one dimension (1-D) from the front, although the image shown is from the side. SOURCE: Courtesy of Matthew Dierking, Air Force Research Laboratory, Dayton, Ohio.

the large scattering signal from the nose, the dimensions of the rest of the airplane can be calculated by the relative time delay.

Tomographic Imaging Using Multiple 1-D Profiles from Various Angles

Tomographic methods, which reconstruct an image from a set of its projections, are well known in radio astronomy and medical imaging. The goal of reflective tomography is to estimate an object's surface features using a set of reflective projections (as opposed to transmissive projections). These projections are measured in angular increments around the object to obtain 3-D information.[7] Each angular resolution cell in the projection represents the energy reflected off the corresponding illuminated surface of the object at that angle. For any laser radar, the received signal represents information about the surface of the object illuminated by the radar from a given line of sight. A series of signals along the range resolution coordinate produces a reflective projection of the object that can be used to reconstruct the object.[8] Figure 2-1 shows one such 1-D image.

[7] F.K. Knight, S.R. Kulkarni, R.M. Marino, and J.K. Parker, 1989," Tomographic techniques applied to laser radar reflective measurements," *Lincoln Laboratory Journal* 2 (2): 143.

[8] X. Jin and R. Levine, 2009, "Bidirectional reflectance distribution function effects in ladar-based reflection tomography," *Appl. Opt.* 48: 4191-4200.

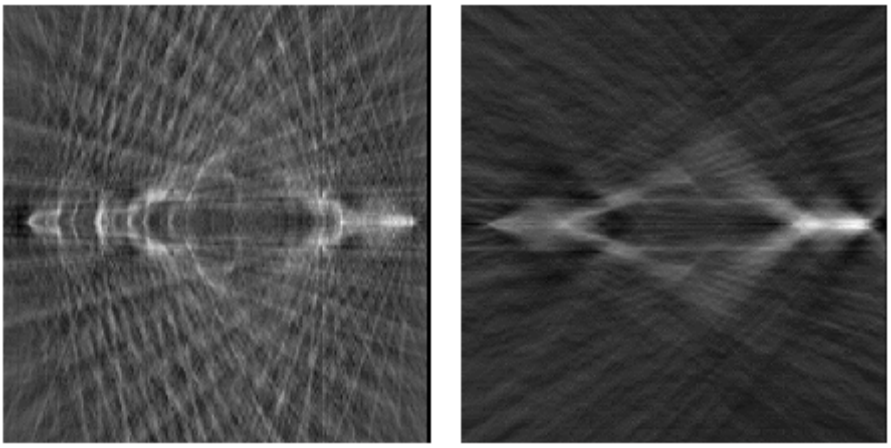

FIGURE 2-2 Two-dimensional projection reconstruction of an F-4 fighter for (a) measurements from 0 to 180 degrees and (b) measurements within 20 degrees around broadside. SOURCE: M.P. Dierking, F. Heitkamp, and L. Barnes, 1998, "High temporal resolution laser radar tomography for long range target identification," *OSA Signal Synthesis & Reconstruction Conference,* June.

In this technique, a pulse whose physical length is short compared to the range extent of the object is first reflected off an object. The resulting time-dependent return signal is collected by single detector in a 1-D optical system (single-element detector), which provides signal as a function of range.[9] This 1-D signal relates to a 1-D slice of the spatial 3-D profile of the object.[10] When the object rotates, different slices of the 3-D profile can be obtained.[11] If a sufficient number of 1-D slices are collected, a 2-D image can be reconstructed. A scanning ladar can be employed as long as the 2-D images are registered in angle.[12]

Figure 2-2 shows a 2-D image, obtained by Air Force researchers in 1998, using range profiles determined at multiple angles. Various algorithm approaches can be used to combine multiple range profiles to make a more complete image from multiple 1-D profile images at different angles to the target. Wider-angle reconstruction yielded better detail, as expected.

Range profiles from various angles can be obtained when the object moves with respect to the 1-D ladar. The object may be stationary, as a target on the ground, or moving, as an airplane, or rotating, as a satellite. In each case suitable mathematics has been developed to provide images. The more complex the object, the more interest there is in obtaining range profile information over multiple angles, but even with a single range profile 1-D imaging can be useful for recognizing some objects, as discussed in the previous section.

[9] J.B. Lasche, C.L. Matson, S.D. Ford, W.L. Thweatt, K.B. Rowland, V.N. Benham, 2009, "Reflective tomography for imaging satellites: Experimental results," *Proc. SPIE* 3815: 178, Digital Image Recovery and Synthesis IV.

[10] Ibid.

[11] J. B. Lasche, C. L. Matson, S. D. Ford, W. L. Thweatt, K. B. Rowland ; V. N. Benham, 2009, "Reflective tomography for imaging satellites: experimental results," *Proc. SPIE* 3815: 178, Digital Image Recovery and Synthesis IV.

[12] R.M. Marino, T. Stephens, R.E. Hatch, J.L. McLaughlin, J.G. Mooney, M.E. O'Brian, G.S. Rowe, J.S. Adams, L. Skelly, R.C. Knowlton, S.E. Forman, and W.R. Davis, 2003, "A compact 3-D imaging laser radar system using Geiger-mode APD array," *Proceedings SPIE*, 5086: 1.

When a coherent detection system is employed and data are collected from multiple views, an object's Doppler spectrum can be used to help determine the angle from which a given 1-D image has been collected.[13] With Doppler techniques, the rotation speed of an object can be precisely measured.

The ability of range-resolved reflective tomography to image satellites was demonstrated in 2001.[14] The researchers used a CO_2 laser mode-locked to produce nanosecond pulses, and the data were collected using a coherent laser radar system. "Key components of the image reconstruction process included the calculation of tomographic angles and the alignment of the intensity projections to an appropriate center of rotation."[15]

Researchers from the Electronic Engineering Institute of the State Key Laboratories of Pulsed Power Laser Technology and Electronic Restriction in Anhui Province, Hefei, China recently reported their verification by computer simulation of ladar reflective tomography imaging of space objects.[16] This paper concluded that atmospheric turbulence has little effect on the coherent detection signal at 10.6 μm, because the atmospheric coherence length is on the order of meters. They found that chirped pulse signals are advantageous for reflective tomography imaging of space targets and that the "optimal pulse repetition is decided by the target altitude, target size, and the cutoff frequency in the spectrum domain."[17] Background work on reflective tomography has been ongoing at Key Laboratory of Space Laser Communication and Testing Technology at the Shanghai Institute of Optics and Fine Mechanics in China.[18]

Most early work on 1-D ladar, and tomographic imaging with 1-D ladar used low-duty-cycle pulsed ladar sensors. More advanced range measurements techniques can be used, such as linear frequency-modulated (FM) chirp or pseudo-random coded range measurements. This provided another method of achieving both 1-D imaging and tomographic imaging.[19]

TWO-DIMENSIONAL ACTIVE/GATED IMAGING

Two dimensional (2-D) gated imaging is one of the more straightforward and mature methods of active EO imaging. It is a useful imaging technique for seeing through haze and smoke, for seeing objects in shadow, for imaging at higher angular resolution in darkness and at longer range (approximately 10-20 km) than can be achieved with passive MWIR systems, and for seeing objects more clearly against a cluttered background. "Range-gated active imaging is widely used in night vision, underwater imaging, three-dimensional scene imaging, and other applications because of its ability to suppress backscatter from fog and other obscurants, as well as its high signal-to-noise ratio, high angular resolution, long detection range, and direct visualization."[20]

---

[13] C. Matson and D. Mosley, 2001, "Reflective tomography reconstruction of satellite features? Field Results," *Appl. Opt.* 40(14): 2290-2296.

[14] C. Matson and D. Mosley, 2001, "Reflective tomography reconstruction of satellite features? Field Results," *Appl. Opt.* 40(14): 2290-2296.

[15] C. Matson and D. Mosley, 2001, "Reflective tomography reconstruction of satellite features? Field Results," *Appl. Opt.* 40(14): 2290-2296.

[16] F. Qu, Y. H., and D. Wang, 2011, "Lidar reflective tomography imaging for space object," *Proc. of SPIE* 8200: 820015.

[17] Ibid

[18] X. Jin, J. Sun, Y. Yan, Y. Zhou, and L. Liu, 2012, "Imaging resolution analysis in limited-view laser radar reflective tomography, *Optics Communications*, 285 (10-11): 2575.

[19] J. Murry, J. Triscari, G. Fetzer, R. Epstein, J. Plath, W. Ryder, and N.V. Lieu, 2010, "Tomographic Lidar," OSA Conference Paper, Applications of Lasers for Sensing and Free Space Communications, San Diego, CA, February.

[20] X. Wang, Y. Zhou, S. Fan, Y. Liu, H. Liu, 2009, "Echo broadening effect in the range-gated active imaging technique." *Proc. SPIE* 7382:738211, International Symposium on Photoelectronic Detection and Imaging 2009: Laser Sensing and Imaging.

2-D gated imagers are typically used in conjunction with laser illuminators and often share the same pulsed laser. They use the pulsed laser to illuminate a scene, then collect the returning reflected light at a predetermined delay time based on the range of the target. These sensors are gated both to eliminate the reflection from obscurations and to reduce the noise coming into the detector (Figure 2-3). The range information is typically first acquired with the co-located laser illuminator. The gated imager uses this information to set the start of the integration window of a high-speed focal plane receiver to coincide with light returning from the range of interest in the scene. This provides a range slice image each time the scene is illuminated. The scene can be interrogated at a series of ranges by varying the delay in turning on the receiver. An example is shown in Figure 2-4.

Collecting imagery using a 2-D gated camera has three primary advantages over passive imaging:

1. For scenes with clutter (such as particles in the air or haze) or obstructions near the camera or clutter in the background behind the area of interest, 2-D gated imaging effectively eliminates the visually confusing effects of clutter by only imaging the slice of the scene of interest. In these cases the object of interest is either not visible or very indistinct when viewed with standard visible or IR cameras.
2. 2-D gated imaging is sometimes referred to as "bring your own flashlight." It effectively brings out features in shadowed or recessed areas, such as inside doorways or other openings or under overhanging structures. By shining the laser into those areas, it is easy to see what is not otherwise visible. It also works when thermal imaging has limited signal, such as thermal crossover.
3. An active EO eye-safe laser 2-D gated camera operating at 1.55 µm will have approximately 3 times better angular resolution than an MWIR sensor. Higher F-number optics allow the user to make use of the improved diffraction limit resulting from operating at shorter wavelengths to see finer details, perform better target detection and recognition, and to see details that have contrast at the laser wavelength but not in the thermal region.

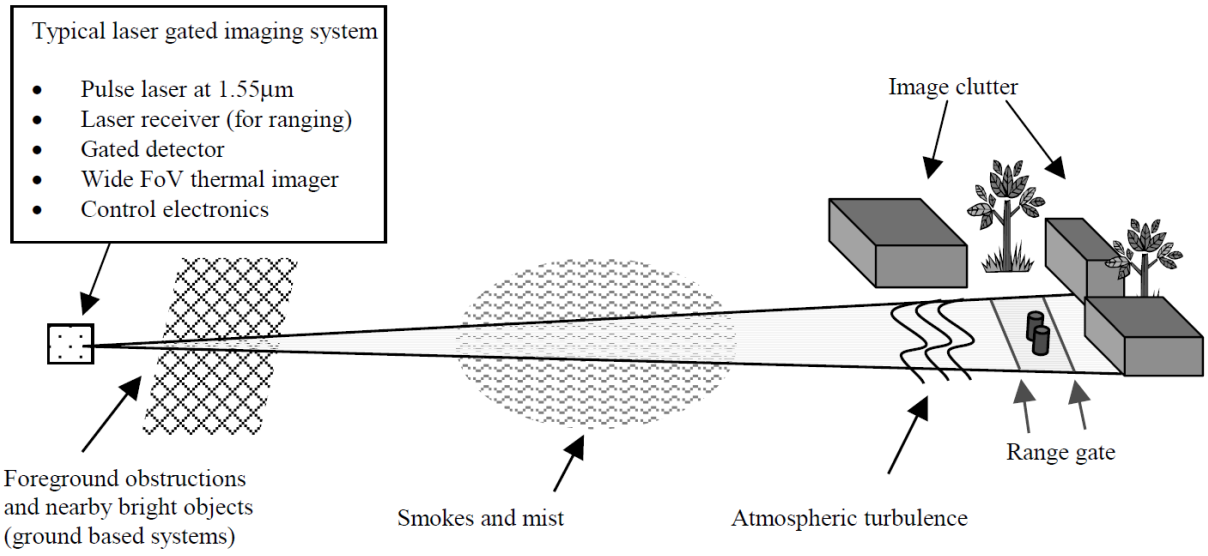

FIGURE 2-3 A gated active EO imaging system can eliminate clutter and obscurants that might otherwise degrade the image. Although laser returns from image clutter, smoke, mist and foreground obstructions can be range-gated, effects of atmospheric turbulence remain. SOURCE: I. Baker, S. Duncan, and J. Copley, 2004, "A Low Noise, Laser-Gated Imaging System for Long Range Target Identification," *Proc. of SPIE*, Vol. 5406. Courtesy of Selex ES Ltd.

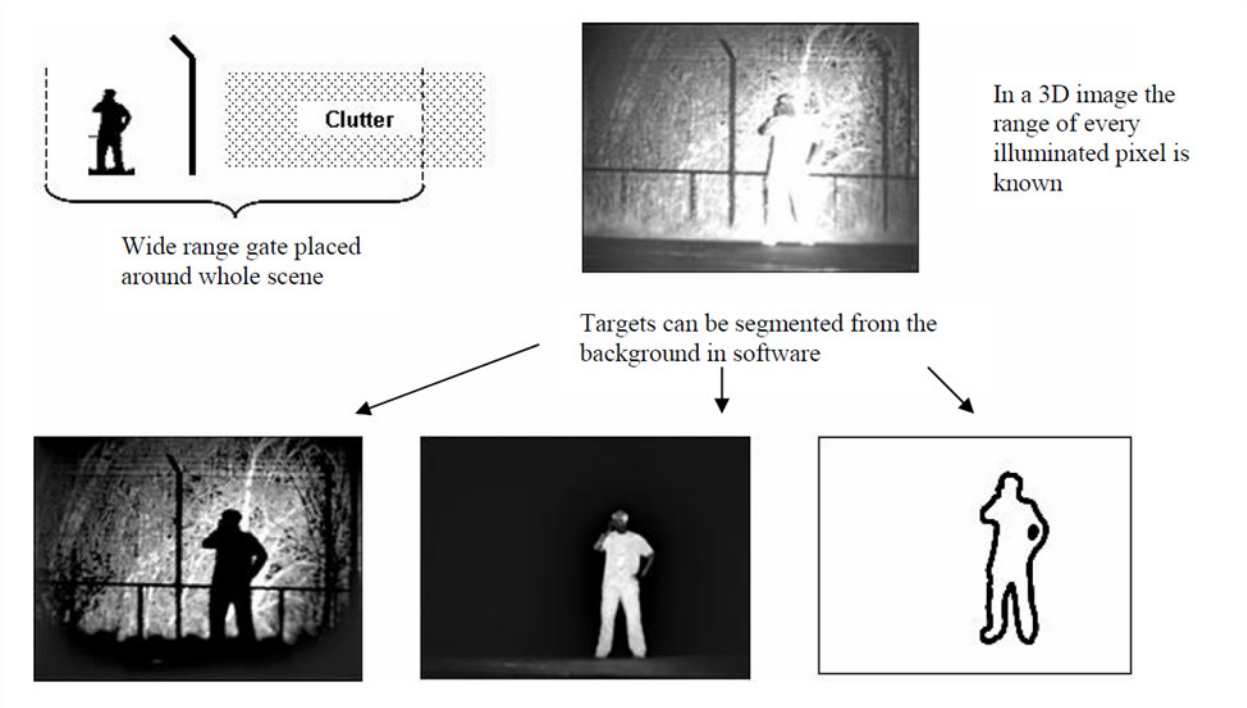

FIGURE 2-4 Illustration of how the outline of a man can be extracted from the deep gate by knowledge of the individual pixel ranges. SOURCE: I. Baker et al., 2008, "Advanced infrared detectors for multimode active and passive imaging applications," *Infrared Technology and Applications XXXIV*, edited by Bjørn F. Andresen, Gabor F. Fulop, Paul R. Norton, *Proc. of SPIE* 6940: 69402L. Courtesy of Selex ES Ltd.

2-D gated imagers have processing requirements equivalent to passive imaging systems, which are significantly less than those required for 3-D imaging. An inexpensive 2-D active imager that can form both passive and active images has recently been reported.[21] However, moderate spatial resolution 3-D images can be constructed using gated imaging in the case where a large number of images are sequentially collected and integrated.[22]

2-D gated imagers are currently fielded in the United States and in Europe. Performance characteristics of two types are given in Table 4-4. In addition to U.S. Air Force work in the late 1990s on the ERASER Program,[23] much of the early development work was done in the United Kingdom at BAE Systems and in Denmark.[24] Commercial units are relatively compact and are used on almost all types of platforms, from handheld and tripod-mounted viewers to ground vehicles and ships, and in aircraft. Current providers of component hardware and systems include Intevac (see Figure 2-5), which is working on an export-approved unit, Sensors Unlimited, SELEX (UK), DRS Technologies, and Northrop Grumman. China's interest is shown in publications referred to above, which are mostly theoretical and do not appear yet to be state of the art.

[21] R.H. Vollmerhausen, 2013, "Solid state active/passive night vision imager using continuous-wave laser diodes and silicon focal plane arrays," *Optical Engineering*, 52: 043201, April.

[22] J. Busck and H. Heiselberg, 2004, "Gated viewing and high-accuracy three-dimensional laser radar," *Applied Optics* 43 (24): 4705.

[23] P.F. McManamon, M. Huffaker, and G. Kamerman, 2010, op. cit.

[24] McManamon et al., op. cit., and Busck and Heiselberg, op. cit.

FIGURE 2-5 Intevac Laser Illuminated Viewing and Ranging (LIVAR) camera. SOURCE: Courtesy of Intevac, Inc.

The fundamental limit to this technology is atmospheric transmission, which limits the range of the sensor. Increasing the laser pulse energy and detector sensitivity and/or reducing the illuminated area can extend the range of the sensor. Atmospheric turbulence is not usually a significant concern as long as the transmission is adequate to close the link budget. The snapshot time duration is short enough that the atmosphere is essentially frozen for each image, and the atmospheric distortion is usually not the limitation in angle/angle resolution. In general the short-pulse, high-energy lasers used for illuminators operate, by their nature, on a number of longitudinal (and often spatial) modes, and thus may be broadband enough in wavelength to eliminate speckle in a single pulse. Techniques such as broadband seeding of the high-energy source by a low-power semiconductor laser may be used to further spoil the laser coherence. If necessary, any remaining speckle can be averaged out using multiple 2-D images.

While not a physical fundamental limit, gated imaging requires high per-pulse energy to illuminate the scene (tens to hundreds of mJ) at any useful range. Therefore, the operating wavelength must be in the eye-safer region of the spectrum where optical radiation cannot focus on the retina (see Box 1-4), requiring solid state lasers with high per-pulse energy and detectors in the 1.55-2 μm range, which are more challenging to manufacture than 1 μm or visible lasers. Single fiber lasers cannot achieve this illuminator performance. Intevac cameras have limited life due to plasma etching of internal components, which limits them to low-repetition-rate systems.

Over the next decade, the committee expects to see the proliferation of 2-D gated imagers, particularly for marine and airborne systems. Their advantages are relative simplicity, technical maturity, good utility, and the ability to utilize existing laser rangefinder transmitters that are already in targeting systems for added functionality. Improvements in detector sensitivity and laser efficiency will extend the useful range of 2-D active imaging systems.

THREE-DIMENSIONAL DIRECT-DETECTION ACTIVE IMAGING

Ladar has been used extensively to create accurate and precise high-resolution angle/angle/range-resolution digital elevation models of tactically significant geographic areas. These can provide true 3-D point clouds[25] with a high density of information. The digital terrain elevation data (DTED) point spacing depends practically on the position of the sensor: 1-10 m angle/angle resolution for airborne laser

[25] A point cloud is a set of data points in some coordinate system, often intended to represent the external surface of an object.

scanning (ALS), and millimeter to centimeter angle/angle resolution for terrestrial laser scanning (TLS). Helicopter-based ALS can give a higher angle/angle resolution than aircraft-based ALS and allows orienting the scanner in all directions. A recent book[26] provides a comprehensive compilation of the principles and fundamentals of ladar technologies as well as the state-of-the-art of the performance aspects of the sensors (components, calibration, waveform analysis, quality control of ladar data, filtering and feature extraction techniques). In most references the term lidar has been used for 3-D mapping, but since the convention in this study is to use the word ladar for active EO imaging of surface reflecting targets, the committee will use ladar for 3-D mapping.

Scanning 3-D Ladar

A scanned laser radar is one technology for collecting 3-D images. Figure 2-6 is a block diagram of a typical bistatic,[27] direct-detection scanned ladar. A laser generates an optical pulse that is shaped and expanded to reduce its divergence. It is then directed toward the scene to be interrogated by a scanner. The backscattered light is directed into a collector (e.g., a telescope), where it is focused onto a photosensitive element. The resulting electronic signal is filtered to remove noise and analyzed to determine the time of arrival of the reflected optical signal. The distance to the scene at the point where the transmitted light was reflected is determined by the round-trip time of the radiation. The scanner repositions the interrogation point and the process is repeated.

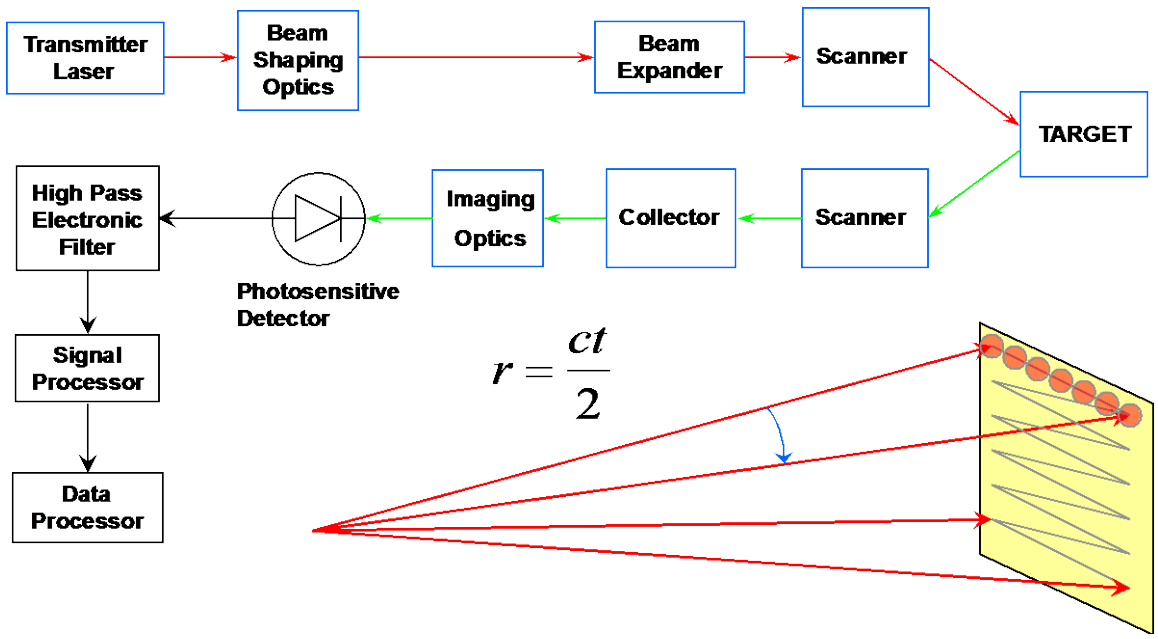

FIGURE 2-6 Block diagram of a typical bistatic, scanned 3-D imaging laser radar.

[26] J. Shan and C.K. Toth, eds., 2008, *Topographic laser ranging and scanning: principles and processing*, CRC Press.

[27] Bistatic means that separate apertures are used to transmit and receive the optical radiation.

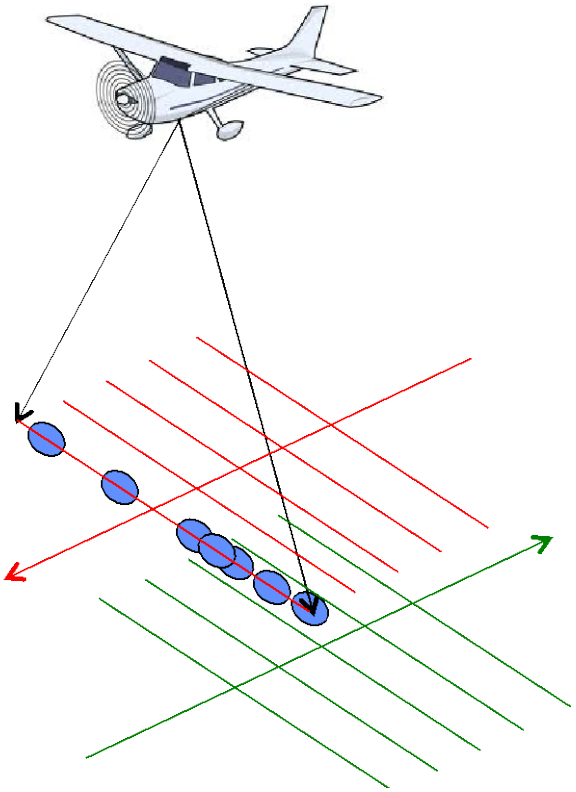

FIGURE 2-7 One scanning 3-D ladar scan pattern.

If the scanner follows a raster or line pattern, as shown in Figure 2-6, then the data will resemble a television image. Here, however, instead of the data consisting of two angles and intensity, it is composed of two angles and range. It is a representation of the geometric shape of the scene portrayed in spherical coordinates originating at the sensor. By trigonometric coordinate transformations, these data can be converted into another format such as Cartesian coordinates. If the location and orientation of the sensor are known, then the Cartesian representation closely resembles a high-resolution digital topographic map.

Such "flying spot" scanned systems are currently the rule for commercial airborne topographic mapping as well as ground-based (tripod) systems. These commonly use a line scan in one dimension, typically across the aircraft flight path. The physical motion of the aircraft itself accomplishes scanning along the flight path (Figure 2-7). A swath is then mapped. Subsequently, overlapping swaths are stitched together to map wider areas, as shown in the figure. Frequently, the ground spatial resolution or ground sampling distance in flying spot systems is limited by the transmitter spot size at the target, and not by the receiver optics. The instantaneous field of view (IFOV) of the receiver optics is often designed to be larger than the transmitter spot size in order to simplify transmitter to receiver alignment. This is in contrast to the focal plane systems to be described below. A typical commercial airborne topographic mapping system is shown in Figure 2-8.

This class of instrument typically uses only a single photosensitive element to receive the reflected radiation. The acquisition cost of sensors that employ this "one pulse-one pixel" methodology is lower than sensors with multiple detector elements but the methodology limits the data collection rate, which in turn limits the area coverage rate at the desired ground sampling distance. The commercial collection strategy requires multiple passes and additional flight time to cover a given collection area. This is acceptable for commercial topographic mapping systems but may not be acceptable for military

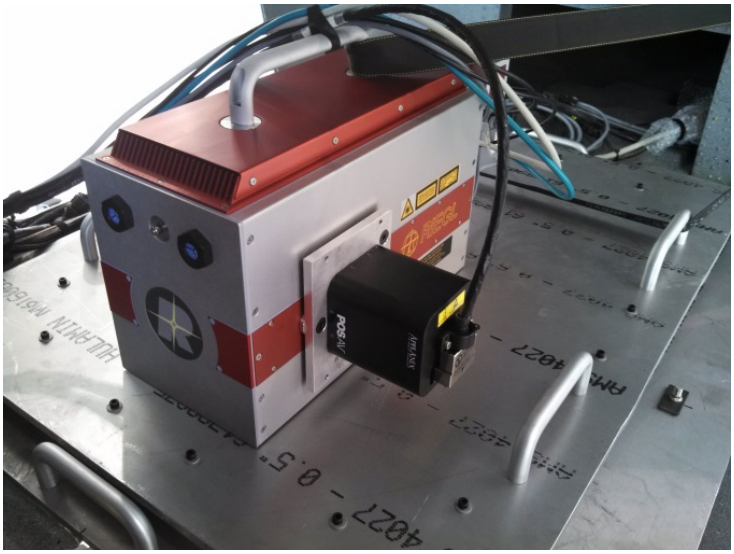

FIGURE 2-8 RIEGL LMS-Q780 Airborne Laser Scanner.
SOURCE: Courtesy of Riegl USA.

applications. Area coverage rate is a key attribute of military sensors as multiple passes are frequently not permissible. Multiple passes are not practical in a hostile airspace or where this collection strategy might alert opposing forces.

Commercial flying spot scanners, or airborne line scanners, usually operate at lower altitudes than do military systems, which must operate at higher altitudes in order to avoid hostile fire. Operating at lower altitudes permits the use of lower sensitivity detectors, smaller aperture sizes, and lower power lasers. Overall, this reduces the acquisition cost of the laser radar and limits eye-safety concerns. This lower cost is highly desirable in a commercial environment but these commercial sensors cannot be applied to many military applications as they require altitudes lower than the effective altitude of many common air defense threats.

Commercial, ground-based survey laser radar systems generally use a second, orthogonal scan element instead of gross sensor motion to generate a raster. The area coverage rate of these ground-based laser radar systems is also limited by the one-pulse-one-pixel collection strategy. It is notable that even though flying spot scanned 3-D imaging laser radars were first demonstrated in the United States, U.S. industries have not taken advantage of this early technological advantage. The majority of commercial airborne mapping and all ground-based surveying laser radars are currently produced outside the United States. Based on total operating units, the commercial market is currently dominated by Optech (Canada), Riegl (Austria), and Leica Geosystems (overall headquarters in Switzerland).[28]

3-D Flash Imaging

As illustrated in Figure 2-9, flash imaging is achieved by flood-illuminating a target scene or a portion of a target scene. The receiver collects the backscattered light and directs an image of the target

[28] M.S. Renslow, ed., 2012, *Manual of Topographic Lidar*, American Society for Photogrammetry and Remote Sensing, 34.

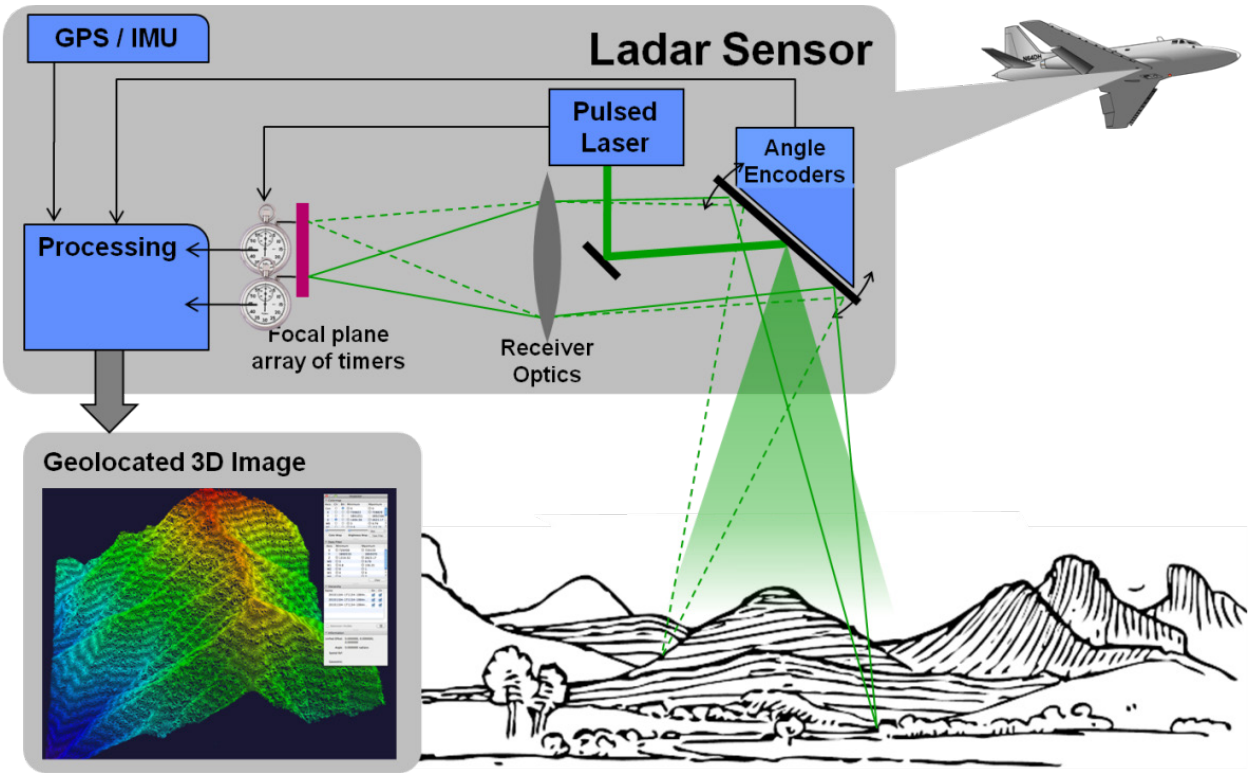

FIGURE 2-9 Concept for 3-D laser radar that creates a 3-D image from a single laser pulse. IMU, inertial measurement unit. SOURCE: Lincoln Laboratory MIT, 2011, *Tech Notes: Airborne Ladar Imaging Research Testbed,* www.ll.mit.edu/publications/technotes/TechNote_ALIRT.pdf.. Reprinted with permission of MIT Lincoln Laboratory, Lexington Massachusetts.

scene onto a 2-D array. This array pattern can then also be scanned for further scene coverage. Using the location and pointing information from the sensor platform, the range for each pixel can be converted into a location in absolute space.

Imaging over a larger area presents the challenge of achieving both detector sensitivity and speed. For flash imagers, the energy returning to the receiver is divided among multiple detectors. Thus, imaging of targets containing many pixels in angle/angle space requires high peak illumination power and/or very sensitive detectors compared to single-pixel scanners. If multiple range returns are contained in a single detector angular subtense (DAS), then it can divide up the returned signal even more, further increasing the required energy per pulse. One method to increase sensitivity of a direct detection receiver is to immediately amplify the return signal, so that this signal level is above noise sources introduced downstream from the initial detection. Avalanche photodiodes (APDs), discussed further in Chapter 4, are often used to increase sensitivity. Linear mode APD detectors have a gain that provides a linear relationship between the number of photons received and the amplitude of the signal output from the detector. Geiger-mode APDs have a huge gain regardless of the number of photons input. Fiber preamplification before detection is also possible, but to date has not been used with arrays of ladar detectors, though it is used with single detectors to increase sensitivity.

Intensity-encoded imaging addresses a different issue. Instead of having high-bandwidth readout circuits, intensity-encoded imagery uses polarization rotation to measure range, which means that simple framing cameras can be used to do 3-D flash imaging. Each of these technologies is described below.

Figure 2-10 shows an early 3-D flash imager. As described above, scanning limits the speed of image acquisition and adds to the size, weight, and power (SWaP) requirements as well as the cost of the system.

FIGURE 2-10 An early flash imager, ASC's 3-D FLC ladar. SOURCE: TigerEye 3D Flash LIDAR™ Camera (45° fixed lens); 2006 for robotics. Used by permission Advanced Scientific Concepts, Inc. (left). The TigerCub 3D camera captures 16,384 pixels with each laser pulse up to 30 frames per second. Used by permission Advanced Scientific Concepts, Inc. (right).

ASC's 3-D ladar, shown in Figure 2-10, "illuminates the scene, records time-of-flight laser pulse data onto a detector array, and generates precise "point cloud" data (video streams) on a per frame basis."[29] "Each pixel is "triggered" independently, allowing capture of 16,384 range data points to generate the 3-D point cloud image."[30]

High-speed and stable optics are needed to steer the beam since any pointing errors can cause gaps or overlap in the image or cause the image to jump from frame to frame. Optical stability needs to be a fraction of the DAS. Also, since it takes a finite time to scan a complete frame, any relative motion between the target and the sensor platform will cause the image to be distorted.[31] Therefore, for applications requiring coverage of larger areas, or real-time imaging of moving targets, it is desirable to capture the 3-D image with a large array of detectors over a very short period of time (<30 Hz). Ideally, these collections could be carried out with a single laser pulse.[32] Systems that utilize arrays of detectors to collect in this manner are referred to as flash imaging systems. 3-D flash imaging can be ideal for these applications and has been implemented in several operational airborne ladar systems (e.g., ALIRT, HALOE). 3-D flash imaging can be divided up into 3-D imagers that collect all the information on a single pulse and 3-D imagers that must take multiple snapshots of the same area many times. The multiple-snapshot flash imagers can suffer from some of the same motion limitations suffered by 3-D scanning ladars, since multiple pulses over time are required to form an image. They do however gain the benefit of not requiring high-resolution tracking to a fraction of a pixel.

APD-Based Imaging

Arrays of Geiger- and linear-mode APDs are often used to amplify the return signal to bring it above other noise sources in imaging applications. The fundamental differences between these systems revolve around how the gain is achieved by the two types of detectors. A comprehensive description of

[29] V. Molebny, G. Kamerman, O. Steinvall, 2010, "Laser radar: from early history to new trends," *Proc. SPIE* 7835:783502, Electro-Optical Remote Sensing, Photonic Technologies, and Applications IV.

[30] Ibid.

[31] R. Stettner, H. Bailey, and R. Richmond, 2004, "Eye-Safe Laser Radar 3-D Imaging," RTO-MP-SCI-145.

[32] B. Aull et al., 2002, "Geiger-mode avalanche photodiodes for three-dimensional imaging," *Lincoln Laboratory Journal* 13 (2): 335.

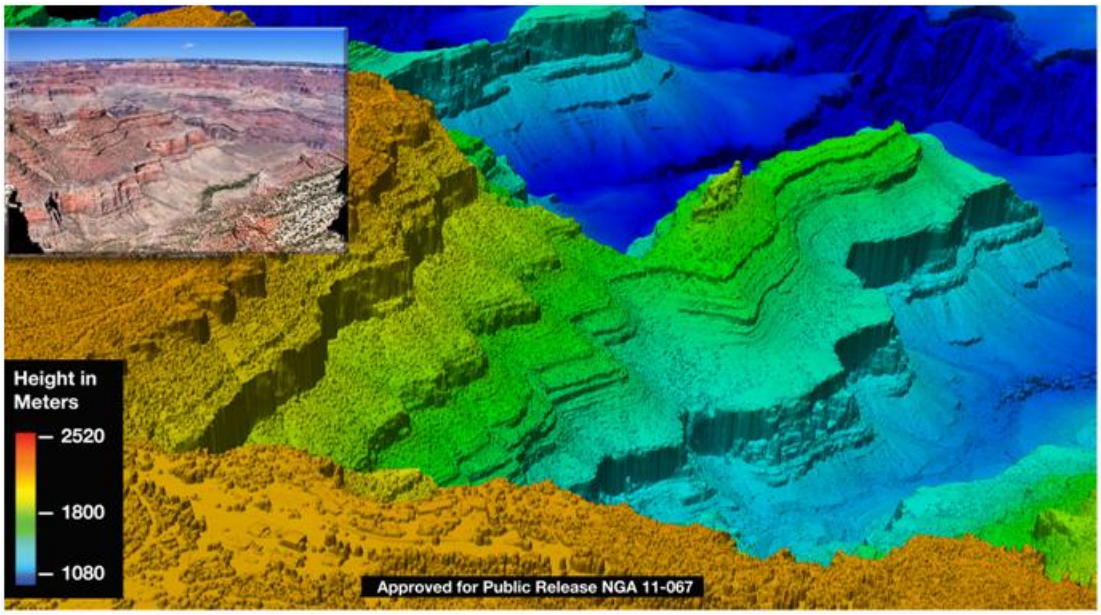

FIGURE 2-11 ALIRT imagery of the Grand Canyon obtained during a single flight. Colors represent heights over a 1,000+ meter range. The inset displays a wide-angle visible-camera image of a similar view. SOURCE: Lincoln Laboratory MIT, 2011, *Tech Notes: Airborne Ladar Imaging Research Testbed,* www.ll.mit.edu/publications/technotes/TechNote_ALIRT.pdf. Reprinted with permission of MIT Lincoln Laboratory, Lexington Massachusetts.

linear- and Geiger-mode APD (GM-APD) arrays, including discussion of materials and fabrication techniques, can be found in Chapter 4. An example of an APD-based 3-D flash image is shown in Figure 2-11.

The ultimate performance of APD-based flash ladar systems depends on the interplay of a number of parameters, including the number and sensitivity of the pixels in the APD array, the power and pulse repetition frequency (PRF) of the laser, and the operating range/altitude of the platform carrying the system. This interplay is illustrated in Figure 2-12, which shows the area collection rate as a function of the aircraft altitude above the ground (or range) for a fixed ground sample distance (GSD) of 30 cm. While this example uses a GM-APD based system to illustrate the trades, the same fundamental trade-offs (but not the exact numbers) hold for linear-mode systems.

The horizontal lines in Figure 2-12 show the performance for three different APD formats (from bottom to top, 32 × 128, 64 × 256, and four 64 × 256 arrays). As the size of the array increases, so does the instantaneous area collection rate. The dashed diagonal lines show the performance for different laser powers and aperture sizes. As expected, as the power and aperture increase the area coverage increases as well, while the performance at a fixed GSD degrades as the altitude increases. The power lines in this chart each assume a 30 cm GSD. A final limitation is imposed by the scan width limitations (most flash imaging ladars still scan to increase their area rate) and the speed of the platforms carrying the system.

Consider a baseline laser design with a 2 W, 20 kHz PRF laser and a single 32 × 128 GM-APD behind a 10 cm aperture. As shown by the red dot in Figure 2-13, the optimal performance point for this system is at approximately 12,000 ft. Increasing the laser power to 30 W alone will allow operation at higher altitudes but will not increase the overall area coverage (green dot). At this point the APD size is limited. If the APD size is increased to 64 × 256, the area rate can be increased, but the aircraft must operate at a lower altitude to achieve this coverage (blue dot). In this case the limit is the laser power. For

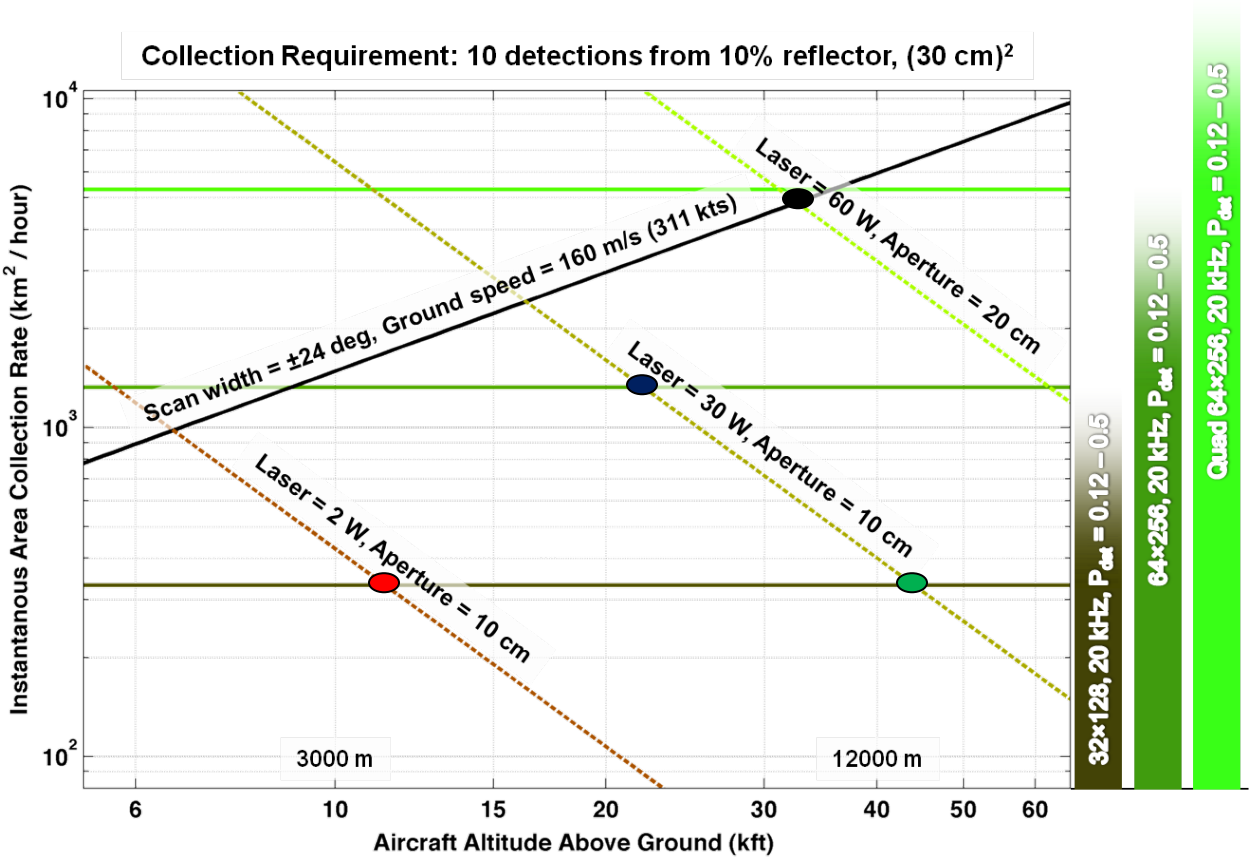

FIGURE 2-12 Example laser radar performance trade-offs for fixed GSD. Geiger mode is assumed here for illustrative purposes. SOURCE: MIT Lincoln Laboratory.

some scenarios, operation at lower altitudes is acceptable, but high altitude operation is desirable where survivability is a concern. Finally, increasing the size of the APD to four 64 × 256 arrays in parallel and increasing the laser power and aperture allows operation at the optimal point in the scan-limited regime at higher altitudes (black dot).

Geiger-Mode Imaging

Geiger-mode imaging refers to imaging based on a detector technology that utilizes a massive amplification of signal no matter how many photons are received. This makes GM-APDs very sensitive photon-counting detector arrays. The term photon-counting is often used interchangeably with detector sensitivity capable of single-photoelectron detection. In fact photon counting is exactly that—the ability to count the exact number of received photons. This means that GM-APDs have the ability to detect the arrival of a single photon. Active sensors that can detect very weak signals of interest can have relatively superior performance in mission areas that are SWaP-constrained and/or require extended ranges or rapid decisions, because an increase in detector sensitivity directly reduces required laser power. There are several methods of implementing photon-counting detector arrays, but Geiger-mode APD array-based systems are currently the most mature and widely used.

As discussed in Chapter 4, Geiger-mode imaging takes advantage of the fact that for low incident photon rates per laser pulse (~0.2 photons per pulse), the measured photon rates for a GM-APD ladar system approach that of a "perfect" detector (i.e., one that detects every incident photon at every pulse). Hence, Geiger-mode ladar systems operate most efficiently at a low per-pulse probability of detection where they collect multiple pulses to obtain high probabilities of detection of a target surface. The output of the detector is then an ensemble of independent measurements that are histogrammed to give a single range measurement. An example of imagery obtained with ALIRT, a fielded GM-APD system, is shown in Figure 2-11.

The assembly of a device with a high probability of detection using multiple pulses creates a sensitivity to motion not unlike the sensitivity scanning sensors have; however, the timescale for the measurements is much shorter than for scanning systems. Still, any scene motion must be compensated for prior to adding detections. GM-APD arrays can operate with 10-200 kHz frame rates, which enables the ensemble detections to be made very quickly (e.g., 1-10 msec for 10-100 measurements at 10 kHz). The fast frame rates also allow the use of lasers that run at rather high PRF with low energy per pulse. "This low energy requirement allows the use of 1.06 μm radiation without eye hazard because of low single-pulse intensities."[33] This is a significant feature since it enables high-performance active sensing with SWaP-constrained platforms.[34] High frame rates also increase the laser duty cycle, which is favorable from a point of view of laser cost, weight, and efficiency and enables fiber laser solutions. However, operation at high pulse repetition rates reduces the unambiguous range[35] for the system.

Initial Geiger-mode flash ladar systems were pioneered by MIT Lincoln Laboratory (MIT/LL), which has made Geiger-mode APD cameras with up to 64 × 256 detectors.[36] MIT/LL developed the ALIRT laser radar system (shown in Figure 2-13), which was deployed in 2010 to Afghanistan and provides wide area mapping and point targets. The DARPA High Altitude Lidar Operational Experiment (HALOE) also uses Geiger-mode APD arrays to provide high angle/angle resolution (20 cm), wide area, off-nadir 3-D mapping and target information. This system was also deployed in the Afghanistan theater.[37,38] These two systems have been responsible for wide area 3-D maps of Afghanistan, providing digital elevation data of the country with unprecedented resolution. More recently, MIT/LL has been developing the MACHETE ladar system for USSOUTHCOM. This ladar, which was based on the ALIRT system, represents the state of the art for GM-APD-based flash ladar systems with 16 times more pixels and a much more capable laser developed by Raytheon Space and Airborne Systems.

Two U.S. companies have commercialized Geiger-mode APD arrays. Both Princeton Lightwave and Boeing Spectrolab have 32 × 32 array based cameras available at 1.06 μm and at 1.55 μm (see, for example, Figure 2-14), and both are developing 32 × 128 format cameras.[39] Initial Geiger-mode APD cameras operated at a PRF of about 20 kHz, but advances in the readout integrated circuits (ROICs) by

[33] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[34] P.F. McManamon et al, op. cit.

[35] The unambiguous range is the range at which the returned signal can be known to be unquestionably at that range.

[36] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[37] DARPA 2011 Congressional testimony, available at http://www.darpa.mil/WorkArea/DownloadAsset.aspx?id=2929.

[38] B. Thompson, 2011, "AFRL plays pivotal role in response to urgent operational need in Afghanistan," *Inside WPAFB*, http://www.wpafb.af.mil/news/story.asp?id=123281012.

[39] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

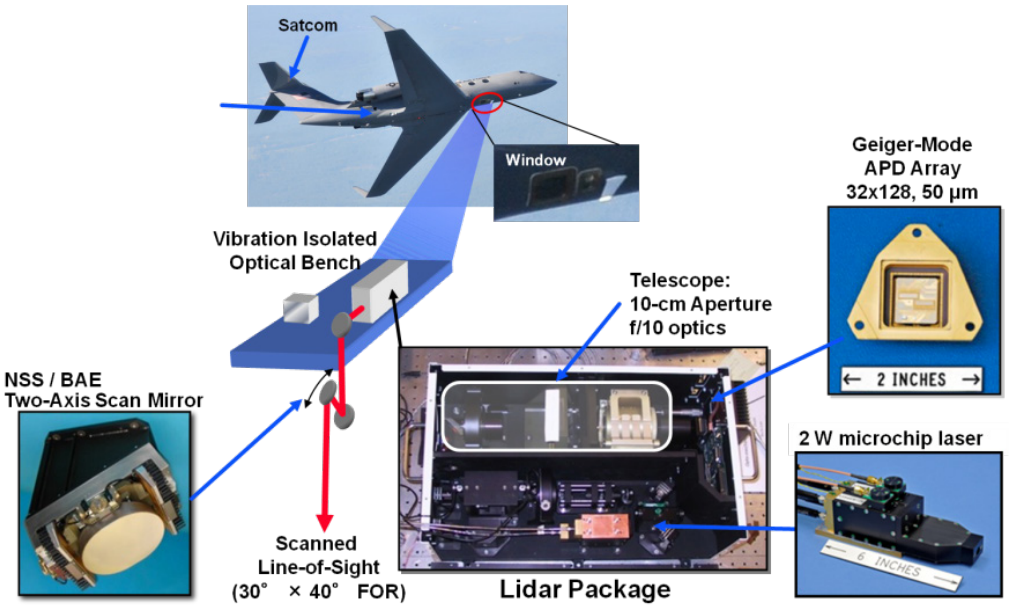

FIGURE 2-13 The ALIRT 3-D laser radar system and associated hardware. SOURCE: MIT Lincoln Laboratory.

FIGURE 2-14 Princeton Lightwave 32 × 32 GM-APD-based flash ladar camera. SOURCE: Courtesy of Princeton Lightwave, Inc.

Princeton Lightwave allow operations up to 186 KHz for the 32 × 32 array.[40] The 32 × 128 array will have a lower maximum frame rate, probably about 105 kHz.[41] The commercial availability of such cameras may lower the barrier to entry for others seeking this type of capability.

Linear-Mode Active Imaging

Another method for flash imaging uses the same overall system architecture shown in Figure 2-13, but uses linear-mode APD arrays instead of Geiger-mode APDs. Linear-mode APDs are designed with the goal of operating below the breakdown voltage with high gain and low noise. The operation of

[40] P.F. McManamon et al., op. cit.
[41] Personal communication from Mark Itzler, Princeton Lightwave.

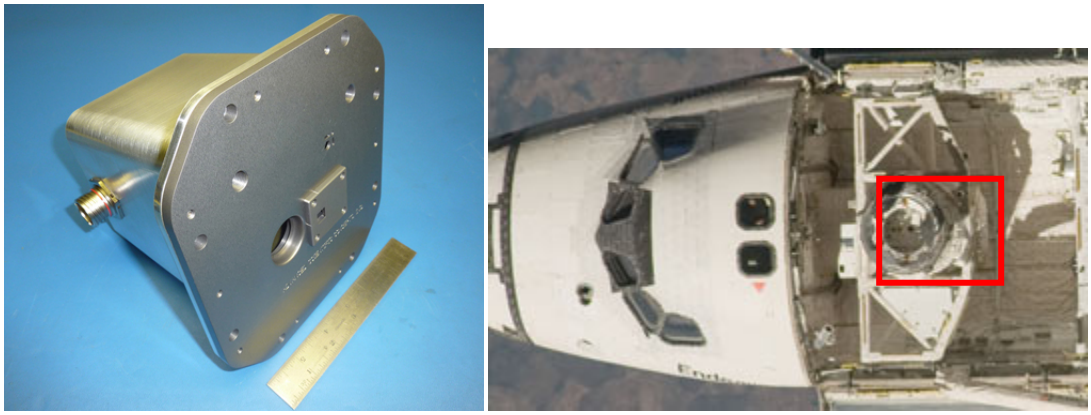

FIGURE 2-15 ASC's 128 × 128 pixel DragonEye 3-D flash ladar camera shown alone and as installed on the NASA Endeavour space shuttle. SOURCE: *Left:* DragonEye 3D Flash LIDAR™ Space Camera (45°FOV) 2011 for International Space Station rendezvous and berthing. Used by permission Advanced Scientific Concepts, Inc. *Right:* NASA.

these APDs produces an average photocurrent that is linearly proportional to the incident optical flux (hence the name linear mode).[42] A detailed description of linear-mode APD arrays can be found in Chapter 4.

Unlike Geiger-mode systems, which must geo-register and aggregate multiple-high-frame rate returns, linear-mode systems have the advantage of collecting range and intensity information in a single frame, assuming the ROICs are made to support this (making them true "flash imaging" systems). However, these systems usually operate at lower frame rates (~30 Hz), requiring higher laser energy per pulse. Historically, linear-mode APDs also have not been as sensitive as GM-APDs, requiring additional total laser energy per sample. Both factors currently combine to limit the ability to use lower SWaP, efficient, fiber lasers. More sensitive linear-mode APDs operating at higher frame rates may be able to use fiber lasers for sources. As single-photon-counting linear-mode APDs that operate at higher frame rates become widely available, the SWaP for future linear-mode APD 3-D ladars will be reduced.

Linear-mode 3-D flash imaging systems and cameras have become commercially available. ASC pioneered this approach, especially for commercial applications, and offer several different camera models for different applications. All of the available cameras utilize 128 × 128 pixel arrays that frame at 1 to 20 Hz or at 30 Hz in burst mode.[43] The DragonEye system (shown in Figure 2-15) was developed for space applications and was flown on the Endeavour Space Shuttle in 2009.

Recently, there has been significant progress in developing high sensitivity linear-mode APD arrays by companies such as Raytheon and DRS Technologies.[44] High gain in the APD reduces the effect of any noise introduced after the amplification stage, as is discussed in more detail in Chapter 4. This increased sensitivity enables photon detection and counting in single-to-few pulses, thus ultimately reducing the discussion of APD-based flash ladars to photon-counting and non-photon-counting systems rather than to Geiger-mode and linear-mode APDs.

[42] B. Aull et al., op. cit.

[43] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[44] Ibid.

Comparison of Linear- and Geiger-Mode Systems

In this section the relative qualitative advantages and disadvantages of Geiger-mode and linear-mode 3-D flash imaging systems are discussed (a comparison at the detector level is deferred to Chapter 4).

Long-Range Applications

The single-photon sensitivity, low readout noise, and accurate timing enabled by the Geiger-mode ladar design mean that high-precision and high-accuracy spatial information (range or time-of-flight) can be measured with minimal return optical signal, making it well-suited to long range and/or high altitude applications. These situations also tend to have limited angular motion, making the required motion compensation less difficult.

Currently fielded linear-mode flash ladar systems require hundreds of photons to create images. The requirement for more photons drives the need for higher laser energy per pulse at relatively low repetition rates and makes it harder to use fiber lasers. Current systems are not designed for precision wide-area mapping at long ranges. The ongoing development of highly sensitive, linear-mode APD arrays will reduce the need for these high-energy-per-pulse lasers and provide high-quality, long-range sensing capability.

High Background Noise Environments

In high background noise environments (e.g., reflections from sunlight, etc.), the photons from the background will be captured along with the target returns. Since linear-mode systems capture the entire return in a signal pulse, the laser power and detection threshold can be turned up so that only the signal return is increased and the background noise remains constant on a per pulse basis.

As described above, Geiger mode systems operate best with low per-pulse probability of detections, so the practice of "turning up the volume" to overcome the background noise is not an effective solution. Instead, Geiger-mode systems spread the energy across multiple (tens to hundreds of) pulses and histogram the returns to create a detection. In a high-background noise environment, the noise will be present in every pulse and therefore may be binned instead of the signal. This challenge may be overcome in many situations by implementing range gating, since smaller range gates offer less time and fewer opportunities for background photons to be collected. If there is a priori knowledge of where to set the range gates (which is often true for ground target imaging in the open), the range gate can be set to be very small (e.g., nanoseconds to microseconds[45]). This is generally sufficient to block out most background events, especially in the open. In practice, fielded Geiger-mode APD systems have not found blocking losses to be a significant issue in actual operation. High background noise may still present challenges for the Geiger-mode systems in foliage penetrating or underwater imaging regimes, but again, this has not been found to be an issue in operations. In fact, as described in the "Imaging Through Obscurants" section below, Geiger-mode systems work well in foliage penetrating environments.

Moving Targets

The binning of multiple pulses for Geiger-mode systems drives the need for some platform motion compensation before the returns are summed. If the target itself is moving, there may be blurring in the 3-D image as the target moves during the image collection time. However, due to the high frame rates of the Geiger-mode systems, the speed at which the vehicle must move to induce significant blurring is generally high and will usually only be an issue for fast-moving targets in regimes requiring long collection times (e.g., in very dense foliage).

[45] M. Entwistle, M.A. Itzler, et. al., "Geiger-mode APD Camera System for Single Photon 3-D Ladar Imaging," *Proc. SPIE* 8375, Advanced Photon Counting Techniques VI, 83750D (May 1, 2012).

Since linear-mode systems collect all of the return information on a single pulse, it is possible to create a snapshot of a moving target with minimal blurring even for fast movers. This snapshot imaging capability also reduces the processing required to form images. Due to motion compensation issues, forming an image with a GM-APD sensor is more complex than forming the same image with an LM-APD sensor.

Resolution

As with the 3-D scanning systems described in the preceding section, the range resolution of a flash ladar system is driven by the pulse width of the laser and cross range resolution by the detector angular subtense of the detectors and possibly the diffraction limit of the receive optics.

In Geiger-mode systems, the timing readout circuitry is realized digitally. It is read out noiselessly and hence can handle larger bandwidths (~1 GHz) without being subject to increased noise. This enables operation with pulse widths of ~1 ns to obtain 10-20 cm range resolution.

Because linear-mode systems use analog electronics for the readout circuitry, larger bandwidths introduce more noise to the system. Hence, these systems tend to operate in the hundreds of MHz regime, resulting in lower range resolution (although the range precision can be quite high with thresholding). Voxtel has reported the development of 128 × 128 multi-gain stage InGaAs APD arrays with gigahertz bandwidth,[46] but these have not yet been put into a ladar system. There is ongoing work to develop linear-mode systems with gigahertz bandwidths.[47]

Gray Scale Intensity Images

It is often desirable to form an intensity or gray scale image of an area or target in addition to the typical range/height images. Overlaying this type of information on a 3-D image provides a more camera-like picture that may be more intuitive to interpret. No matter what type of sensor is used, measuring a gray scale dynamic range requires enough samples to form that level of gray scale.

Linear-mode systems can measure gray scale on a single pulse, since the output is proportional to the reflected light.[48] However, in order to measure gray scale, enough photons must be returned in order to see the dynamic range associated with a given gray scale. Therefore, the required energy per frame for a photon-counting LM-APD sensor will increase as the desired gray scale dynamic range increases.

With Geiger-mode imaging there is no ability to measure the signal intensity per pixel (or gray scale image) on each pulse, because any received photon causes the same massive signal response.[49] However, in the regime where the probability of detection per pulse is kept low, a higher reflectance area will have a higher probability of return, causing more events to trigger in those areas.[50] Accounting (histogramming) for the total number of counts (signal trigger events) from each pixel across multiple pulses results in an effective gray scale but reduces the effective frame rate of the system. It also increases the energy required for a frame.[51] Geiger-mode flash imaging becomes less efficient for gray scale or range profiles because multiple (typically ~1,000) looks are needed to bin enough pulses to create these products. However, the high frame rates used in these systems still make this an efficient way to gather these data, and for typical operating parameters the number of samples (and hence "frame" energy)

[46] See for example http://voxtel-inc.com/products/single-element-detectors-and-pixelated-detector-arrays/.

[47] J. Asbrocket et al., 2008," Ultra-high sensitivity APD based 3-D LADAR sensors: linear-mode photon counting LADAR camera for the Ultra-Sensitive Detector program," *Proc. of SPIE* 6940: 2-3.

[48] P. McManamon, 2012, "Review of ladar: a historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901

[49] Ibid.

[50] Ibid.

[51] Ibid.

required to form the gray scale image is well below what is required to measure the fundamental limit of dynamic range.

Linear-mode cameras can measure multiple range returns on a single pulse but require a more complex ROIC in order to collect the multiple range bins. Once multiple range returns are collected, images can be range gated to show only the returns beyond an obscurant (e.g., foliage, camouflage, or dust). Figure 2-16 shows gated imagery through a sand cloud, from an ASC 3-D imager.[52]

Geiger-mode systems can also be operated in a way to enable penetration of obscurants. However, the emitted laser energy must be set so that the system can detect the returns from bright scatterers (e.g., treetops or other obscurants) without saturating while still detecting weaker signals beyond the obscurant with the right probability of detection. If one has pixels with mixed range returns and collects enough samples to detect both the weak and strong target returns with some probability, "one can play essentially the same trick as used with gray scale in to map the returns as a function of range…If the probability of triggering is low for any event, one will get events triggering at longer ranges" through the holes in the trees.[53] By peeling away the canopy layer in processing, the features of the underlying layer can be seen, as shown in Figure 2-17. In this case, the high-PRF operation of the Geiger-mode systems are an advantage because there are increased opportunities "per frame" to see through different holes in the canopy, thus increasing the probability of detecting the target as well as increasing the aspect diversity on the target, making the target easier to pull out of the background clutter. Unlike linear-mode sensors, which require a more complicated ROIC to measure multiple range returns, the ROIC complexity for a GM-APD does not change.

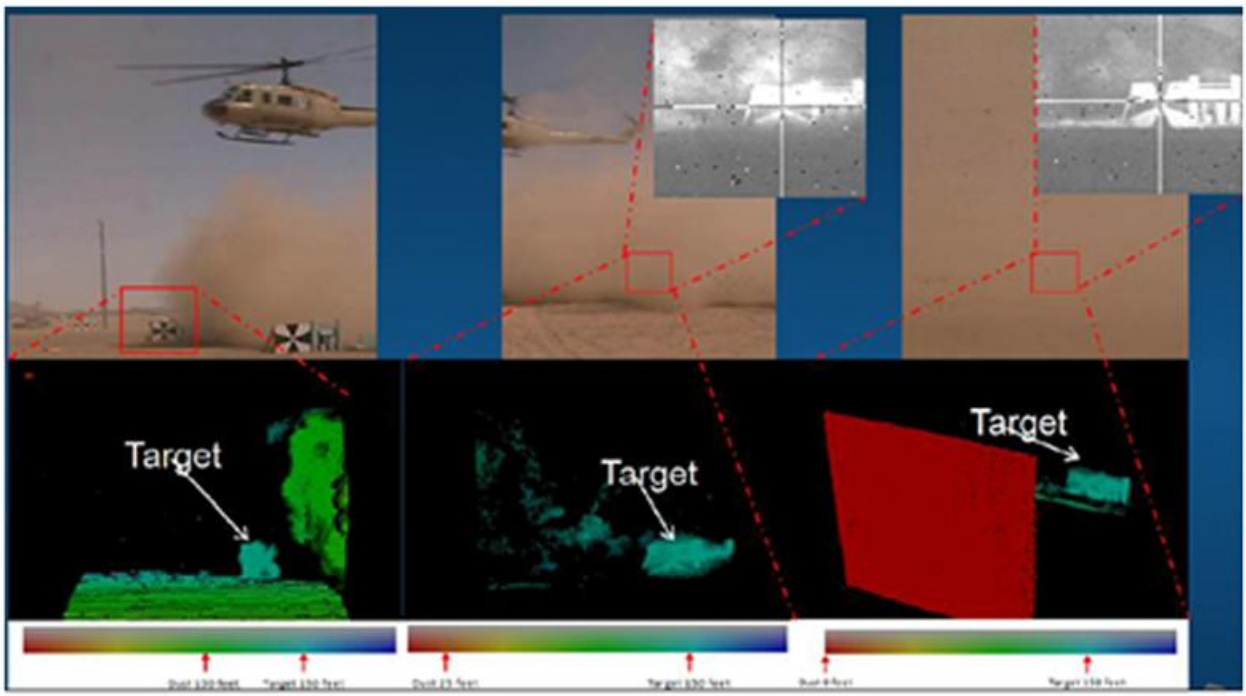

FIGURE 2-16 Flash ladar image showing penetration through dust. Visible imagery is completely obscured, while ladar imagery shows the potential hazards. SOURCE: Helicopter Brownout Landing. Used by permission Advanced Scientific Concepts, Inc.

[52] NAVAIR Public Release 11-033, "3-D Flash LADAR helicopter landing sensor for brownout and reduced visual cue."

[53] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

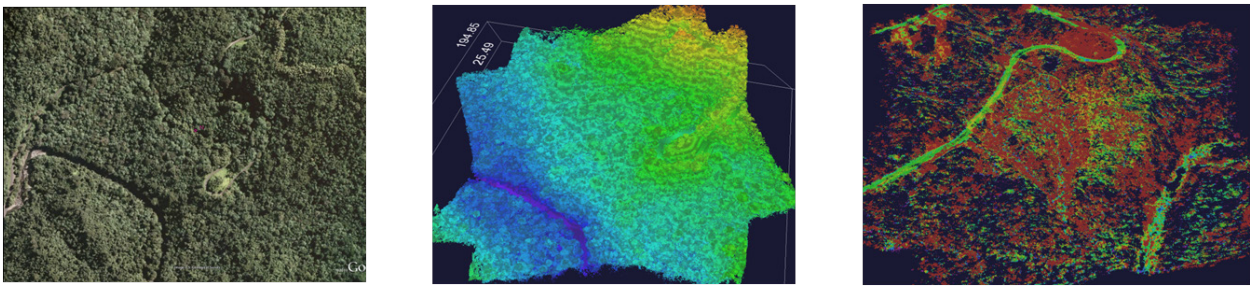

FIGURE 2-17 Puerto Rico foliage penetration (FOPEN) data collected by ALIRT over El Yunque National Forest, Puerto Rico. *Left:* EO image of the scene; *Center:* Ladar image with canopy (color indicates height above mean sea level); *Right:* Ladar image with canopy removed (color indicates height above ground). SOURCE: Dale Fried, 2012, "Photon counting laser radar," presentation at 2012 SPIE Defense Security and Sensing Conference (April 25). Approved for public release, unlimited distribution. Post-Detection Processing. MIT Lincoln Laboratory.

The high frame rates used by Geiger-mode APDs produce a large amount of data that must be read out and processed in order to produce 3-D images. This data processing increases the SWaP of the ladar systems and limits the current utility of Geiger-mode flash ladar systems for real-time operations. Instead, the current airborne systems store the data and process it later.

The processing requirements are much less demanding for linear-mode sensors because more processing is done on the ROIC, and there are fewer pulses per image to process. The image is formed directly out of the ROIC for most linear-mode sensors. However, as previously described, the more complicated ROIC impacts the performance in other areas.

Many of the limitations of the APD-based ladar systems arise from the operational limits of the APD array itself. For Geiger-mode systems, "there is a dead time after each triggered event. During the dead time, the detector will not detect any received photons."[54] Crosstalk and dark currents can also affect the performance of the ladar system. Fielded linear-mode arrays have a noise floor that is significantly higher than a single photon. Therefore, the energy required to image a given area at a given range is significantly higher for the linear-mode receiver than for a Geiger-mode receiver. Also, the bandwidth for linear-mode receivers currently limits the range resolution achievable with these systems. These attributes, and the potential advancements to overcome these limitations, are discussed in the various references previously provided and in Chapter 4.[55]

The area that can be imaged in a single "flash" will depend on the detector and the laser. While larger format and higher frame rate detector arrays will continue to develop to support coverage of wider areas, the development cycle for lasers that can provide the same energy per pulse over a larger area may be the limiting factor. SWaP may also limit the desirability of going to large flash illumination areas for longer ranges.

Another limitation of Geiger-mode laser radar is the processing needed for the ladar data. Because the data are being collected at such a high frame rate and the detector arrays are large, the data processing needs can be quite high. For example, a 64 × 256 array being read out at 20 kHz will produce on the order of 300 MB per second of data. This drives the need for large on-board data storage and/or data processing that can increase the SWaP of the sensor, and it also limits the ability to obtain real-time data from these types of systems. As the size of both linear and Geiger-mode arrays increases and linear-mode arrays are able to capture more data per pulse at higher pulse rates, the need for this kind of processing will continue to grow.

As described above and in Chapter 4, improvements in the ladar components will improve the APD-based ladar system performance but must be coupled in a smart way to optimize that performance.

[54] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[55] P.F. McManamon et al., op. cit.

Improvements in Geiger-mode APD array technology (larger arrays with better quantum efficiency, faster reset, less crosstalk, higher timing precision, faster or asynchronous ROICs, and so on) and laser technology improvements—for example high-power, short-pulse, medium-PRF (20-200 kHz) systems with high wall-plug efficiencies—are required for longer range capability. Reducing the processing and data readout times would enable production of real-time images. The development of on-chip processing (histogramming) and/or compression would reduce the amount of data that need to be read out. Conversely, the development of wideband off-chip communications paths (20-200 GB/s) would allow more data to be read out.

Continued progress in the development of high-sensitivity, photon-counting, linear-mode APDs will improve the linear-mode system performance. As described previously, there is significant ongoing work in this direction. Improvements in the APDs to reduce the readout noise will enable larger bandwidth measurements and hence, better range resolution. Like Geiger-mode systems, performance (area coverage, cross-range angle/angle resolution) will also improve with the development of larger arrays with smaller pixel pitch.

With the exception of MWIR HgCdTe-based APDs, most of the detectors discussed so far, linear- or Geiger-mode, operate in the SWIR and near infrared (NIR) spectral bands. HgCdTe APDs in MWIR and LWIR bands have been demonstrated in recent years that exhibit sensitivities comparable to the SWIR band. Lasers suitable for active sensing, such as quantum cascade lasers (QCLs), are becoming increasingly available. However, active sensing in the MWIR and LWIR bands under standard terrestrial conditions is hampered by the high thermal background flux. Thus the laser flux required is significantly greater to operate in these bands. Furthermore, optical resolution that can be achieved in LWIR and MWIR is also reduced. The advantage of operating in the MWIR and LWIR bands is that laser illumination in 2-D and 3-D modes in a narrow wavelength region can be selectively tuned for gas and or chemical sensing. These bands also enable reduced turbulence effects and atmospheric scattering compared to the SWIR band.

As both types of APD-based ladar systems also utilize scanners to increase their area rates, the development of large, lightweight, agile pointing apertures would also improve the system performance.

Geiger- and linear-mode ladar technology development is an active area of R&D, and advances such as those described above would be published in literature and would probably be presented at conferences, although some advances may be held as proprietary.

Conclusion 2-1: The distinction between linear- and Geiger-mode systems is likely to blur as highly sensitive linear-mode arrays become more mature. Instead of focusing on one type or the other, it is more important to focus on the photon counting performance, since this drives the best achievable sensitivity and therefore the system size, weight, and power.

This technology is well suited for high-resolution 3-D imaging applications from the ground, air, or space. Like all direct-detect 3-D ladar systems, it is well suited for application requiring "geometrical" information, such as 3-D mapping, line-of-sight maps, and geometric change detection. It is also the best technology to obtain high-resolution images through obscurants and under foliage over relatively wide areas. It is anticipated that a global, 3-D database derived from ladar measurements will support civil, commercial, and military needs as the foundation layer for organizing geospatial information of all types.

The United States is the leader in the development of 3-D flash ladar systems, but there is a great deal of foreign development work in APD arrays. CEA-LETI/Sofradir (France/Israel) have demonstrated HgCdTe (MCT) arrays for flash ladar in "near photon counting" mode. Selex Galileo (U.K.) has done flash ladar with MCT APDs.[56] The Milan group (Milano Politechnico) has also done work in the development of single photon sensitive APD arrays. First Sensor A.G. (Germany) sells small APD arrays

[56] See http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1342598.

for visible and near-infrared ladar,[57] but it only makes the arrays and not the electronics required for time-of-flight measurements.

Hamamatsu (Japan) also makes linear-mode and Geiger-mode APD arrays[58] (silicon for visible/near-IR) but neither is directly useful for flash ladar. In principle, the company could produce imaging arrays. The leading GM-APD group in China is at East China Normal University,[59] though their array work is still early in development.

With the development and commercialization of ladar cameras and arrays, it is likely that flash ladar systems will become more widely proliferated and less expensive. It is anticipated that commercial applications will drive this proliferation.

Intensity-Encoded 3-D Flash Imaging

As mentioned previously, one of the challenges of flash imaging is having a large enough focal plane array to detect an area-based object with a single pulse.[60] Like Geiger-mode systems, performance (area coverage, cross-range angle/angle resolution) will also improve with the development of larger arrays with smaller pixel pitch. While APD-based flash imagers use arrays with high bandwidth readout circuits to measure high-resolution 3-D images of an area of interest, intensity-encoded flash imaging leverages commercially available large framing CCD arrays to obtain 3-D images. This technique was developed to take advantage of existing laser illuminators and CCD arrays at a time when APD technology was in its infancy. In the early 1990s, multiple patents were awarded for the Laser Imaging and Ranging System (LIMARS),[61,62] which replaces the high-speed camera with a Pockels cell and two low-frame-rate cameras for flash 3-D imaging.[63] In the LIMARS receiver, temporal (range) resolution is provided by a high-speed Pockels cell, so that high-bandwidth cameras are not required.[64] Figure 2-18 shows a diagram of the LIMARS receiver.

This concept uses a high-power, short-pulse laser to illuminate a scene of interest. The reflected light enters the receiver and is passed through a polarizer, where a single polarization of return light is isolated (alternatively, twice as many cameras could be used to detect both polarizations). The light next passes into a Pockels cell, which is key in converting the time of flight of the pulse to intensity information.[65] A ramp is placed on the Pockels cell, where a phase shift between the components of the optical field is proportional to the applied voltage, as shown in the bottom of Figure 2-18. The output of the Pockels cell is in general elliptically polarized, and is passed into a polarizing beam splitter which splits the polarized energy between two CCD cameras.[66] In any given detector, representing a certain solid angle, the ratio of power in one camera to power in the other camera provides range information,

[57] See http://www.first-sensor.com/en/news/newspress/2010-11-09-matrix-apd-detector-arrays-lidar, and http://www.first-sensor.com/en/products/optical-sensors/rd.

[58] See http://www.hamamatsu.com/us/en/index.html.

[59] See, for example, Min Ren et al., 2011, "Laser ranging at 1550 nm with 1-GHz sine-wave gated InGaAs/InP APD single-photon detector," *Optics Express*, 19 (14): 13497.

[60] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[61] L. Tamborino and J. Taboda, 1992, "Laser imaging and ranging system, one camera," U.S. Patent No. 5,162,861.

[62] J. Taboda and L. Tamborino, 1992, "Laser imaging and ranging system using two cameras," U.S. Patent No. 5,157,451.

[63] P. McManamon, op. cit.

[64] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[65] R. Goldstein, 1968, "Pockels cell primer," *Laser Focus,* Feb.: 21.

[66] M.B. Mark, 1992, "Laser Imaging and Ranging System (LIMARS) Range Accuracy Analyses," WL-TR-92-1053

since targets at different ranges will have returns to the Pockels cell at different times in the waveform. If the cameras are aligned so that the same pixels on each camera image the same part of the target, spatially and range-resolved measurements are obtained.

The range information within each detector is driven by the slope of the waveform on the Pockels cell. A steeper slope provides more accurate range information but also repeats more quickly[67] and results in a smaller ambiguous range. This problem is not unique to polarization-based imaging, and a number of standard techniques (e.g., chirping the length of the ramps) can be applied to expand the unambiguous range. Any number of available CCD cameras can work for this application, including silicon-based TV cameras, NIR cameras, and military hardened SWIR cameras. All of these cameras have formats that are larger than those of the high-bandwidth cameras described in the preceding section. This technique has been used in several programs. Figure 2-19 shows two images from the DARPA SPI 3-D effort—a program aimed at developing airborne active EO sensors capable of positive target identification at standoff ranges—which uses this approach.[68] Figure 2-20 shows a very high resolution short-range image from a small company, Tetravue, again using this technique.[69]

Comparison of Intensity-Encoded and APD-Based 3-D Imaging

The main advantage of intensity-encoded flash imaging is the ability to create high- range-resolution 3-D images using standard framing cameras. These framing cameras often have a larger format than the APD arrays used in the systems above and can cover a wider area with a single pulse. In addition, the readout of these arrays can be very fast, enabling real-time imaging and even the formation of 3-D videos.

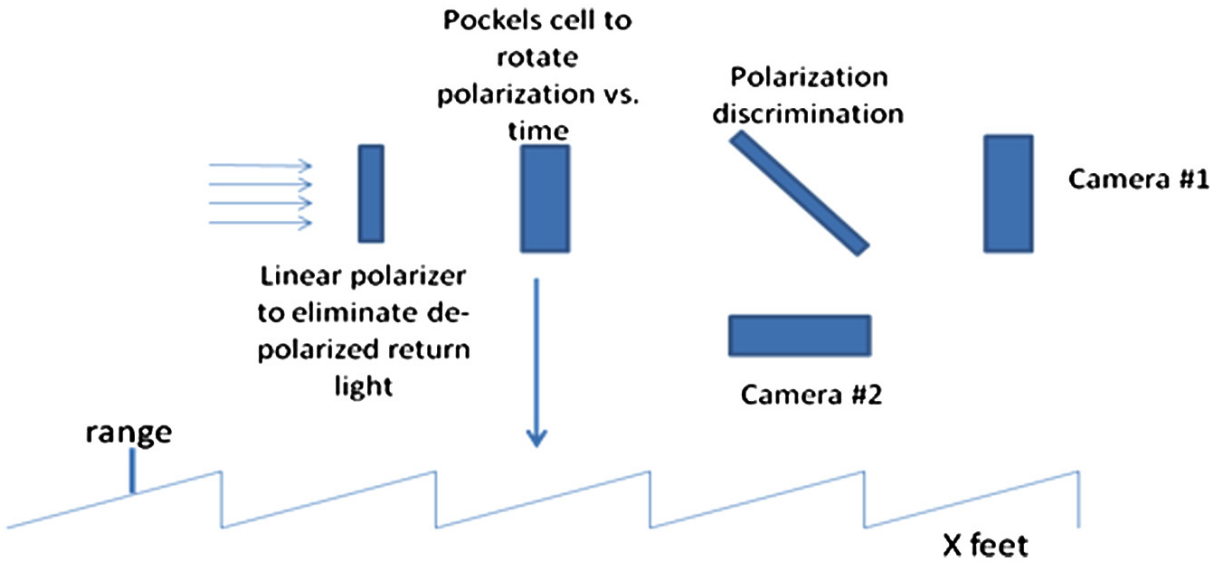

FIGURE 2-18 Diagram of the LIMARS intensity-encoded 3-D flash ladar concept. The bottom portion of the diagram shows a typical Pockels cell voltage. SOURCE: P. McManamon, 2012, "Review of ladar: a historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[67] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[68] T. Tether, 2004, testimony before the House Armed Services Committee Subcommittee On Terrorism, Unconventional Threats And Capabilities, http://www.globalsecurity.org/military/library/congress/2004_hr/04-03-25tether.htm.

[69] See http://www.tetravue.com.

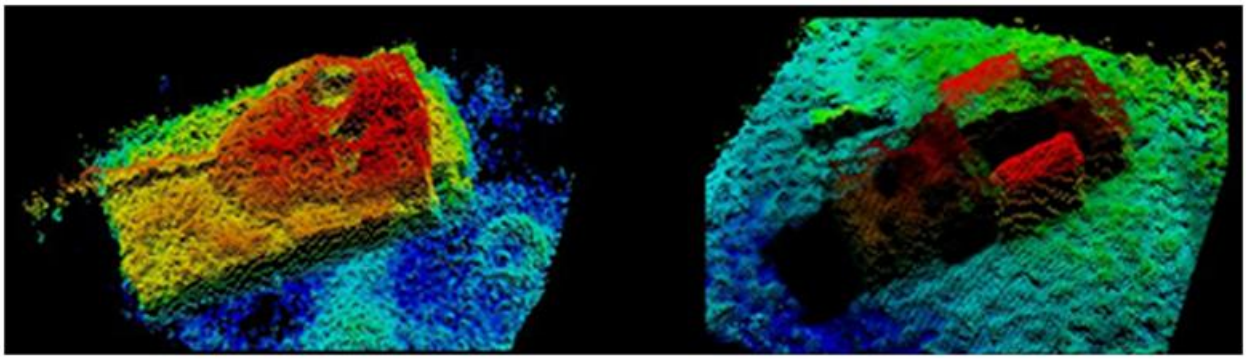

FIGURE 2-19 SPI 3-D images. SOURCE: P. McManamon, op. cit.

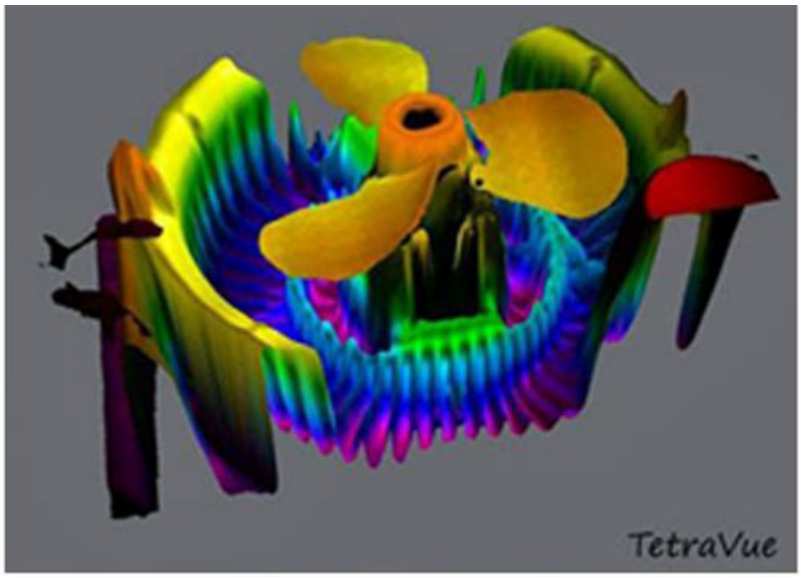

FIGURE 2-20 Polarization-based 3-D ladar image. SOURCE: High resolution 3D image, TetraVue, Inc. 2011.

The use of the Pockels cell to rotate polarization does introduce some challenges. As described above, there is a trade-off between ambiguous range and range resolution. Pockels cells traditionally require high voltage and have a narrow field of view.[70] Like linear-mode sensors, these systems operate at low frame rates (~30 Hz). However, these systems have lower sensitivity than the APD arrays and require higher SNR and, hence, high-power lasers that operate at low PRF. Therefore, polarization-based imaging sensors may be disadvantageous in SWaP-constrained applications. The high SNR requirement will also limit the suitability of this technique for long-range applications. Using detectors with low detector noise levels or increasing the aperture size would improve the ability to carry out long-range measurements. Another option is to effectively concentrate the received energy onto fewer pixels by reducing the beam divergence to only illuminate the target while reducing the focal length to have the target occupy a reduced number of pixels. However, this will reduce the coverage area. At this writing the committee is aware of only two companies pursuing the 3-D ladar polarization imaging approach—one for the military market, and one for the short-range commercial market.

[70] K. Ayer and W. Martin, "Laser Imaging and Ranging System (LIMARS): A proof of concept experiment," Proc. SPIE 1633, 54-69, (1992).

Conclusion 2-2: Polarization imaging offers the potential for high-resolution 3-D imaging without the need for large avalanche photodiode arrays and is well suited for close-range imaging. However, this technique is less likely to be used for long-range applications, particularly those that require low size, weight, and power.

ACTIVE POLARIMETRY

Light is a transverse electromagnetic wave (i.e., the oscillations are perpendicular to the direction of propagation), similar in some respects to the waves that propagate down a taunt string when it is shaken. Consequently, the wave may oscillate horizontally, vertically, in a right- or left-handed spiral or in some combination of the four. Mathematically, the four components of the polarization are represented by a set of order quadruples known as the Stokes vector.[71]

The polarization of the light reflected from a surface contains information not contained in the intensity or even the spectral reflectivity. For example, measurements of polarization of ambient reflected or emitted light have shown the potential to differentiate man-made from naturally occurring materials, even when the two materials are perfectly spectrally matched.

One of the more common approaches to polarization measurement is to pass the received light through one or more polarization analyzers (e.g., a Polaroid filter) and measure the intensity of the filtered light that emerges. Multiple measurements with filters rotated with respect to each other can identify the degree of polarization (DOP) of the light received. The DOP[72] is strongly affected by the roughness of the surface. Since the roughness may be different on different surfaces having the same reflectivity, the DOP provides an additional degree of contrast when analyzing imagery. Surface roughness seems to be one of the parameters that influences the ratio of light reflected at different polarization states. People often wear polarized sunglasses to reduce glare from smooth surfaces. Passive DOP measurements have been limited by the uncertainty in the initial polarization state of the ambient illumination and the angle of incidence of the light. In addition, if more than one material is present within the field of view (FOV) of a single pixel (e.g., when leaves or camouflage are in front of a vehicle), the intensity and the DOP measured will be the result of the summation of all of the surfaces within the FOV and each surface will contribute in the proportion to which it is visible to the receiver. Since these proportions are not known, the "mixed pixel" effect adds to the uncertainty.

Active EO polarimetric measurements can eliminate this uncertainty by controlling direction of illumination and the initial state of its polarization. In addition, the bandwidth of the laser modulation allows high range resolution imaging. The signal generated by foliage or camouflage can be separated in range from the signal generated by whatever lies behind or in front of the object of interest. This minimizes the mixed pixel effect unless the two surfaces are separated by less than the range resolution of the ladar. Active polarimetry has not been used operationally, but is being investigated as part of emerging ladar sensors.

Conclusion 2-3: Active polarimetric imaging can provide much more powerful characterization of surface roughness than passive polarimetric imaging, and this can aid target detection and automated target recognition.

---

[71] Any nondegenerate representation of the polarization states can be used to represent the Stokes vector. Commonly, the four elements used to represent the Stokes vector are the horizontal linear polarization, vertical linear polarization, 45° linear polarization, and right hand circular polarization.

[72] The DOP measurements usually are based on the ratio of only a subset of the elements of the Stokes vector. DOP includes degree of linear polarization (DOLP), degree of circular polarization (DOCP) and ellipticity, which is the ratio of DOLP to DOCP.

UNDERWATER SENSING

An important application for underwater active EO sensing is detection of underwater mines. This technology represents some of the opportunities and challenges for active electro-active sensing. Four wavelengths are useful simultaneously for mine detection: 1,064 nm, 532 nm, one wavelength band in the red region (~650-670 nm), and one wavelength band in the near infrared (~750-810 nm). For simplicity the latter two will be referred to as ~650 nm and ~810 nm. These four wavelengths enable the application of multi-modal detection strategies to address the complex issues involved in mine detection, including target contrast with the background and false detections. Table 2-1 summarizes the laser wavelengths and their applications.

The 532-nm wavelength provides the primary imaging capability for both in- and above-water detection. The choice of 532 nm leverages prior Navy experience in mine detection, relatively good seawater transmission in the green spectral region, and the well-proven, frequency-doubled laser wavelength and high pulse energy that can be produced using a Nd:YAG laser source. The fundamental laser wavelength output from the Nd:YAG laser is 1,064 nm. Efficient frequency doubling to produce 532-nm laser radiation leaves approximately 40 to 50 percent of the 1,064-nm energy available. This can be used both for (1) interrogation of changes to the reflectivity of the ocean surface (due to waves and surface roughness) for signal normalization and (2) frequency conversion to produce two additional wavelengths that will aid in target discrimination. The rapidly absorbed 1,064-nm light provides only near-surface backscatter signals.

The final two wavelengths, ~650-670 nm and ~750-810 nm, are primarily used for target discrimination using both reflectance and fluorescence spectroscopy. A metric that is commonly used in remote sensing for determining the presence of vegetation (a common cause of false detections) is the Spectral Vegetation Index (SVI), which takes advantage of the "NIR vegetation rise" in reflectance. There is a distinct difference in reflectance between red and NIR light associated with the presence of chlorophyll. Chlorophyll absorbs strongly in the red (<700 nm) and reflects strongly in the near infrared

TABLE 2-1 Four Laser System Output Wavelengths and Their Application for Mine Detection

Laser Wavelength	Properties / Application
1,064 nm	• Very shallow water penetration depth (cm) • Measurement of ocean surface effects (waves/ripples) on reflectivity • Normalization/interpretation of signals at other wavelengths
532 nm	• Deep water penetration depth (tens of meters) • In-water imaging and target identification • Above-water imaging
~650 nm	• Moderate water penetration depth (m) • Excitation wavelength for chlorophyll fluorescence-based detection of vegetation • Low reflection wavelength for spectral vegetation index (SVI) • Application in very shallow water (VSW), surf zone (SZ), and beach detection
~810 nm	• Poor water penetration depth, but located at a minimum in the water NIR absorption, allowing some water penetration • High reflection wavelength for SVI-based detection of vegetation • Application in VSW, SZ, and beach detection

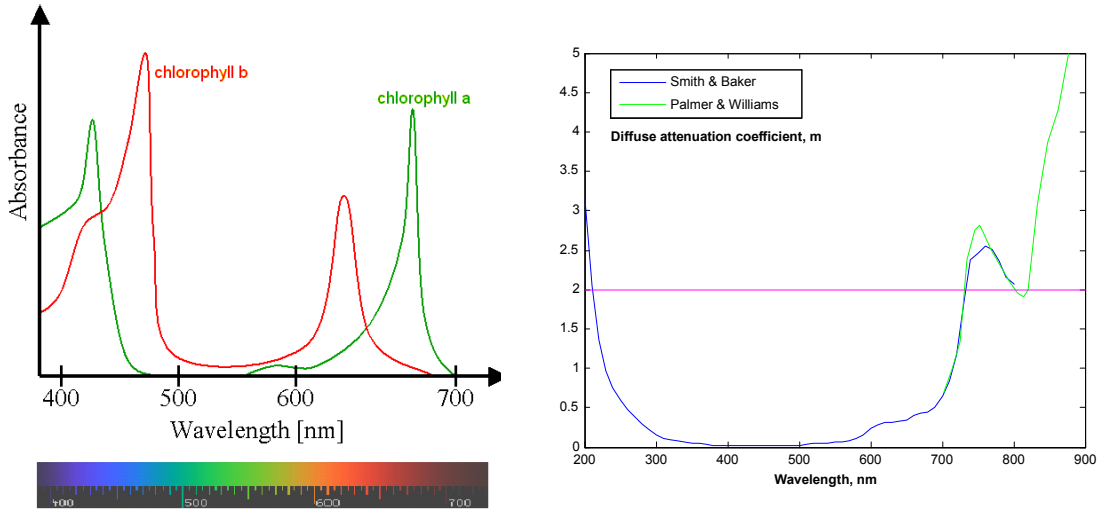

FIGURE 2-21 (a) Chlorophyll *a* and *b* absorption spectra; (b) Water transmission throughout the visible and infrared spectral regions. Note the sharp rise starting just above 700 nm and the dip near 810 nm. SOURCE: (a) By Chlorophyll_ab_spectra2.PNG: Aushulz derivative work: M0tty [CC-BY-SA-3.0 (http://creativecommons.org/ licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], via Wikimedia Commons. See http://upload.wikimedia.org/wikipedia/commons/6/68/Chlorophyll_ab_ spectra2.png.(b) Q-Peak, Inc.

(>700 nm). Most other common materials, such as sand, bare soil, exposed rock, or concrete, generally show a steady rise in reflectance (with no dramatic jumps) as wavelength increases from the visible to the NIR. Water, in contrast, has much higher reflectance in the red than in the NIR.[73,74]

In satellite remote-sensing applications, the measurement takes advantage of ambient sunlight, but for laser-based applications it can be done through selection of appropriate wavelengths to conduct the measurement of interest. A laser wavelength pair at ~650 and ~810 nm (red and NIR) can be used to measure the SVI based on either the spectral reflectance (SR) or the normalized difference vegetation index (NDVI). As shown in Figure 2-21a, absorption by both chlorophyll *a* and *b* are elevated at 650 nm, meaning the reflectance would be low, while at 810 nm, reflectance will be high. The ~650 nm wavelength will also excite chlorophyll fluorescence that can be detected in the vicinity of both 700 and 735 nm, the two chlorophyll emission peaks. This dual approach—vegetation detection by both fluorescence and SVI—has advantages for the application proposed here, for which the system should work in very shallow water (VSW), surf zone (SZ), and beach environments. Vegetation detection by SVI in water of any significant depth is degraded due to strong NIR absorption by water (Figure 2-21b). While 650 nm is not an ideal wavelength for penetration of the water column, it can penetrate far enough to reach shallow targets and excite fluorescence. The deeper red fluorescence emission then only has to travel a one-way path exiting the water, as opposed to reflectance signals, which require a two-way transit.

The selection of a laser wavelength near 810 nm as the NIR wavelength for SVI detection works as the high-reflectance band for vegetation, and it is also the NIR wavelength with the greatest penetration in the water column, enabling limited in-water use of SVI to complement fluorescence detection of

[73] See http://www.deepocean.net/deepocean/index.php?science07.php. Also see http://extension.usu.edu/nasa/ntm/on-target/near-infrared-tutorial.

[74] J. Weiqi, C. Fengmei, W. Xia, L. Guangrong, H. Youwei, Q. Huaichuan, and S. Fei, 2008, "Range-gated underwater laser imaging system based on intensified gate imaging technology," *Proc. of SPIE* 6621: 66210L.

chlorophyll-containing targets. Absorption by water goes up strongly in the infrared, but there is a local minimum at 810 nm, with steep rises on either side of this wavelength. The 810-nm wavelength is also well separated from the chlorophyll fluorescence wavelengths, enabling the use of broader detection bandpass filters if required.

Active Illumination and Detection from Above the Water Surface

Considerable effort has been spent investigating the possibility of sensing underwater from above the water, with many applications being relevant to ecological assessment. When incident from above the surface of the water, a fraction of incoming light is reflected away; the amount depends on the state of the water itself. If it is calm and smooth, theory determines that the fraction of light reflected depends on the polarization of the incident light. When vertically polarized light is incident at 57° to the surface, no light will be reflected. At angles more vertical than this, the reflectivity increases to a maximum of 5 percent. Angles more shallow are not very practical, as the reflectivity increases dramatically.

Incident illumination can easily be aimed at an underwater target when the water is calm. If the surface of the water is turbulent, however, with many waves, light will refract into the water at different time-varying angles, effectively spreading out the beam. With enough froth, most of the illumination may be reflected rather than refracted into the water. Light returning from under the water to a sensor above the water can also be modulated by the changing waves on the surface. This can reduce resolution, unless motion of the surface can be compensated for.

Light may be scattered or absorbed by solid particles. Most of the visible light spectrum is absorbed within 10 meters of the water's surface, and almost none penetrates below 150 meters of water depth, even when the water is very clear. Shallow water near a shore is typically more turbid (cloudy) due to particles and will show a decrease in light transmission. Large numbers of particles are brought in by river systems, by biological production of microorganisms, and by waves, tides, and other water movement that picks up debris on the ocean floor.

As the light enters the water and begins to scatter, the beam increases in size, as shown in Figure 2-22. It has been suggested that the nonlinear self-filamentation of intense laser light might keep the beam from spreading out, but it is unclear how fundamental losses in water would affect such behavior.[75] Experimental demonstration with filaments indicated that they remained at the same size in pure water for a distance of 4.5 m from the iris.

Range-Gating to Mitigate Backscatter

The impact of scattering is not only the reduction in signal reaching the target, but also the possibility of backscatter blinding the detector. Figure 2-23 shows an example of a backreflected signal as a function of time, measured by naval engineers in China.[76] It is clear that gating the detector to measure signal only in the expected range will substantially increase the SNR. When a laser pulse is emitted, a clock is started to trigger the detector array at the proper distance, as discussed in the section on 2-D active/gated imaging, above.

[75] K. Wang, B.D. Strycker, D.V. Voronine, P.K. Jha, M.O. Scully, R.E. Meyers, P. Hemmer, and A.V. Sokolov, 2012, "Remote sub-diffraction imaging with femtosecond laser filaments," *Optics Letters* 37 (8): 1343.

[76] G. Wei-Long, H. Hong-Weei, Z. Siao-Hui, and X. Xiang, 1999, "A new kind of underwater photoelectric imaging system," *Proc. of SPIE* 7382: 73824T.

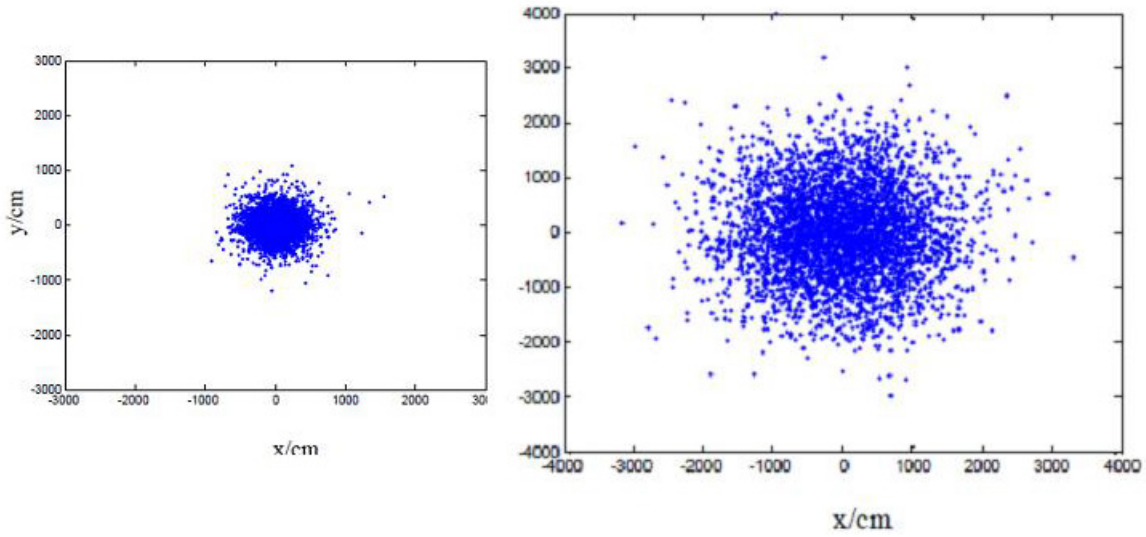

FIGURE 2-22 On the same scale, a light beam at 20 m below the surface of the ocean (left) and at 200 m depth (right). The assumed scattering attenuation coefficient was 0.4 m^{-1}. SOURCE: Y. Xiao-li, Y. Hong, X. Li-ming, and W. Fu, 2009, "Analysis of characteristics of blue-green laser propagation through ocean water," *Proc. of SPIE* 7382: 738212. © SPIE. Reprinted with permission.

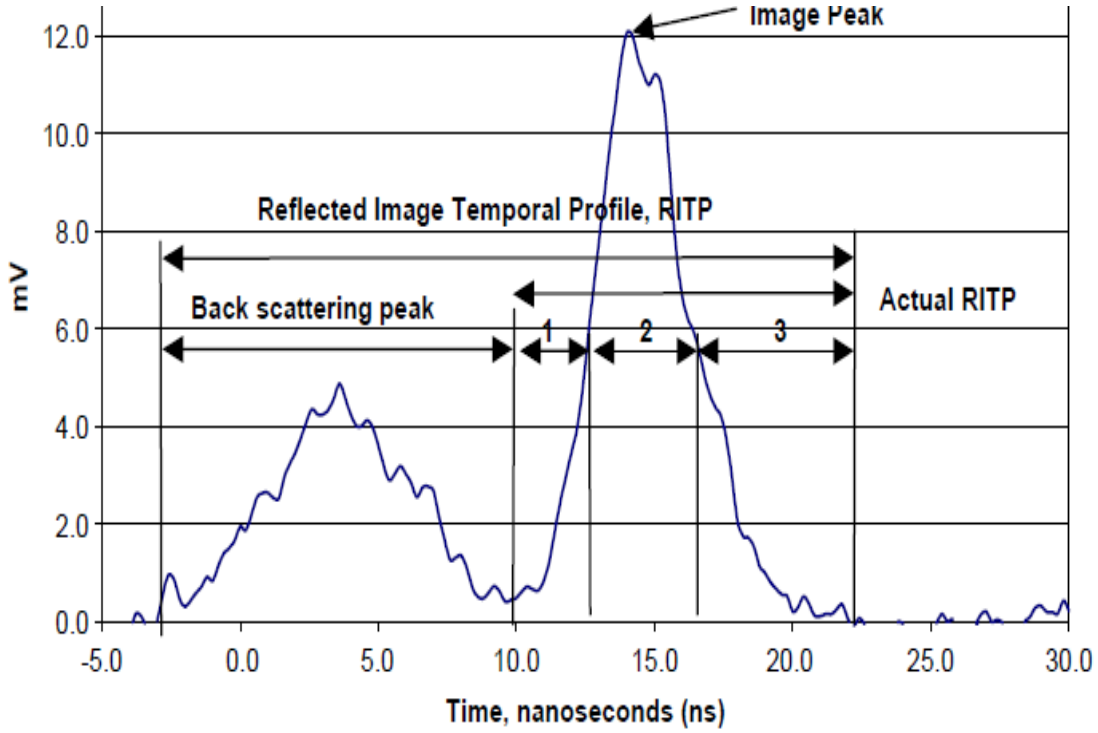

FIGURE 2-23 Reflected image temporal profile in the time domain, for clear water with attenuation coefficient α = 0.26 m^{-1} and absorption coefficient 0.04 m^{-1}. The front, middle, and tail of the actual signal are identified as 1, 2, and 3. SOURCE: G. Wei-Long, H. Hong-Weei, Z. Siao-Hui, and X. Xiang, 1999, "A new kind of underwater photoelectric imaging system," *Proc. of SPIE* 7382: 73824T. © SPIE. Reprinted with permission.

Practical distances are still limited to where water is very transparent. Jaffe[77] reports typical values for minimum attenuation of deep-ocean water, coastal water, and bay water as 0.05 m^{-1},0.2 m^{-1} and 0.33 m^{-1}, respectively. An acceptable attenuation for a completely reflected frequency-doubled, Q-switched Nd:YAG laser with 80 mJ per pulse energy, assuming suitably gated measurements, could be the detection of one photon per pulse. Because the input corresponds to 2.1×10^{17} photons per pulse, an acceptable round-trip attenuation would be $\exp(-40) = 4.2 \times 10^{-18}$. This corresponds to a one-way distance of $L = 20/\alpha$. Using Jaffe's suggestions for loss coefficients, visibility in deep ocean water should be 400 m; in coastal water the depth would be 100 m, and in bay water would be 60 m for a single-element detector. For imaging arrays, the distances would be reduced by the logarithm of the number of elements in the array.

In recent years, many research groups around the world including in Sweden, Denmark and Israel, have engaged in this technology. A special symposium of SPIE, held in 2009 to explore laser sensing and imaging, examined many of these technologies.[78]

A variety of methods have been tried to decrease the problems with scattering of the illumination signal. Rejecting backscatter by properly designing polarizing filters has been extensively investigated, but none of the methods seems to have increased the effectiveness of long-range underwater sensing. It appears that researchers in the United States are no longer pursuing advanced polarization techniques, at least in the open literature.

Coastal Sea Floor Imaging

Coastal waters provide another important application of active EO sensing, with aerial systems common since 1995. Recent systems under development are designed to be operated by helicopters or unmanned aerial vehicles (UAVs). Active optical range-imaging sensors collect three-dimensional coordinate data from object surfaces and can be useful along the sea coast, in littoral waters up to 50 meters deep, along the surf-line, and onto the beach. Performance of ocean-sensing ladar sensors depends on laser propagation effects in a scattering medium and the manner in which a particular system collects the ladar data. A few examples are discussed below.

The Scanning Hydrographic Operational Airborne Lidar Survey (SHOALS) system, developed for the U.S. Army Corps of Engineers, was fully operational by 1995. Employing scanning ladar to remotely collect accurate, high-density measurements of both bathymetry and topography in coastal regions, SHOALS had by 1999 completed 230 projects totaling 5000 km.[79] The SHOALS technology is marketed today by Optech.[80] It uses green laser pulses (second harmonic of the Nd:YAG laser) and measures time-of-flight of the reflected light. For bathymetry, the system sends out both the fundamental as well as the harmonic. Since the fundamental is absorbed by water within in a short distance, its sole signal is reflection is from the surface of the water. This means ocean depth is given by the time difference between these two different signals. SHOALS can thus measure both land heights and ocean littoral regions.

Coastal zone mapping and imaging lidar (CZMIL) is a multisensor airborne mapping system, also designed by Optech for the U.S. Army Corps of Engineers. The Optech CZML summary specification sheet cites the system produces simultaneous "high-resolution 3-D data and imagery of the beach and shallow water seafloor, including coastal topography andperforms particularly well in shallow, turbid waters... Its bathymetric ladar is integrated with a hyperspectral

[77] J.S. Jaffe, 1990, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean Eng.* 15: 101.

[78] International Symposium on Photoelectronic Detection and Imaging 2009: Laser Sensing and Imaging, edited by F. Amzajerdian, C.-Q. Gao, and T.-Y. Xie, *Proc. of SPIE* 7382.

[79] J.L. Irish and W.J. Lillycrop, 1999, "Scanning laser mapping of the coastal zone: the SHOALS system," *ISPRS Journal of Photogrammetry & Remote Sensing* 54: 123.

[80] See http://www.optech.ca/pdf/Brochures/shoals_shoals.pdf.

imaging system and digital metric camera."[81] OptechHydroFusion, an end-to-end software suite, handles all three sensors—from mission planning through to fused ladar and imagery data sets. For optimal object detection, particularly around surf, with its breaking waves and foam, a circular scan pattern provides two looks at any target. Tuell et al. describes the

> CZMIL scanner is based on a rotating Fresnel prism having a 20-cm clear aperture. The segmented detector holds the possibility of producing sub-meter, 3-D seafloor images in very shallow water… This bathymetric ladar is integrated with a commercial imaging spectrometer and digital metric camera. The data processing system employs new algorithms and software designed to automatically produce environmental image products by combining data from the three sensors within a data fusion paradigm.[82]

Northrop Grumman's active 3-D ocean sensor system is the Airborne Laser Mine Detection System (ALMDS), shown in Figure 1-11, whose initial roll-out was in 2007. The pulsed laser radar in this system, made by Arete Associates, scans a wide swath of ocean and is designed to locate and identify mines accurately in a single pass over a target area. "It transmits a fan-shaped beam of laser light to establish its swath width, and then relies on the forward motion of the helicopter to sweep the light over the water in a push broom manner."[83] "Four cameras are arranged to cover the same swath illuminated by the laser fan beam."[84] Time-resolution is obtained by streak-tube imaging light (STIL) detection. The streak tube uses a photocathode to detect the 1-D spatial information that comes from the slice of ground illuminated by each fan-shaped laser pulse and converts it to a line swath of electrons whose lateral density is proportional to the lateral variations in light intensity. The electron swath within the tube is accelerated and deflected by high voltage applied to plates. As the electron swath is rapidly swept vertically, so that when the electrons hit a phosphor screen, each portion forms a streak that provides time resolution along the third dimension. Thus, for any portion of the initial illumination slice, upward deviation along the streak represents depth into the ocean. Arete Associates have also developed a multiple-slit STIL system that uses several slits to provide additional capabilities, used for flash mode, 3-D polarimetry, and multispectral 3-D fluorescence imaging.[85]

Arete was also involved in an experimental program of the Office of Naval Research (ONR) called Anti-Invasion Mine Signature Measurement and Assessment for Remote Targeting (AIMSMART), which investigated 3-D flash ladar and 3-D polarimetric data on sea and land mines scattered throughout the surf and beach zones. Moran et al. describes the Rapid Overt Airborne Reconnaissance (ROAR) system:

> Under the Office of Naval Research's Organic Mine Countermeasures Future Naval Capabilities (OMCM FNC) program, Lite Cycles, Inc. is developing ROAR, a highly compact airborne active sensor system for mine and obstacle detection in very shallow water, through the surf-zone and onto the beach. The system uses a proprietary integrated scanner, detector, and telescope receiver architecture that tightly couples all receiver components and ladar electronics for small size while providing a large aperture. The system also includes an advanced compact multifunction laser transmitter, a compact 3-D camera, a scanner for wide area search, and temporally displaced

[81] Optech CZMIL. Available at http://www.optech.com/wp-content/uploads/specification_czmil.pdf. Acessed March 14, 2014.

[82] G. Tuell, K. Barbor, and J. Wozencraft, 2010, "Overview of the coastal zone mapping and imaging lidar (CZMIL): A new multisensor airborne mapping system for the U.S. Army Corps of Engineers," *Proc. SPIE* 7695, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI, 76950R.

[83] Northrop Grumman, Photo Release-Japan Maritime Self-Defense Force Orders Northrop Grumman's Airborne Laser Mine Detection Systems, http://www.irconnect.com/noc/press/pages/news_releases.html?d=244673. Accessed on March 14, 2014

[84] Ibid.

[85] See http://www.arete.com/arete_innovation/streak_tube_imaging_lidar_stil.aspx.

> multiple looks on the fly over the ocean surface for clutter reduction. Additionally, the laser provides time-multiplexed multi-color output to perform day or night multispectral imaging for beach surveillance. New processing algorithms for mine detection in the very challenging surf-zone clutter environment are under development for significant processing gains. The laser-illuminated, multi-spectral imaging system uses three time-stepped spectral bands with a single stabilized and scanned intensified CMOS camera. The 3-D imaging sensor utilizes a single stabilized and scanned 3-D focal plane array to provide a 3-D image of a volume of water with a single green laser.
>
> The primary sensor for use over the beach zone is a range-gated ICMOS camera that is used in conjunction with a multispectral laser illuminator that provides time-interlaced, three-color laser pulses. It generates the multispectral imagery for detecting individual mines, mine lines, and obstacles on the beach. The primary sensor for use over the surf zone is the 3-D camera used in conjunction with the laser operated in the green illumination mode. It provides the 3-D ladar data required to detect single bottom mines, volume mines, and obstacles, as well as mine and obstacle-lines. Both receivers operate simultaneously in all zones; adding the ICMOS camera for surf zone operation provides additional high-spatial-resolution, range-gated imagery to complement the lower transverse resolution imagery generated by the 3-D camera. This helps overcome the high level of clutter associated with the very challenging surf zone environment. The ICMOS camera can also be used in the shallow water zone to provide high-spatial-resolution imagery at specified depths and to assist in 3-D camera automatic gain control. The 3-D camera can also be used on the beach to provide topographic information for obstacle detection, and for detection of targets under camouflage and foliage.[86],,

Finally, the well-known military coastal program Coastal Battlefield Reconnaissance and Analysis (COBRA), is not classified as an active electro-optic system because this Northrop Grumman technology uses passive multi-spectral mine detection.

Sensing of Bubble Laser Scattering

A 2009 publication originating in China saw value in the laser sensing of ocean bubbles.[87]The authors described an application to wake-homing torpedoes that are designed to home in on ships by identifying their wake. The basic principle is to use the difference between laser scattering properties of the ship's wake, which contains a myriad of bubbles, and scattering from seawater in the nonwake region. Signal comparison can identify the ship's wake and determine the relative position of the ship. Underwater target detection is based on the scattering angle of the scattered light intensity distribution of bubbles in the water.

Conclusion 2-4: Active underwater sensing is likely to be limited to much less than 400 meters distance, reaching this value only in the deep ocean and with a single element detector.

When the technology is applicable, removal of backscatter from the detector by gating becomes important. The technology will undoubtedly have specific applications, but no practical solution to the transparency problem of water seems to be on the horizon.

[86] S.E. Moran, W.L. Austin, J.T. Murray, N. Roddier, R. Bridges, R. Vercillo, R. Stettner, D. Phillips, A. Bisbee, and N. Witherspoon, 2013, "Rapid overt airborne reconnaissance (ROAR^{TM}) for mines and obstacles in very shallow water, surf zone and beach," *Proc. of SPIE* 5089: 224.

[87] S. Liping, Z. Weijiang, R. Deming, Q. Yanchen, and H. Xiaoyong, 2009, "Processing methods in frequency domain for bubble laser scattering signals," *Proc. of SPIE* 7382: 73822X.

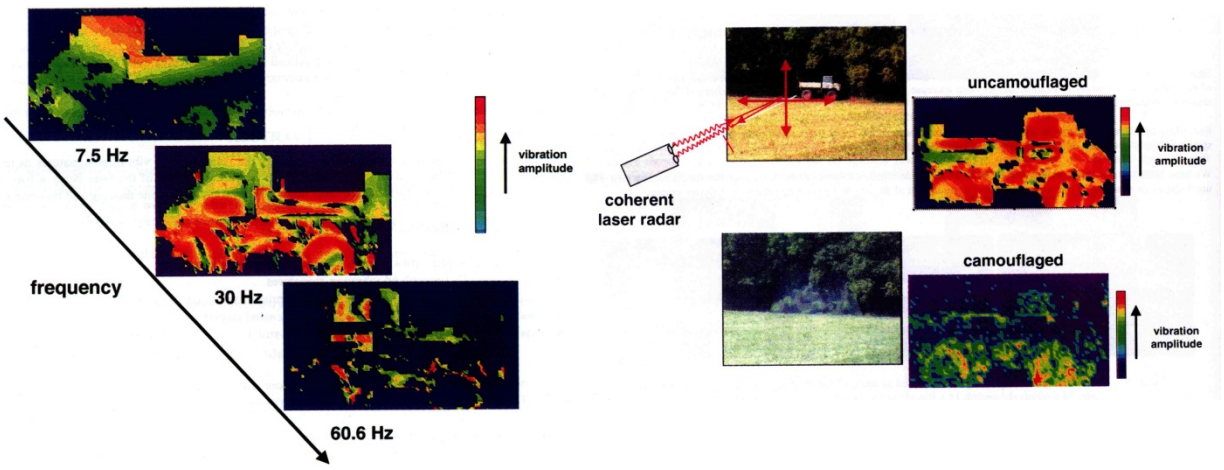

FIGURE 2-24 German vibrational imaging system. SOURCE: P. Lutzmann, R. Frank and R. Ebert, 2000, "Laser radar based vibration imaging of remote objects," In Laser Radar Technology and Applications V, *Proc. SPIE* 4035: 436.

VIBRATION SENSING

Listening to voices from far away, monitoring the health of machinery bearings, determining the operational status of electrical equipment, detecting underground tunneling activities, fingerprinting a particular operating vehicle—these are some of the applications enabled by remote laser vibrometry. The short wavelength of light, combined with sophisticated processing techniques, enables remote measurement of submicron displacement and sub-micron-per-second velocities. For many industrial applications, this capability is readily available commercially from companies such as Polytec-PI and Ometron. Military and intelligence applications have additional requirements, however, including the ability to operate at long range.

There are a variety of technological approaches to remote laser vibrometry, and this section will cover these alternative approaches and their relative merits for different applications, as well as limitations to performance.

Many laser vibrometer systems are point sensors, and spatial information is derived by scanning the laser spot across a surface. This results in a "data cube" similar to those in hyperspectral imaging, where at each *x, y* point in an image one associates a complete vibrational spectrum. This data cube can then be processed using algorithms similar to those used in hyperspectral data exploitation, such as creating an image corresponding to vibrational amplitude at a given frequency or even corresponding to a complete vibrational target signature. Laser vibrometers can also be designed and operated in a range-resolved mode, which can be useful in obscurant penetration applications.

Figure 2-24 presents results from a German vibrational imaging system. It shows frequency-resolved processing of the data cube as well as operation in a camouflage penetration scenario.

Beyond scanning spot systems, there is increasing attention being paid to multichannel systems for (a) improved area coverage and finding antinodes of vibrational modes and (b) averaging and common-mode rejection of noise and platform motion.

Commercial applications of laser vibrometers include monitoring of plant machinery, modal analysis of structures (from small parts to gas turbines to buildings), noise reduction, non-destructive analysis (finding manufacturing defects, cracks, and the like). These applications are usually short standoff, as the operator usually has ready access to the target of interest.

Law-enforcement applications include "laser microphones" that can monitor acoustic signals from windows and shades, objects in a room, or directly from a speaker's body. This even includes listening to both sides of a cellphone conversation by targeting the body of the phone. An important

feature is that unlike acoustic microphones that are strongly impacted by background noise, wind, and such, laser vibrometry targets the vibrating surface source directly, and is much less susceptible to such perturbations. This type of application tends to have a moderate standoff range (tens to hundreds of feet).

Military and intelligence applications include vehicle identification and tracking, plant monitoring (industrial intelligence, analysis of enigma facilities, etc.), equipment operational status (including air handlers, transformers, power lines) battle damage assessment, and interrogation of subsurface activity. These applications tend to have a long standoff range, up to many kilometers. One exception is laser vibrometry for buried landmine detection, in which a strong acoustic signal is broadcast from a ground vehicle and a laser vibrometer scans the ground surface for an anomalous response.

Vibrational sensitivity is quantified by both the noise-equivalent vibrational amplitude and the noise-equivalent vibrational velocity, and these quantities are related by the vibrational frequency of interest. (A higher frequency vibration has a higher surface velocity for a given vibrational amplitude). Because most laser vibration sensing technology is based on coherent detection, it can be made to be shot-noise limited in the optical detection by suitable choice of local oscillator power. A limitation (for cases where there is relative motion between the transceiver and the target) is usually laser speckle, which imposes a multiplicative noise on the signal. For long-range operation in the atmosphere, the limitation is atmospheric turbulence. Turbulence-induced "piston" fluctuations impose a phase noise on the optical signal that is indistinguishable from relative distance changes. Both of these limitations can be mitigated through multibeam and multiwavelength approaches that allow rejection of common-mode noise.

Standoff range is limited ultimately by laser power and receiver efficiency and is described by the usual ladar link budget for a shot-noise-limited coherent detection system. Another limitation that must be taken into account in designing system CONOPS is the requirement on the time a vibrating surface must be interrogated to discern the vibrational frequency (or to distinguish a specific vibrational signature from noise). In order to measure a vibrational frequency, one must typically dwell for several vibrational cycles. Bayesian processing techniques have been shown to achieve high frequency precision with very short data records.[88]

Of course one limitation for laser vibration sensing is that (unlike microphones) in most cases a direct line of sight/optical access to the vibrating surface is required, but acoustic and mechanical transmission of vibrations can be exploited and secondary vibrating surfaces can be used.

The most common commercial laser vibrometer technology is Doppler vibrometry (sometimes called micro-Doppler). In this approach the return signal is mixed with the local oscillator, and the beat frequency between the two electric fields creates a sinusoidally varying photocurrent. The motion of the target surface imposes a modulation of the frequency of that photocurrent. The signal is then processed to convert that FM-modulated signal into the vibrational spectrum. The most common approach is spectrogram processing (see Figure 2-25).

Alternatively, the surface displacement rather than the surface velocity can be measured directly, by measuring the phase of the return signal against that of the local oscillator in a Michelson or Mach-Zehnder interferometer. The local oscillator is often frequency shifted from the transmit wavelength by a modulator, and balanced or double-balanced heterodyne detection retrieves the optical phase. The evolution of this phase then measures the surface motion in time, and the data are processed to produce a vibrational spectrum, as above.

As noted above, there is a relationship between the surface displacement and velocity based on the vibrational frequency. Thus the micro-Doppler technique is best suited for high-frequency vibrations and the direct displacement technique is often best suited to lower-frequency vibrations.

[88] W.F. Buell, B.A. Shadwick, and R.W. Farley, 2000, "Bayesian spectrum analysis for laser vibrometry processing," Proceedings SPIE The International Society For Optical Engineering, 4035: 444.

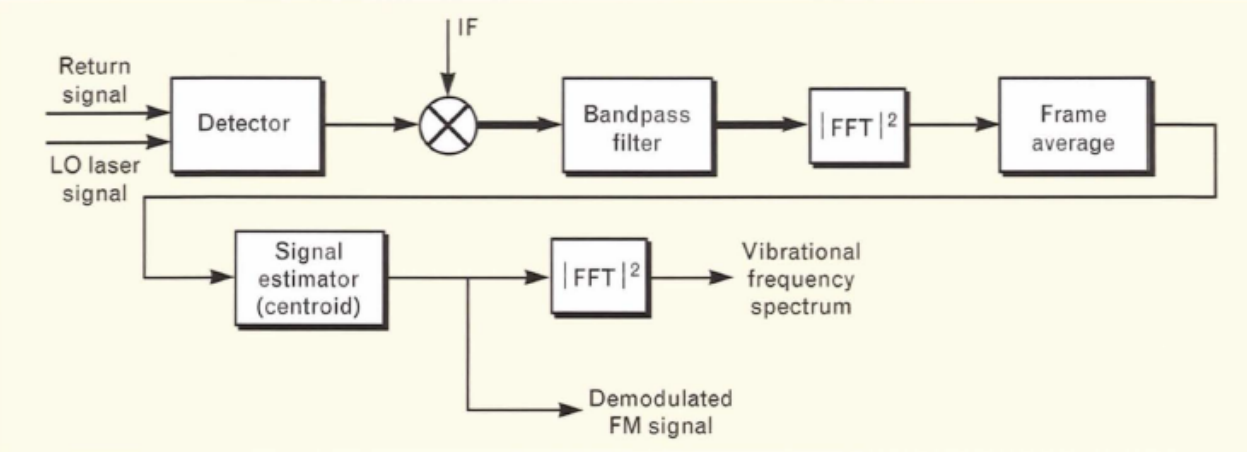

FIGURE 2-25 Block diagram of spectrogram processing. FFT, fast Fourier transform; LO, local oscillator. SOURCE: A.L. Kachelmyer and K.I. Schultz, 1995, "Laser vibration sensing," *MIT-LL Journal* 8 (1): 3.

With both techniques, there is a premium placed on the phase stability (coherence length) of the local oscillator. Both techniques can also be operated with a variety of transmit waveforms, including continuous wave (CW) tones, frequency-modulated continuous wave (FMCW) ramps, and pulse-pair or poly-pulse coherent detection.[89]

Another technique that has attracted interest in recent years is direct-detection laser vibrometry based on time-resolved imaging of laser speckle patterns. In this approach, no local oscillator is involved, and the detection is performed on a high-frame-rate camera. Surface motions induce small shifts in the speckle pattern at the receiver, and by sizing the spatial sampling at the focal plane appropriately, that small speckle motion can be tracked and processed to derive surface vibration information.[90]

Coherent detection of vibrometry signals can be accomplished with GM-APDs. As described in Chapter 3 of this report, heterodyne detection typically uses a LO strong enough to ensure that the measurement is limited only by the shot noise of the signal from the object. This heterodyne approach to reaching the shot noise limit (SNL) of photodetection allows the use of low tech p-doped-intrinsic-n-doped (p-i-n) photodiodes while enabling sensitive detection. Some measurement scenarios, however, require a large array of coherent detection channels and substantial LO power across the array, which poses challenges to thermal management of the focal plane. Larger arrays become progressively more unrealistic.

An alternative approach is to reach the SNL using photon-counting detectors, such as Geiger-mode avalanche photodiodes. These detectors of course cannot operate with strong LO powers—they would be immediately saturated and never see the signal light. The solution is to use a weak LO, on the order of the amplitude of the return signal. In this situation, the fields of the LO and signal interfere just as in the strong LO case, and this interference is manifest as a modulation of the photocount rate from the GM-APD. This modulation of the photocount rate occurs at the difference frequency between the LO and signal fields as expected, or is manifest spatially across the array if the signal and LO fields are not co-aligned (do not have parallel phase fronts).

[89] J. Totems, V. Jolivet, J.-P. Ovarlez, and N. Martin, 2010, "Advanced Laser Vibrometry In Pulsed Mode Using Poly-Pulse Waveforms And Time-Frequency Processing," OPTRO 2010, 4th International Symposium on Optronics in Defence and Security, Paris, France.

[90] Z. Zalevsky, Y. Beiderman, I. Margalit, S. Gingold, M. Teicher, V. Mico, and J. Garcia, 2009, "Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern," *Optics Express*, 17 (24): 21566.

There are important limitations to this technique that limit its applicability and the measurement CONOPS. Principal among these are (1) the dynamic range and visibility constraints and requirements on LO power control and (2) the timescale constraints that link photocount rate, beat frequency, and vibrational frequency.

- *LO power constraints*. In this technique it is necessary to keep the LO power as close to the signal power as possible. Naturally the LO power should not be strong enough to saturate the array, or even pose a blocking loss for the signal. An additional constraint, however, is imposed by beat frequency (fringe) visibility. In linear-mode strong-LO coherent detection, the detector is AC-coupled and gain is adjusted to optimize measurement of the beat signal. In weak-LO GM-APD coherent detection, AC coupling does not apply, and the fringe visibility (strength of the beat note) falls off as the LO and signal amplitudes become mismatched. Specifically, the fringe visibility is just the ratio of the field amplitudes.
- *Timescale considerations*. A 100 Hz signal will be detected as a 100 Hz modulation of the interference frequency of the count rate. Clearly the slowest part of the beat frequency needs to be larger than the vibrational frequency. Furthermore, the count rate must be high enough to resolve the beat note but small enough to (1) not saturate the detector and (2) not exceed the maximum readout rate.

As noted above, military and intelligence applications (and to a lesser extent, law enforcement) tend to be longer range than commercial and industrial applications. Thus, key indicators include the development of high-power phase-stable lasers and long-coherence-length LOs.

Further development of very low noise linear-mode APD arrays will reduce the power requirement on LOs, and enable higher-pixel-count "flash imaging laser vibrometers" that will compete with the GM-APD approach.

Future development of very-long-coherence-length LOs will enable extension of laser vibration sensing to even longer ranges. It is also likely that laser vibrometry will be incorporated in multisensor fusion suites to derive more complete information about a scene under observation.

LASER-INDUCED BREAKDOWN SPECTROSCOPY

Laser-induced breakdown spectroscopy (LIBS) uses intense pulses from a laser to induce optical breakdown in a material, typically a solid surface, and then detects the emission spectrum from the breakdown (Figure 2-26).[91,92,93] For nanosecond-duration pulses, the initial breakdown creates a plasma (ionized atoms and electrons) that then absorbs a good portion of the energy in the pulse. Shorter pulses in the femtosecond to picosecond range provide weaker signals because they just create surface breakdown and ejected, ionized atoms, and the pulses are over before the plasma absorbs more of the pulse energy. Recent advances in LIBS technology employ two pulses, spaced apart by one to tens of microseconds, where the second pulse acts to efficiently heat the plasma created by the first pulse.[94] The detection system then observes the emission created from the irradiation process, with time-gating of the detector to optimize the desired emission spectrum signal. If the pulse or pulses are powerful enough, the emission contains components from the electronic transitions of individual, ionized atomic constituents of the irradiated material. Through spectroscopic analysis of the emission wavelengths, typically in the 170-

[91] L.J. Radziemski and D.A. Cremers, 1989, "Spectrochemical analysis using laser plasma excitation," in *Laser-Induced Plasmas and Applications*, L.J. Radziemski and D.A. Cremers, eds., Marcel Dekker, New York, 295.

[92] L.J. Radziemski and D.A. Cremers, 2013, *Handbook of Laser-Induced Breakdown Spectroscopy,* John Wiley, New York, 2006.

[93] I. Schechter, A.W. Miziolek, and V. Palleschi, 2006, "Laser-induced breakdown spectroscopy (LIBS): fundamentals and applications," Cambridge University Press, Cambridge, U.K.

[94] R. Ahmed and M.A. Baig, 2009, "A comparative study of single and double pulse laser induced breakdown spectroscopy," *J. App.Phys.* 106: 033307.

1,100-nm range, one can determine the atomic constituents of the material from the characteristic wavelengths of their emission lines. In addition, from the strength of the signals at each line, it is possible to determine their relative concentration and, from this, the chemical makeup of the irradiated material.

LIBS provides a unique capability for in situ remote measurements of materials composition with nonionizing radiation. Although it does not strictly qualify as a non-destructive measurement, the amount of material sampled can be minimal. It is possible, by scanning the beam over a region, to get the spatial distribution of material composition with submillimeter resolution, another unique attribute.

Originally, LIBS was considered a short-range technology (< 1 m), developed for applications such as the detection of lead in paint. With more powerful, and low-divergence lasers, as well as a sufficiently large receiver telescope, it is possible to detect LIBS signals from targets 50-100 m away, thereby making the technology a form of remote sensing.[95] The typical laser source is a pulsed, 1-200-mJ, 5-20-ns-duration, Q-switched Nd:YAG laser at 1,064 nm, but other, more eye-safe wavelengths have been considered for long-range systems. These can be produced by third-harmonic generation of the Nd:YAG laser in the UV, or OPO-based shifting of the same laser to the 1,500-1,600-nm region. The typical receiver employs a charge-coupled device (CCD)-array detector, along with a dispersive element such as a diffraction grating, to image the spectrum onto the array. Typical limits of detection for each element are between 0.1 and 200 parts per million, depending on the sample and the element of interest. By employing advances in real-time spectral recognition software (chemometrics), it is possible with LIBS systems to provide a full analysis of the material composition on a 1-s timescale.[96] No significant range resolution is required for this technique, although if multiple range returns were in a given pixel's range resolution capability, this could have an advantage in clutter discrimination. A framing camera is sufficient for LIBs.

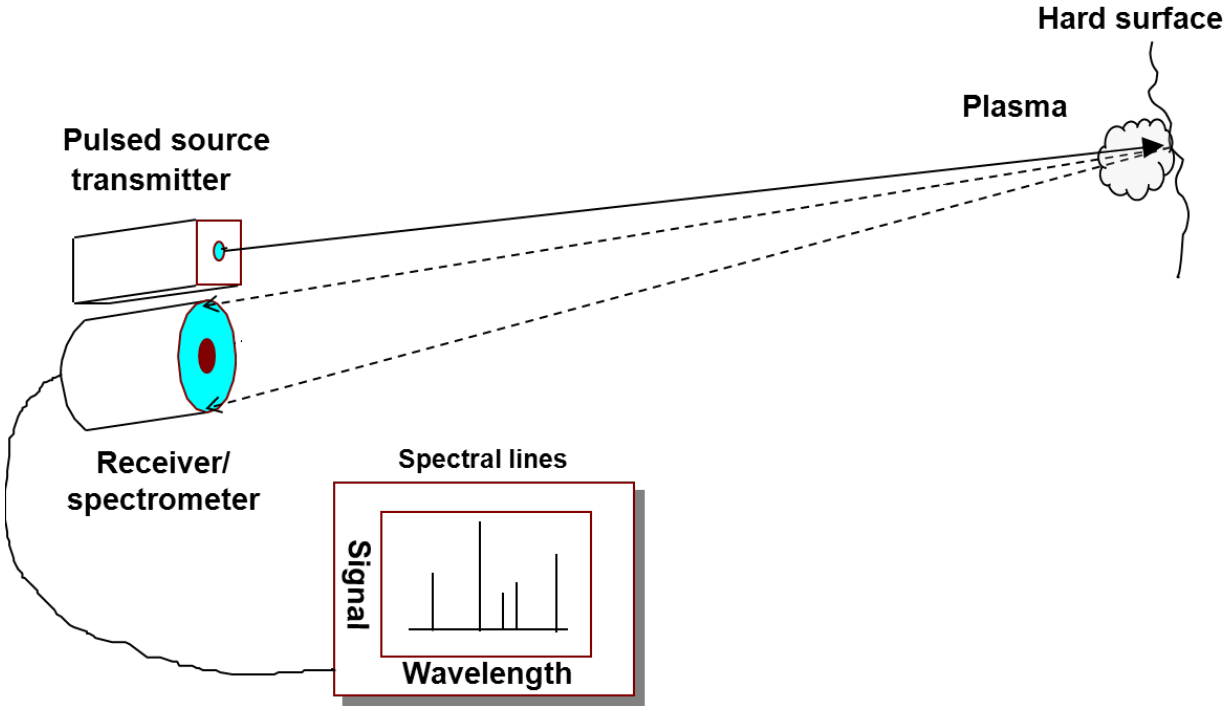

FIGURE 2-26 Generalized schematic diagram of a LIBS system.

[95] A. Miziolek, F. DeLucia, J. Gottfried, and C. Munson, 2010, "Progress in standoff libs detection and identification of residue materials," in *Lasers, Sources and Related Photonic Devices*, OSA Technical Digest Series (CD), Optical Society of America: LWD2.

[96] Ibid.

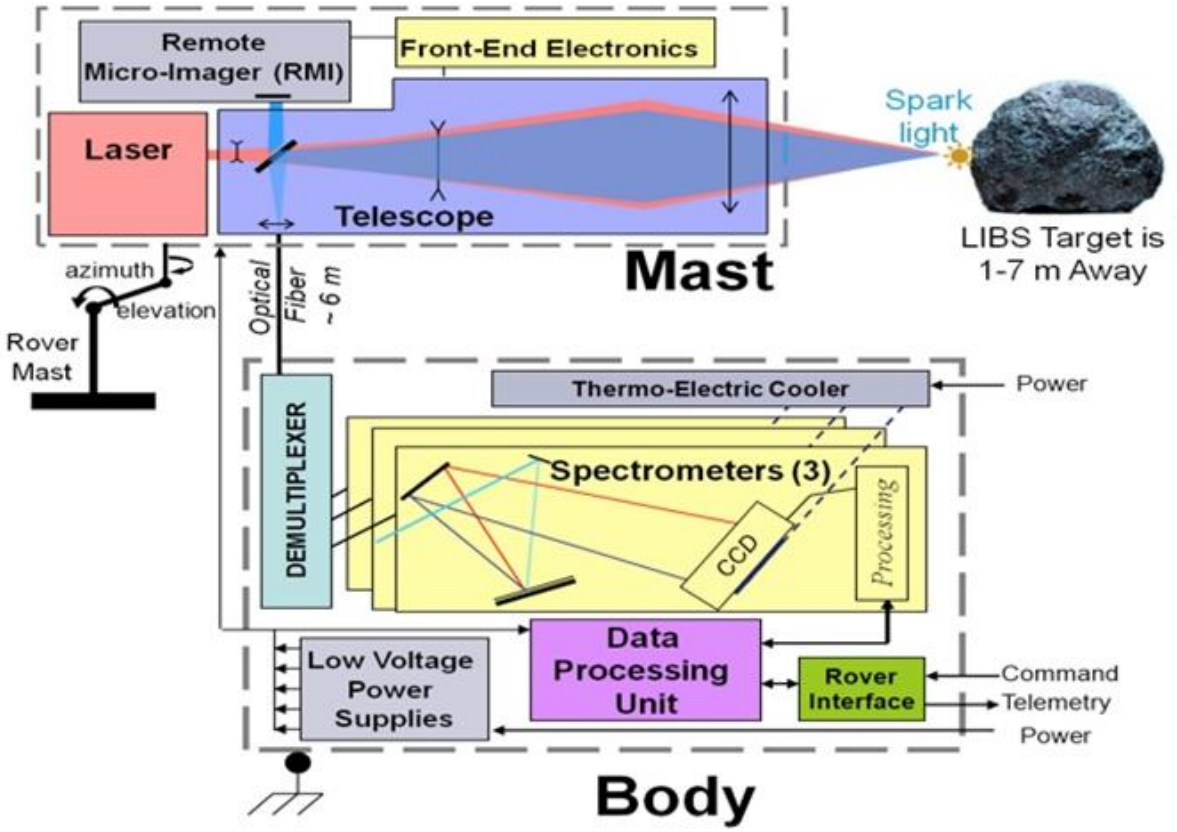

FIGURE 2-27 Schematic of LIBS instrument on Mars rover. SOURCE: See http://msl-scicorner.jpl.nasa.gov/Instruments/ChemCam/.

Figure 2-27 shows a detailed schematic of the LIBS instrument developed and now deployed on the Mars Rover to analyze martian soil. This short-range system employs a 14-mJ, 5-ns pulsewidth, diode-pumped, 1,070-nm, Nd:KGW laser that is mounted along with an 11-cm-diameter receiver telescope, on a rotating mast for beam direction to the desired location on martian soil. An optical fiber transports the received optical signal to a fixed unit in the Rover body, where the signal is optically split into three wavelength bands, each of which is analyzed by a grating/CCD spectrometer. An example of the spectral data product from the device appears in Figure 2-28, with the lines labeled according to the element detected.

The use of femtosecond pulses has shown, as expected, that less pulse energy is required to obtain ablation of material from a surface, and that at the low energies the signal contains much less background continuum emission and line signals from the atmospheric constituents (N and O) surrounding the material. However, the resultant LIBS signal is weaker, and when the pulse energy increases to correct this, the emission signals begin to resemble those from nanosecond pulses.[97] One area of evident advantage for femtosecond pulses is in the LIBS detection of biological entities, where it is important to

[97] F.C. De Lucia, J.L. Gottfried, and A.W. Miziolek, 2009, "Evaluation of femtosecond laser-induced breakdown spectroscopy for explosive residue detection," *Opt. Express* 17: 419.

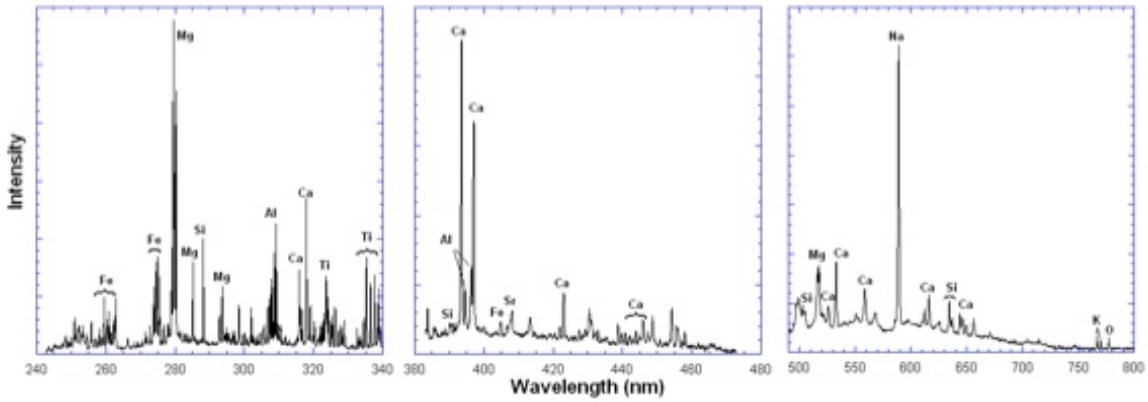

FIGURE 2-28 Spectral data from Mars LIBS instrument. SOURCE: See http://msl-scicorner.jpl.nasa.gov/images/ChemCam_spectra.jpg, http://msl-scicorner.jpl.nasa.gov/Instruments/ChemCam/.

detect the presence of the cyanide group (CN) inherent in the sample rather than the CN created by interactions in the air plasma surrounding the sample.[98] The other advantage of femtosecond-pulse LIBS for long-range standoff detection is discussed below.

LIBS technology has sufficiently advanced beyond basic scientific studies to the commercial market, where it finds uses worldwide in manufacturing process control; one example is monitoring the composition of molten metals. Military applications are primarily defensive in nature; they include remote detection of the presence of chemical, biological, nuclear, or explosive material on surfaces. Future advances will include (1) the use of higher-beam-quality lasers to increase the detection range;(2) the application of femtosecond-generated filaments (discussed in Chapter 3), also for extending the range; and (3) more sophisticated double-pulse, multilaser sources that enhance the emission intensity from the ablated materials and improve the minimum detectable constituent level. Although initially developed for short-range (1 m) sensing, advances in the development of high-beam-quality, nanosecond-pulse lasers have extended ranges to 100 m. Advances in array detectors and chemometrics software provide opportunities to develop real-time analysis of LIBS data to identify the material being sensed.

Conclusion 2-5: Laser-induced breakdown spectroscopy provides a unique sensing capability to determine the elemental makeup of the surface of a solid or liquid material.

AEROSOL SENSING

An aerosol is a suspension of liquids or solids in a gas. In Earth's atmosphere, aerosols are always present at normal altitudes and cause visible laser beams and searchlights to be "seen" as they propagate, due to the scattering of light from the liquid or solid component of the aerosol. Figure 2-29 shows the diameters of a variety of common atmospheric aerosol particles.

In the approximation that the aerosol is a simple sphere with a certain refractive index, one can calculate the fraction of electromagnetic radiation that is reflected directly backwards ("backscattered") from a beam incident on aerosols of all the same size. The fraction for particles with diameters much smaller than the wavelength of the radiation is proportional to the sixth power of the diameter divided by the fourth power of the wavelength (Rayleigh scattering), while for particles with diameters much larger than the wavelength, the fraction scattered back is simply proportional to the area of the particle. For

[98] M. Baudelet, L. Guyon, J. Yu, J-P. Wolf, T. Amodeo, E. Fréjafon, and P. Laloi, 2006, "Spectral signature of native CN bonds for bacterium detection and identification using femtosecond laser-induced breakdown spectroscopy," *Appl. Phys. Lett.* 88: 063901.

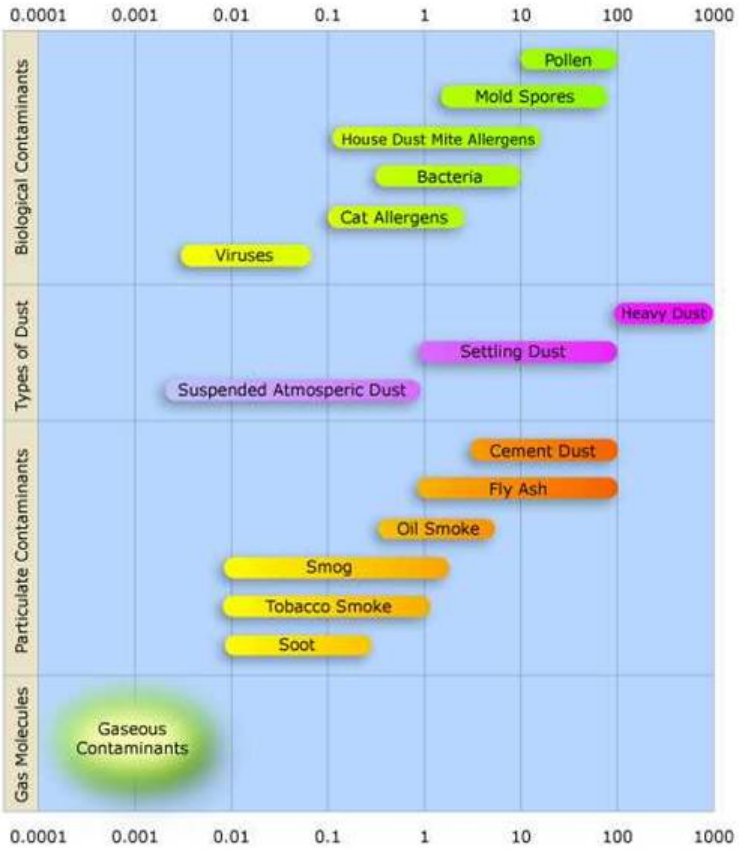

FIGURE 2-29 Distribution of aerosol diameters (μm) in the atmosphere. Water clouds and fogs fall in the 1-100 μm range. SOURCE: By Jisaac9 (Own work) [CC-BY-3.0 (http://creativecommons.org/licenses/by/3.0)], via Wikimedia Commons. See https://en.wikipedia.org/wiki/File:Airborne-particulate-size-chart.jpg.

particles with diameters comparable to the sensing wavelength, the full solution to the scattering problem, based on Maxwell's equations, must be applied. The solutions, first credited to Gustav Mie in 1908, are referred to as Mie scattering.

The scattering probabilities show, as a function of increasing particle size, an oscillation about the eventual scattering value of the fraction given by the area of the particle. Given that common radar systems work in the 1- to 10-cm wavelength range, it is evident that a large percentage of aerosols fall in the Rayleigh approximation and are weakly scattering, making radar, for the most part, useless for detecting atmospheric aerosols. Given lasers operating at a wavelength on the order of 1 μm—near the scattering peak of a typical aerosol—laser-based sensors can generate a much larger backscattered signal and have provided a unique tool for the detection and study of atmospheric aerosols.

Conventional Aerosol Lidar

The simplest aerosol lidar sensor (Figure 2-30) employs an energetic, short-pulse (nanosecond-range) laser, and a high-speed detector to provide a time-dependent return signal that provides a measure of the spatial distribution of the aerosol concentration. For a monostatic lidar, the signal is essentially proportional to the backscattered fraction of laser light.

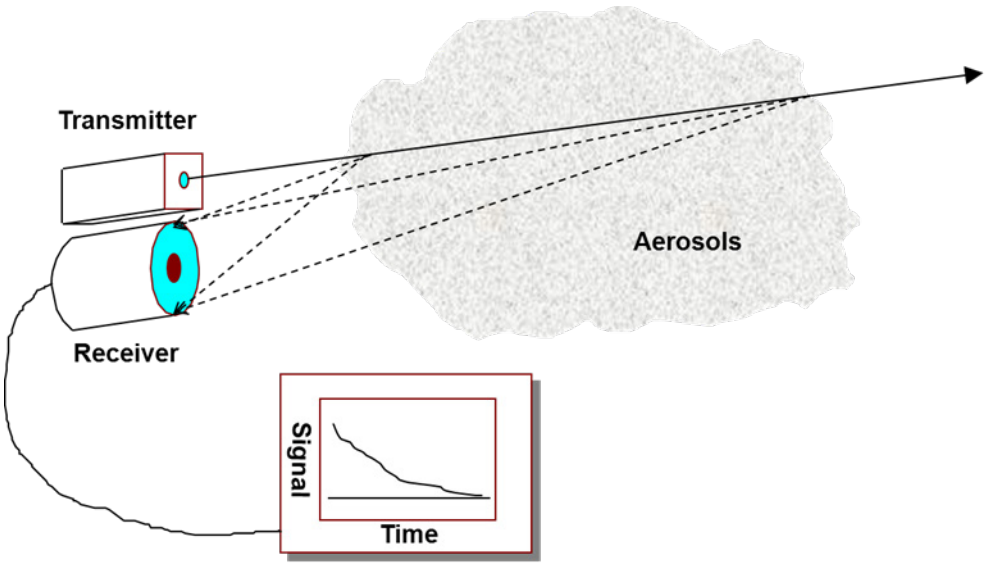

FIGURE 2-30 Schematic of an aerosol lidar.

The lidar equation (see Box 1-3) shows that the expected backscattered signal falls as the inverse square of the distance to the aerosols. The time-dependent detected signal can be corrected to account for this. The return signal as a function of time is determined not only by the intensity of the light returned from a given spatial region, but also by the light lost (from aerosol and other scattering processes, as well as atmospheric absorption) in going to that region and returning, which depends on the distance to the region. If the light lost was due only to the aerosols, and one knew that all the aerosols sensed were of the same diameter, shape and refractive index, it would be possible to accurately determine the aerosol concentrations with a careful calibration of the lidar system parameters and use of the Klett inversion method[99] for analyzing the time-dependent signal. Unfortunately, aerosols are rarely this uniform and instead present a distribution of sizes and shapes. Obtaining a reliable measure of the aerosol concentration and makeup from a single-wavelength laser is a challenge. It has been remarked that aerosol lidar provides an accurate measurement, but of what is uncertain.

The aerosol return signal can vary greatly, and for long ranges and/or small aerosol concentrations, lidar systems employing high pulse energies are desirable, to obtain an adequate SNR, especially in the presence of background skylight. For small-diameter aerosols, shorter wavelengths are desirable up to the point at which atmospheric transmission or the scattering loss caused by larger aerosols in the same path becomes a limit. In addition, the same (Rayleigh) scattering of light by molecules in the atmosphere that gives the sky its blue adds to the return signal at short wavelengths. The Klett inversion method was subsequently modified by Fernald[100] to include the case where the return signal included Rayleigh-scattered light. More sophisticated single-wavelength aerosol lidar systems employ a single-frequency laser, along with narrow-line filters (etalons) that can discriminate between the relative broad linewidth of the molecular-scattered return signal and the much narrower return linewidth from aerosols.[101] This is discussed in greater detail in the "Wind Sensing" section below. An alternative

[99] J.D. Klett, 1981, "Stable analytical inversion solution for processing lidar returns," *Appl. Opt.* 20: 211.
[100] F.G. Fernald, 1984, "Analysis of atmospheric lidar observations: some comments," *Appl. Opt.*, 23: 652.
[101] S.T. Shipley, D.H. Tracy, E.W. Eloranta, J.T. Trauger, J.T. Sroga, F.L. Roesler and J.A. Weinman, 1983, "A high spectral resolution lidar to measure optical scattering properties of atmospheric aerosols, Part I: Instrumentation and theory," *Appl. Opt.* 23: 3716.

approach employs detection of the Raman-scattered signal from atmospheric molecules, which is discussed in the "Raman Sensing" section below.

Given the limitations in measurement, conventional, single-wavelength aerosol lidars can provide accurate information on the spatial distributions of aerosols but not on their precise makeup. Figure 2-31 shows a photograph of the aerosol lidar system originally developed at Los Alamos National Laboratory[102] and subsequently transferred to UCLA as part of the Center for Lidar Environmental and Atmospheric Research (CLEAR). The system was developed to map the spatial characteristics of pollution-generated aerosols in urban environments. It employed a flashlamp-pumped Nd:YAG laser that used an optical parametric oscillator (see Chapter 4) to shift the laser wavelength to an eye-safe 1,570 nm, allowing operation in populated environments. The laser and transmit telescope are mounted to the left of the larger receiver telescope; the laser chiller and power supply are in rack mounts on the left and right cabinets below the telescope; and the signal processing/spatial location electronics are in the top right enclosure, with the system data display. Figure 2-32 presents a data product from this system, showing several scans over the Los Angeles basin; color scales indicate the relative "pollution" level. The distinct elevation boundaries between high and low levels are evident, along with a concentration of pollution in Beverly Hills associated with a cluster of automobile traffic.

China has a severe air pollution problem,[103] and may develop lidar technology to monitor aerosol pollutant concentrations. This could provide a technology springboard to other applications.

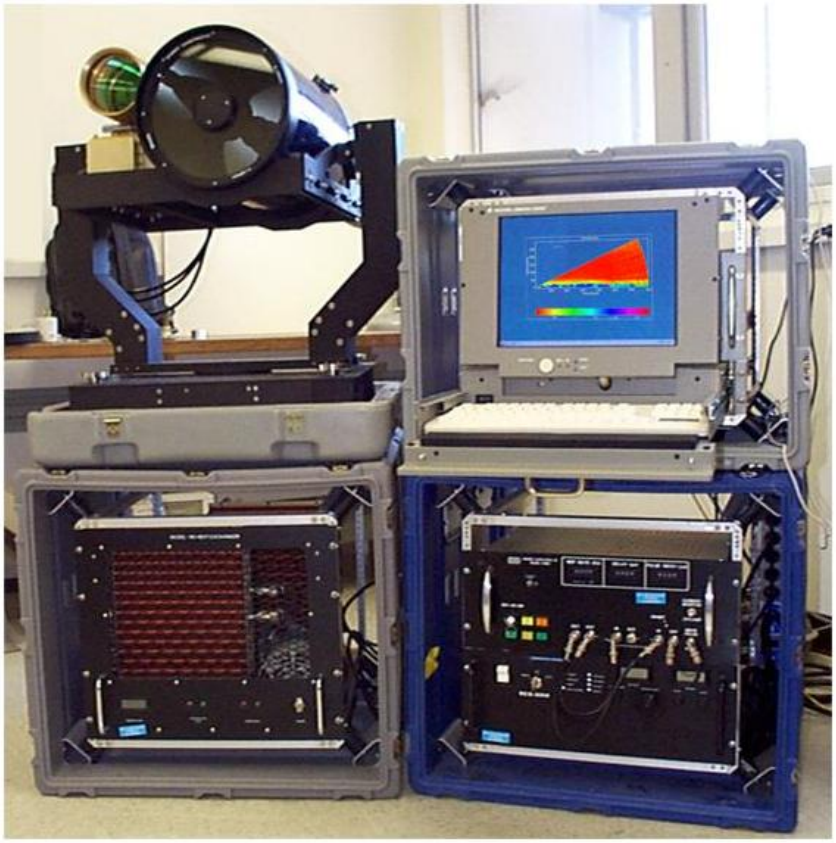

FIGURE 2-31 CLEAR aerosol lidar system developed for urban pollution measurements. SOURCE: Courtesy of Q-Peak, Inc.

[102] S. Barr, W. Buttler, D.A. Clark. W.B. Cottingame, and W.E. Eichinger, 1995, "Lidar-observed wind patterns in the Mexico - New Mexico - Texas border region," Los Alamos National Laboratory Report.

[103] See for example http://www.nytimes.com/2013/10/25/world/asia/smoggy-days-in-harbin-prompt-quick-reaction.html.

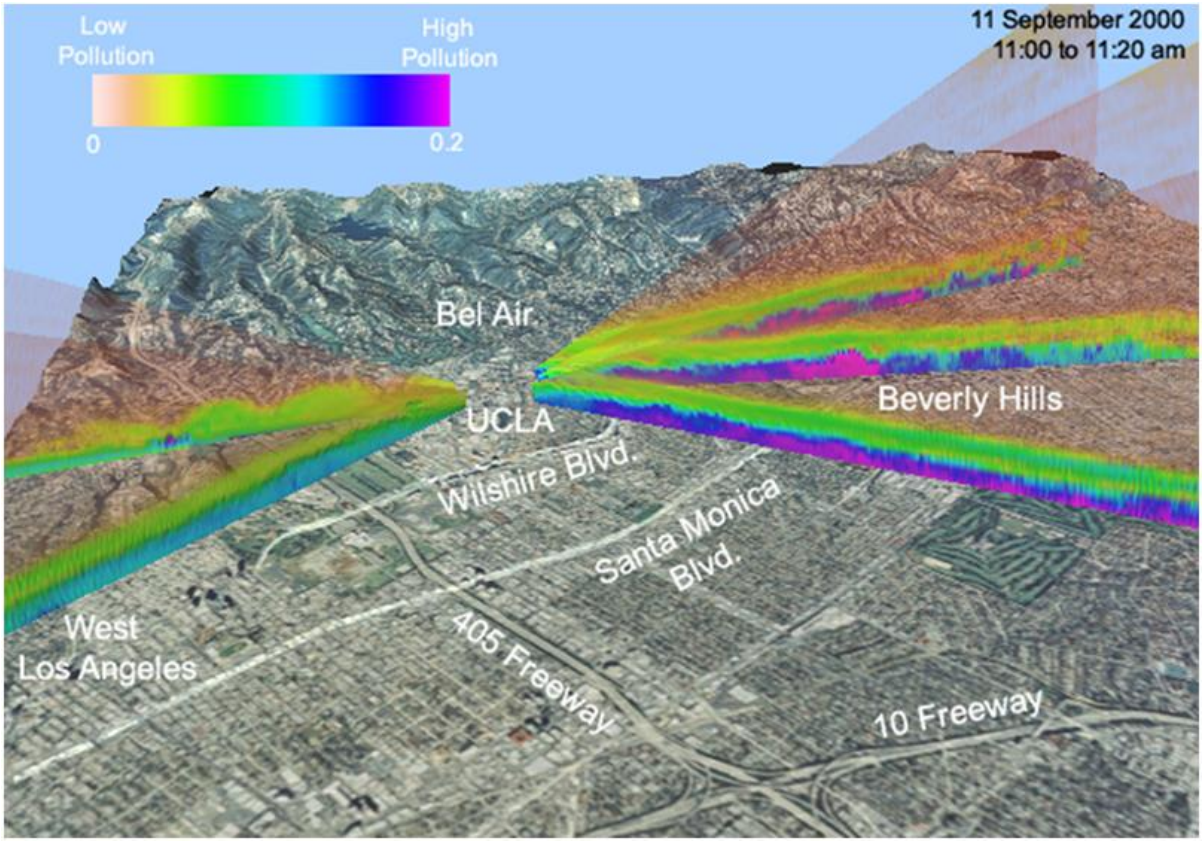

FIGURE 2-32 Data product from CLEAR lidar system, covering the Los Angeles basin. SOURCE: Courtesy of Q-Peak, Inc.

It is evident that a series of measurements at different times based on aerosol scattering can provide data on aerosol transport in a selected area, and hence some indication of the local winds as well as vertical convection currents. Besides providing useful meteorological information, the aerosol data could track the dispersion of toxic fumes from a fire or industrial accident, since they likely have associated aerosols. In addition, in the case of an intentional chemical or biological agent attack in an urban area, an aerosol lidar could provide information on the immediate spread of the agents, and where they might go in the future.

For defense applications, a similar system was developed in the late 1990s by Schwartz Electro-Optics for the U.S. Army Long-Range Biological Standoff Detection System (LR-BSDS).[104] The device was intended to be mounted in a UH-60A (Blackhawk) helicopter platform, for "pop-up" sensing of possible bio-aerosol clouds that would be present in the case of a biological warfare attack. The operational concept employed is sketched in Figure 2-33 and a photograph of the system mounted in the helicopter appears in Figure 2-34.

The LR-BSDS system was specified to detect aerosol clouds up to 50 km away. The transmitter employed a diode-laser-pumped, 1 J/pulse, 20-ns pulsewidth, 100 Hz Nd:YAG laser that was converted to 0.33 J of pulse energy at 1,535 nm by a potassium titanyl arsenate (KTA)-crystal-based optical parametric oscillator (OPO) (see Chapter 4). To be eye-safe according to ANSI standards at the transmit aperture,

[104] L.A. Condatore, Jr., R.B. Guthrie, B.J. Bradshaw, K.E. Logan, L.S. Lingvay, T.H. Smith, T.S. Kaffenberger, B.W. Jezek, V.J. Cannaliato, W.J. Ginley, and W.S. Hungatem, "U.S. Army Soldier and Biological Chemical Command counterproliferation long-range biological standoff detection system (CP LR-BSDS)," *Proc. SPIE* 3707, Laser Radar Technology and Applications IV, 188 (May 28, 1999).

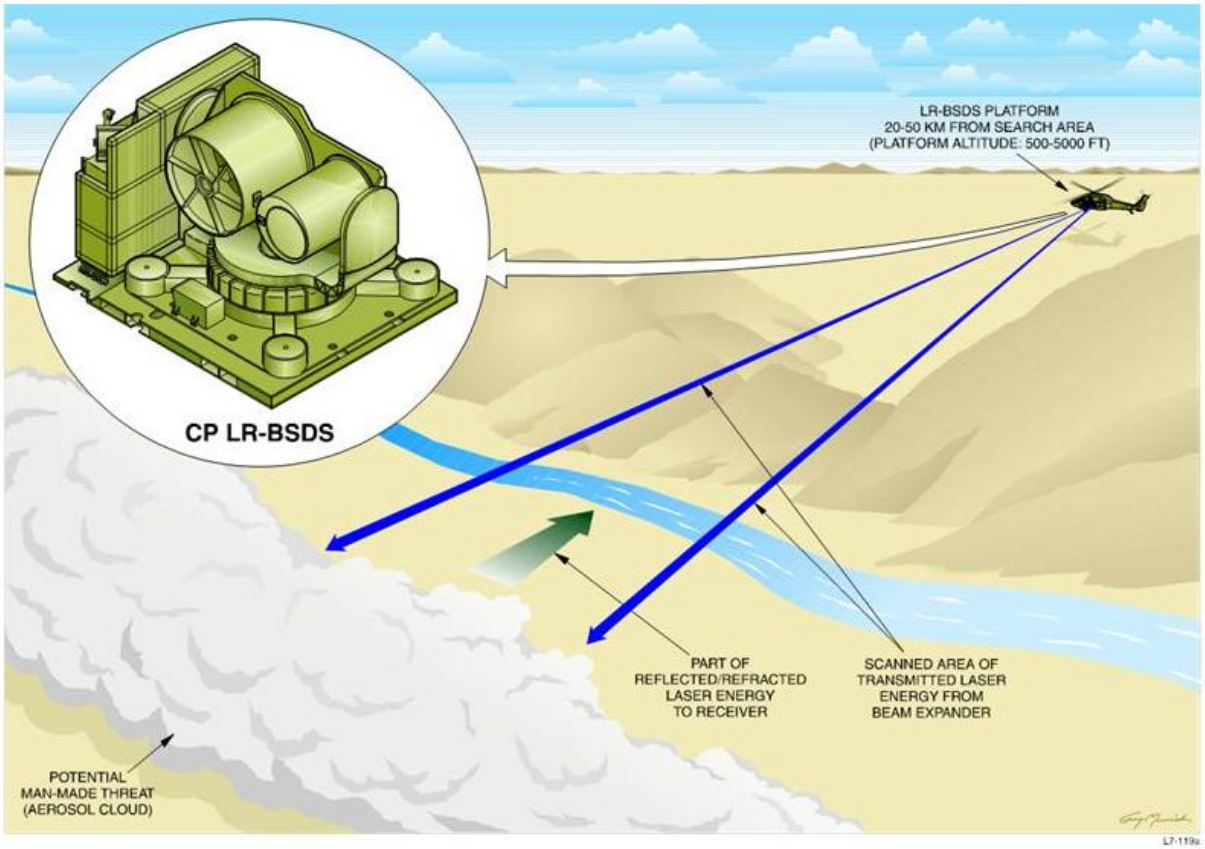

FIGURE 2-33 Drawing of operational concept for LR-BSDS system. SOURCE: Courtesy of Q-Peak, Inc.

FIGURE 2-34 The LR-BSDS system mounted in a UH-60 helicopter. SOURCE: Courtesy of Q-Peak, Inc.

the OPO output was coupled to a 7-in. aperture transmit telescope. The receiver employed a 24-in. aperture telescope and a transferred-electron-intensified photodiode (TE-IPD) was the detector. The system did demonstrate detection of aerosols to a distance of at least 45 km, but a change in U.S. Army tactics regarding biological attacks halted further development for operational uses.

As noted above, conventional, single-wavelength aerosol detectors provide ambiguous data on the nature of the aerosol clouds. As an example, the LR-BSDS system could not distinguish a biological agent from a dust cloud, and required additional data, such as the shape of the cloud or other sensor information. The next section discusses more advanced aerosol sensors that can provide less ambiguous data.

In summary, aerosol mapping with lidar has found widespread scientific applications in helping to understand fundamental atmospheric properties such as solar absorption. It can provide information on the three-dimensional nature of pollution, can identify sources of pollution, and can provide micro-scale mapping of pollution transport in urban environments. It has also been developed as one means of detecting and mapping the dispersion of a biological weapons attack. However, the lack of a widespread need beyond science, pollution monitoring, and defense has limited the development of low-cost, high-production-rate aerosol sensors.

Differential Scatter Lidar—Polarization and Multi-Wavelength

More complete data on the composition of the aerosols producing a return signal can be obtained through a combination of multiple-wavelength transmitters and/or polarized sources and polarization-sensitive detection systems. The comparison of the signals from the different channels to obtain better data is sometimes called Differential Scatter (DISC) lidar.

There are several motivations for obtaining improved aerosol data. From an Earth-science standpoint, atmospheric aerosols feature significantly in our planet's energy balance of incoming solar radiation and energy radiated back to space. Better data on the makeup of the aerosols (including those in water clouds) is necessary for effective atmospheric modeling. As noted above, for biological-agent defense, being able to determine if a cloud is indeed an active agent is crucial to a warning and tracking system.

For aerosols that are very small compared to the laser wavelength, Mie scattering calculations show that there is a very strong variation in backscattered and attenuated energy with wavelength; for large aerosols on the other hand, there is essentially none, and in between, there is a complex relationship. An aerosol lidar employing multiple widely spaced wavelengths provides multiple channels of return data, and the differences in the channels (hence DISC) can be analyzed to better determine the aerosols generating the signal. Figure 2-35 shows atmospheric lidar data obtained with a laser system designed for sensing tropospheric aerosols; the system employs a Nd:YAG laser at 1,064 nm and its second and third harmonics at 532 and 355 nm, along with a Ti:sapphire laser at 750 nm.[105] As expected, the strength of the scattering, expressed as an extinction coefficient (loss per km), increases with decreasing wavelength. Analysis of the data showed a broad distribution of aerosol diameters over the range 0.001-1 μm, peaking in the 0.01-0.1-μm range and essentially falling to zero beyond 1 μm.

Aerosols are not always perfectly spherical, and when they are not, they can change the polarization state of the backscattered signal. Lidar systems employing a polarized transmitter and a multichannel, polarization-sensing receiver, can also sense whether or not the aerosols are nonspherical.

One example of a system that includes both wavelength and polarization sensing is the NASA space-based Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) system that was launched into

[105] M. Yabuki, H. Kinjo, H. Kuze and N. Takeuchi, 2001, "Derivation of aerosol optical properties from four-wavelength lidar observations," *Proc. SPIE* 4153, Lidar Remote Sensing for Industry and Environment Monitoring, 132, February.

a sun-synchronous, 705-km altitude orbit in 2006.[106] The system includes a diode-pumped Nd:YAG laser transmitter that generates two wavelengths, 1,064 and 532 nm, with 110 mJ/pulse at each wavelength, at a 20-Hz rate. The receiver, employing a 1-m telescope, has three channels, one unpolarized at 1,064 nm with an avalanche photodiode detector and two that are orthogonally polarized at 532 nm with photomultiplier detectors. An example of a CALIOP data product is presented in Figure 2-36.

The inclusion of longer wavelengths would be appropriate for characterization of larger aerosol sizes, such as man-made particles (see Figure 2-29) resulting from industrial production or activities around urban areas. Past work on multiple-wavelength aerosol lidar has included wavelengths in the 1-10-μm range, but systems have suffered from reduced sensitivity due to the lack of both photomultiplier tubes and powerful sources, with the exception of the wavelengths between nine and eleven μm generated by pulsed CO_2 lasers. Recent developments in low-noise, high-gain semiconductor detectors as well as high-energy solid state and OPO-based sources make high-sensitivity operation at long wavelengths possible.

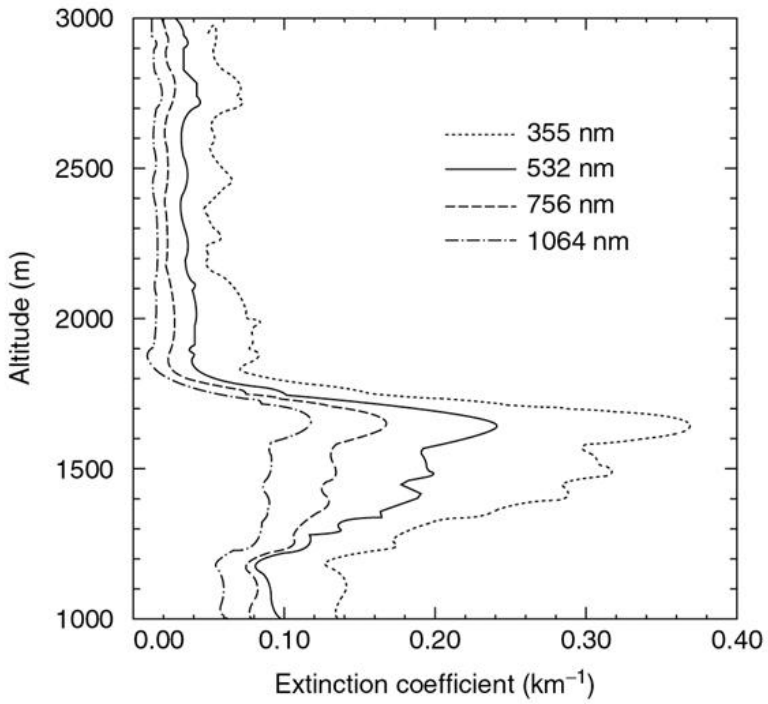

FIGURE 2-35 Atmospheric aerosol data generated by four-wavelength, vertical-pointing lidar system. Some of the data, especially at high altitudes and at short wavelengths, are from Rayleigh scattering from air, rather than aerosols. SOURCE: M. Yabuki, H. Kinjo, H. Kuze and N. Takeuchi, 2001, "Derivation of aerosol optical properties from four-wavelength lidar observations," Lidar Remote Sensing for Industry and Environment Monitoring, *Proc. SPIE* 4153: 132.

[106] W.H. Hunt, D. M. Winker, M. A. Vaughan, K. A. Powell, P. L. Lucker and C. Weimer, 2009, "CALIPSO Lidar Description and Performance Assessment," *J. Atmos. Oceanic Technol.*, 26: 1214.

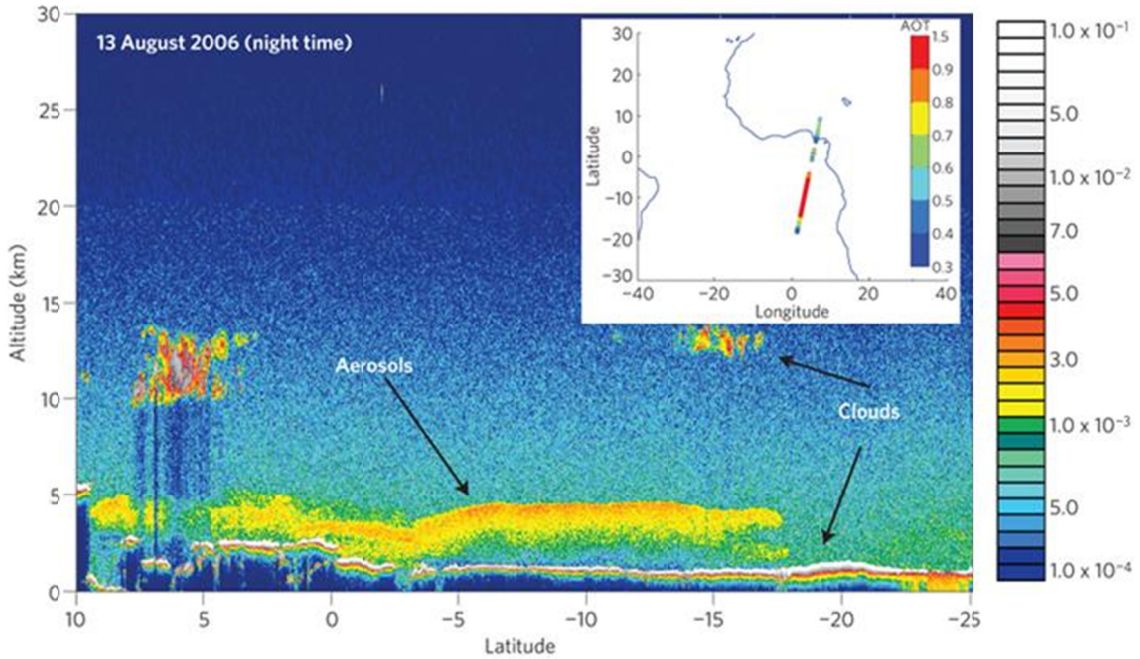

FIGURE 2-36 A vertical profile provided by the CALIOP system reveals a thick plume of aerosol haze from fires over a thinner layer of clouds. This image was captured over the southeast Atlantic Ocean, off southern Africa, during the time period shown. SOURCE: P. Lynch, 2009, "CALIPSO sees through the haze," http://www.nasa.gov/topics/earth/features/calipso-aerosol.html.

The LR-BSDS sensor for long-range detection of bio-aerosol clouds was shown in Figure 2-34; this system provides a limited ability to discriminate threats from other aerosols. There has been considerable interest in, and support for, development of systems that can positively identify bio-aerosols, and modern approaches to this problem combine multiple-sensor approaches, most notably the use of laser-induced fluorescence (LIF), discussed below in the section "Laser- Induced Fluorescence."Since effective bio-agent aerosols have diameters in the 1- to 10-μm range, DISC systems for this application would be expected to use wavelengths extending further into the infrared than sensors designed for scientific studies of the natural atmosphere.

One such device employed a Nd:YAG laser combined with an OPO to generate 1,064, 1,551 and 3,389 nm and used polarization as well as wavelength diversity to better discriminate bio-aerosols signals from others.[107] The discrimination technique employed not only DISC but a technique called wavelength normalized depolarization ratios (WANDER) to compare depolarization results at all three wavelengths. A recent publication reviews work in this area, including the use of CO_2-laser-based, multiwavelength DISC sensors in the 9- to 11-μm wavelength range, and suggests an optimal approach would be as follows.

> Depolarization measurements in SWIR/MWIR for the detection of unusual concentrations of non-background aerosols; elastic backscattering in SWIR for detecting unusual aerosol concentrations,

[107] J.J. Glennon, T. Nichols, P. Gatt, T. Baynard, J.H. Marquardt, and R.G. Vanderbeek, 2009, "System performance and modeling of a bio-aerosol detection lidar sensor utilizing polarization diversity," *Proc. SPIE* 7323, Laser Radar Technology and Applications XIV, 73230T, May.

and cloud mapping and tracking; UV Laser Induced Fluorescence (LIF) for discrimination of aerosols.[108]

In summary, applications of DISC technology have been in basic atmospheric studies as well as in systems to identify potential harmful biological agent attacks. DISC system development will likely be done only for scientific and defense applications.

DIFFERENTIAL ABSORPTION LIDAR

Differential absorption lidar (DIAL) seeks to measure the concentration of a gas (or gases) in the region illuminated by the laser transmitter. There are two general varieties, path-averaged, which measures gas over a specific path with no information on the distribution of gas in the path, and range-resolved, where data on the distribution of gas are generated. The measurement relies on the actual absorption of the transmitter light by the gas, not its scattering, where the absorption can be from either electronic transitions, or in the case of molecules, vibrational transitions. In the case of electronic transitions of molecules, they are actually coupled electronic-vibrational transitions, often called vibronic transitions, while the vibrational transitions in molecules are in fact coupled to the rotational states and called ro-vibrational transitions. Both lead to complex structure in the absorption properties.

Path-Averaged DIAL

An optical schematic of a path-averaged DIAL system appears in Figure 2-37. The transmitter power is reflected back to the receiver either by scatter from some available surface (a "topographic" target) or, for a fixed installation, a retroreflector designed to reflect back a large fraction of the incident light. An alternative "bistatic" approach places the receiver at the desired end of the path. A key element of the system is a tunable-wavelength transmitter that provides at least one wavelength absorbed by the gas or gases of interest ("on-line"), and one wavelength that is not ("off-line"). The key element of the measurement is to evaluate the difference between the received signals for the on-line and off-line wavelengths, which accounts for "differential" in the DIAL acronym. The measurement provides data on the presence of the gas in the path between the source and the reflecting surface but no data on where the gas may be in the path, only its average absorption level (hence the descriptor "path-averaged"). With enough tuning in the source, it is possible to detect multiple gases ("a" and "b" in the figure) along the path.

Conventional, path-averaged gas spectroscopy is often done with incoherent broadband sources, where the receiver employs a dispersive element to do spectral discrimination needed to detect the desired absorption lines of the gas or gases sensed. For industrial fence-line monitoring, incoherent-source systems that use Fourier-transform infrared spectroscopy (FTIR) have been deployed, operating over paths as long as 1 km. The outstanding feature of the FTIR system is the ability to cover a very broad wavelength range, 0.7-30 μm. This range covers the strong vibrational transitions of all molecules exhibiting vibrational absorption features. Defense applications have included detection of chemical weapons, with most of the important distinguishing features of chemical agents falling in the atmospheric window between 8 and 12 μm. Disadvantages of the FTIR system include the relative mechanical complexity and high capital cost of the device; as a result, they have found limited commercial success in recent years for atmospheric measurements, although they are widely used for laboratory work.

[108] S. Lambert-Girard, N. Hô, B. Bourliaguet, P.-F. Paradis, M. Piché, and F. Babin, 2012, "Proposal for a standoff bio-agent detection SWIR/MWIR differential scattering lidar," *Proc. SPIE* 8358, Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XIII, 835805, May 1.

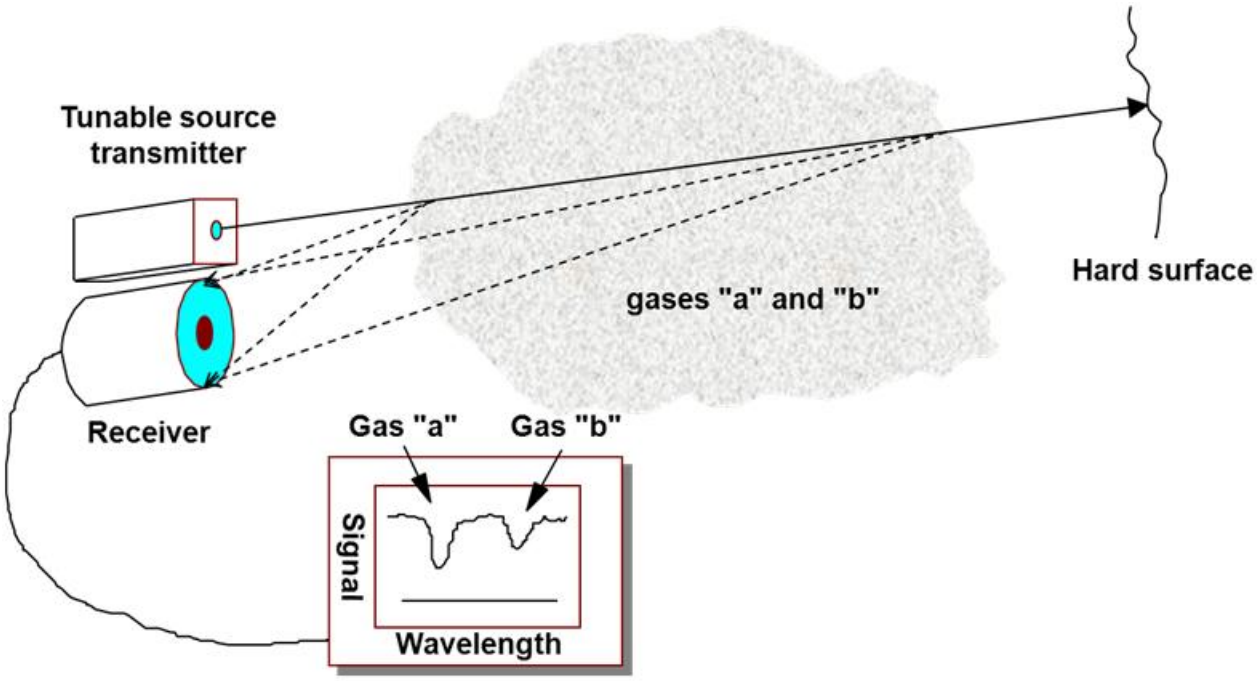

FIGURE 2-37 Schematic of path-averaged DIAL system.

The advantages of lasers for the path-averaged sensing application are as follows:

1. Much higher spectral brightness than incoherent sources, which allows longer paths or, for some systems, the use of available surfaces for reflection of the source back to the receiver rather than purposely located retroreflectors. For defense applications, such as the detection of chemical weapons, the ability to use an available surface rather than a retroreflector is a major advantage.
2. The higher SNR that is possible in the system is important in detecting the small changes with wavelength exhibited by small concentration of gases. With coherent sources, a variety of techniques to extend detection sensitivity have been devised, such as wavelength modulation spectroscopy, and the ultimate detection sensitivity of laser-based systems is many orders of magnitude higher than with incoherent-source systems. In applications involving measurements in the atmosphere, the high sensitivity is often unusable because of absorption from other gases.
3. The higher spectral resolution that is possible is useful when there are multiple interfering absorption features from other gases in the path that interfere with the gas to be sensed.

The main disadvantages of coherent sources have included the limited wavelength coverage, a problem when either the gases to be sensed are unknown or a wide range of gases are to be detected, and the relative cost and complexity of the coherent source. Past work on path-averaged DIAL systems has included the use of tunable, pulsed, high-energy CO_2 lasers in the 9- to 11-μm region to measure vibrational transitions, where the high available powers allowed multikilometer ranges against topographic targets. As for spectral coverage, there have been recent developments in broadband sources based on nonlinear optics that provide significant power over much of the IR region, and future advances in path-averaged systems may be able to exploit these sources, which provide broadband power from a diffraction-limited source. This is discussed in more detail in Chapter 3, in the section on femtosecond lasers.

For scientific applications, one dramatic application of path-averaged DIAL is in space-based measurements of Earth's atmosphere, where the satellite contains the tunable transmitter as well as the receiver and the signal is reflected from Earth's surface. As an example, systems for measurement of the spatial distribution of CO_2 throughout the planet are under development by NASA, probing absorption by CO_2 vibrational levels in the 1,500- and 2,000-nm wavelength regions. By tuning through the absorption lines and analyzing the absorption lineshape, it may be possible to get some understanding of the vertical distribution of the CO_2 concentrations, since the spectral linewidths are strongly dependent on temperatures and hence altitude. Given the long ranges involved, the NASA programs plan to use erbium- or thulium-doped fiber lasers to generate the required power levels.

While path-averaged DIAL does not directly provide spatial information (with the exception of high-altitude or space-based, down-looking sensors, which may be able to take advantage of the absorption lineshape change with altitude), it is possible to employ tomographic techniques to build up spatial information, when it is possible to compare data from a variety of different path angles. The simplest example would be a ground-based system that could make use of multiple reflecting surfaces at different angles or ranges. Another example is an aircraft-based, down-looking sensor that obtains signals from reflections off the ground and can provide wide-area data that can be analyzed to localize sources of the gases sensed. Given advances in source and detector technology, it should also be possible to develop space-based sensors for a variety of gases beyond those, such as CO_2, that are of interest for science.

When semiconductor diodes are employed as sources, the technology is sometimes called tunable diode laser absorption spectroscopy (TDLAS), and a variety of sensitive gas measurement techniques have been developed to leverage the ability to rapidly amplitude- or frequency-modulate the diode output. One important application for long-path measurements is detection of very weak atmospheric absorption features, of interest to basic science spectroscopy as well as possible effects of atmospheric absorption on directed energy laser systems. Figure 2-38 shows a schematic of one TDLAS system, designed to measure atmospheric absorption due to weak electronic transitions of O_2 in the 895-nm region. This is the region where the diode-pumped cesium alkali directed energy laser (DPAL) operates.[109] The system employs 100-kHz amplitude modulation along with electronic coherent detection to improve the SNR. Spectral data from the system for 150- and 1,000-m paths appear in Figure 2-39, which shows the location of the cesium laser transitions as well as the weak absorption lines due to the vibronic O_2 transitions. Owing to the very weak absorption levels, the measurements would not be possible with conventional incoherent sources.

The relatively recent development of cascade semiconductor lasers (Chapter 4) has led to much wider wavelength coverage for TDLAS of the infrared molecular absorption bands, and development of a "killer app" for these devices in gas sensing could lead to a major reduction in price. At present, many commercial TDLAS systems make use of more conventional interband lasers, originally developed for telecommunications applications. These are readily available and relatively low in cost. Table 2-2 lists the detection sensitivities for a variety of gases for TDLAS sensors, primarily based on near-IR diode lasers (wavelengths less than 2.2 μm) which sense the overtone vibrational levels of the molecules. For two gases there is greatly increased sensitivity when one employs longer wavelength QCL-based sources, which sense the much more strongly absorbing fundamental vibrational absorption bands. The sensitivities are expressed in parts-per million meters (ppm-m), so for a path length of 1 km the sensitivity would be greater than listed by a factor of 1,000. For example, for sensing of CO gas, one can detect 40 parts per billion (ppb) over a 1 km path with a near-IR source, while a sensor at 4.8 μm could, in principle, detect levels well below a ppt. (In practice, the ambient levels for CO in a normal atmosphere are well over this level.)

[109] C.A. Rice, G.E. Lott, and G.P. Perram, 2012, "Open-path atmospheric transmission for a diode-pumped cesium laser," *Appl. Opt.* 51: 8102.

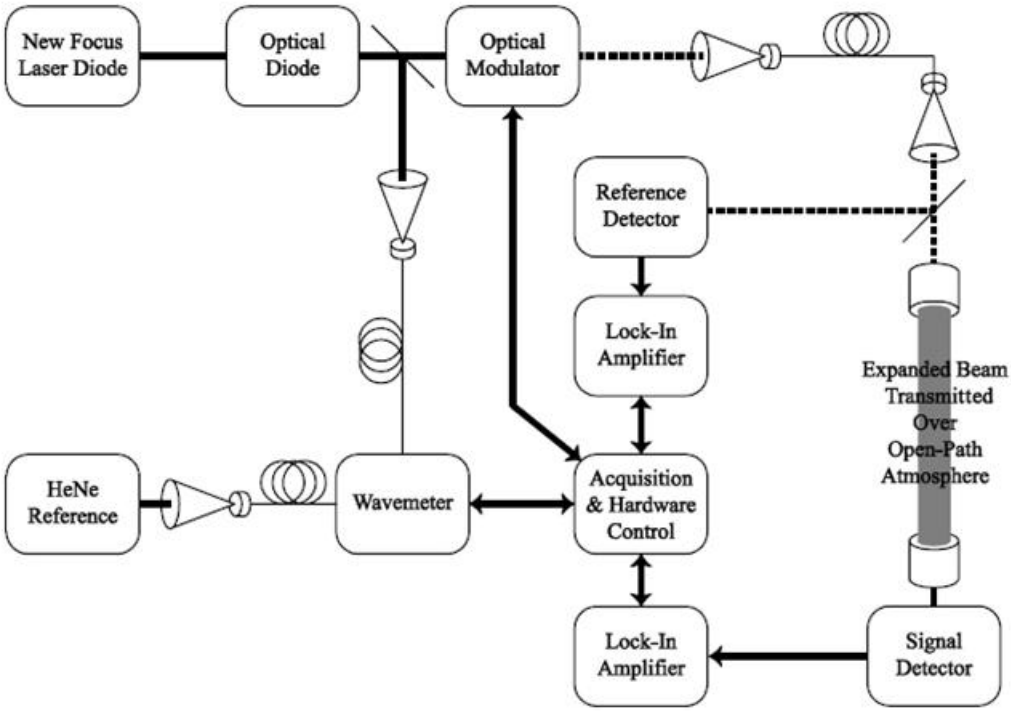

FIGURE 2-38 Schematic of path-averaged TDLAS system operating at 895 nm. SOURCE: C. Rice, G. Lott, and G. Perram, 2012, "Open-path atmospheric transmission for a diode-pumped cesium laser," *Applied Optics* 51**:** 8102.

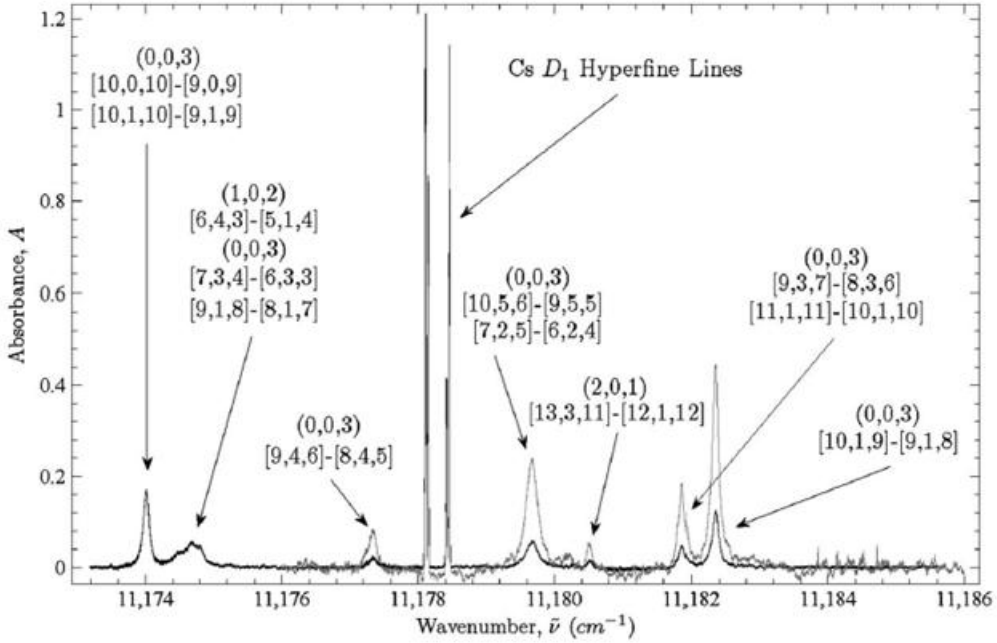

FIGURE 2-39 Data from 895-nm TDLAS system for 150- (black) and 1,000-m (gray) paths. The broad lines, with level identifications, are from O_2 atmospheric absorption. SOURCE: C. Rice, G. Lott, and G. Perram, 2012, "Open-path atmospheric transmission for a diode-pumped cesium laser," *Applied Optics* 51: 8102.

TABLE 2-2 Detection Sensitivities of TDLAS Systems to Various Gases. SOURCE: Mark Allen, Physical Sciences Inc., June 5, 2013.

ppm-m					
Near-IR (<2.2 μm)				2 to 3 μm	4 to 8 μm
Gas	300 K 1 atm	Gas	300 K 1 atm	300 K 1 atm	300 K 1 atm
HF	0.2	HCN	1.0		
H_2S	20.0	CO	40.0	0.02	0.0001
NH_3	5.0	CO_2	1.0		
H_2O	1.0	NO	30.0	0.6	0.03
CH_4	1.0	NO_2	0.2		
HCl	0.15	O_2	50.0		
CH_3CN	10.0	$CH_3CHOHCH_3$	20.0		
CH_2Cl_2	10.0	CH_3CH_2OH	20.0		
		CH_3OH	20.0		

For defense applications, as noted above, the 8-12μm wavelength range is well understood and documented for its usefulness as the "fingerprint region" for chemical weapons, generally heavy molecules with strong fundamental vibrational-rotational absorption bands in the region. It is likely that FTIR-based sensors will be replaced in the future by arrays of QCL sources that provide full coverage of the region and allow longer range and more precise path-averaged sensing, with localization provided via tomographic approaches that could employ UAV-based sensors based on reflection from the ground.

Another defense application is explosives detection, where the goal would be detect gases given off by the solid explosives or those associated with explosives production. Here the challenges are detection sensitivity, as the vapor pressures for most explosives themselves are very low.[110] The production of explosives, particularly low-technology, so-called homemade explosives (HMEs) such as urea nitrate (UN), ammonium nitrate (AN) and nitromethane (NM) may involve materials such as ammonia that produce more detectable gases. Remote explosives detection by path-averaged DIAL is clearly an area of interest for both defense and homeland security.

Compared to sensors based on incoherent sources, DIAL systems have been able to detect much smaller concentrations of gas but, until recent developments in broadband lasers, they could not match the broad wavelength ranges of the incoherent sensors. The development of practical, room-temperature QCLs is set to provide much higher gas-detection sensitivities through operation in the fundamental vibrational absorption bands of molecular gases. Thus far, military applications of DIAL have been primarily defensive in nature and include the detection of chemical agents in the so-called molecular fingerprint region of 8- to 12-μm wavelength. The DIAL technology is well known throughout the world. Although there are differences in development of QCL-based sources, any gaps are likely to narrow, especially if industrial sensors find wider deployment.

[110] J. Steinfeld and J. Wormhoudt, 1998, "Explosives detection: A challenge for physical chemistry," *Annu. Rev. Phys. Chem.* 49(1): 203.

Conclusion 2-6: Path-averaged differential absorption lidar will see a large increase in sensitivity through the use of quantum cascade lasers.

Conclusion 2-7: With lightweight differential absorption lidar systems enabled by semiconductor lasers and emerging broadband coherent sources (see Chapter 3), the ability to place sensors on unmanned aerial vehicles should allow development of three-dimensional data from relatively simple sensors.

Range-Resolved DIAL

While path-averaged DIAL systems can provide significantly improved performance in terms of gas-detection sensitivity compared to systems based on incoherent sources, or allow measurements from space, range-resolved DIAL technology (Figure 2-40) provides a truly unique capability. It generally utilizes a pulsed, tunable source (or sources) and light backscattered by aerosol/Rayleigh scattering and compares the range-gated return signals as a function of wavelength to measure the absorption of the light by a gas or gases in the region probed by the laser. With enough tunability in the transmitter one can detect different gases ("a" and "b" in the drawing) as a function of range, but many DIAL systems are designed to detect only one gas. The simplest system can use one online and one off-line measurement wavelength, but more sophisticated DIAL systems employ multiple off- and online wavelengths.

The sensitivity of range-resolved DIAL to sensing gases depends on the signal-to-noise levels of the returned signals, which must be subtracted from one another. The presence of atmospheric fluctuations from the aerosol distributions and variations in atmospheric refractive index provides another source of noise, and more advanced DIAL systems will send off- and online pulses closely spaced together to minimize changes in the atmosphere, with a spacing less than 1 ms sufficient to "freeze" the atmosphere. Even with this technique, it is difficult for operating DIAL systems to detect less than a 1 percent difference in the strength of the returned signals. The need to reduce fluctuations in the returned signals generally limits DIAL receivers to the use of direct detection, as the pulse-to-pulse speckle fluctuations from coherent detection are large.

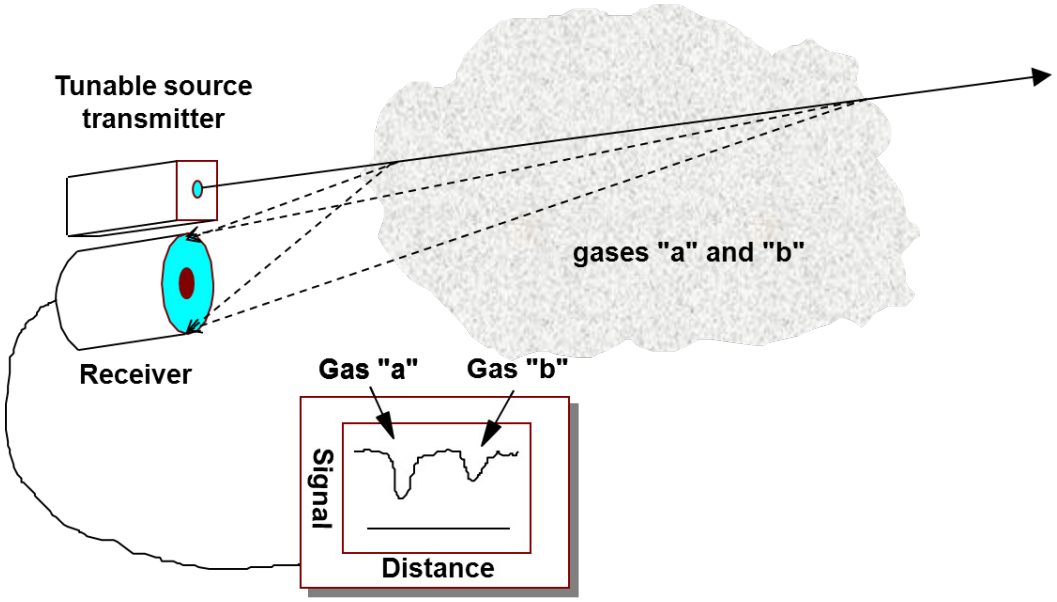

FIGURE 2-40 Schematic of range-resolved DIAL system.

Given the need to employ atmospheric backscatter, range-resolved DIAL systems for long-range use are generally limited to shorter wavelengths and thus probe absorption due to electronic transitions. Many gases of special interest, such as CH_4 and a variety of related hydrocarbons, do not have electronic transitions low enough in energy to be accessed within the UV transmission range of the atmosphere and thus can only be sensed by range-resolved DIAL systems operating in the infrared. Operating ranges are limited compared to the short-wavelength systems, and the system performance may vary depending on the aerosol content of larger-diameter particles.

The applications of range-resolved DIAL can be divided into scientific and applied, the latter including the detection and monitoring of pollution and the detection of leaks in gas pipelines.

Basic atmospheric science programs include monitoring of global distributions of water vapor and ozone. Water vapor is by far the most important greenhouse gas, and understanding its distribution in the atmosphere is of great importance. The NASA LASE system[111] employs a pulsed Ti:sapphire laser at 815 nm sending out pulse pairs separated by 0.4 ms in time and 70 picometer in wavelength to generate the DIAL data, with a pulse energy of 150 mJ. The receiver employs a 40-cm-diameter telescope and a silicon avalanche photodiode. Initial missions were flown based on an ER-2 platform, at altitudes of 16-21 km. The data product from one mission appears in Figure 2-41, showing the complex nature of water vapor distributions in the atmosphere. Before development of the DIAL system, equivalent data, with much less global coverage, had to be generated by balloon-born radiosondes. A ground-based water-vapor DIAL system, situated on the Zugspitze, (2.7-km ASL) a mountain in Germany, also employs a high-energy (250 mJ, 2 ns pulsewidth) Ti:sapphire laser around 817 nm and provides range-resolved data up to an altitude of 12 km.[112]

DIAL operating in the UV wavelength region can take advantage not only of aerosol backscatter but also of the more predictable Rayleigh backscatter. Ozone is of interest from both a scientific and an applied standpoint. At high altitudes, (stratospheric) ozone is crucial in absorbing harmful deep-UV light from the sun, while at low altitudes, (tropospheric) ozone is a pollutant that affects heath and damages vegetation and is one of the components of smog in cities. Ozone absorption has a very broad, generally unstructured band. Operational ozone DIAL systems have been based on the ground for providing data on tropospheric ozone, while aircraft-based systems have pointed the DIAL system downward for tropospheric measurements or upward for stratospheric measurements. For detection in the stratosphere, where ozone concentrations are high, typical on/off wavelengths used are in the range 308-355 nm, while for lower concentrations in the troposphere, 289 and 299 nm are used as on/off wavelengths. For systems to date, detection sensitivities are in the 10 ppbv range, and typical data show ozone levels in a large range around 100 ppbv.[113] In the future, NASA and other agencies are planning to build space-based ozone DIAL systems to provide full global coverage.

The other gases—O_3 (ozone), SO_2, and NO_2—are of interest for monitoring pollution created by burning fossil fuels, various industrial activities, internal combustion engines, or volcanoes. SO_2 resulting primarily from burning coal, is toxic, with death possible for prolonged exposure at 1 ppm and was reportedly one of the gases that caused more than 10,000 deaths during the London Great Smog of 1952. More recently, SO_2 has been identified as one of the major causes of "acid rain." With the development of range-resolved DIAL systems operating around 300 nm it became possible to track the generation and distribution of SO_2. A variety of mobile, ground-based DIAL systems were deployed in the United States and Europe starting in the 1970s to better understand how SO_2 was dispersed from sources, and early results showed that the gas spread over a wider region (100-1,000 km) than expected. This led to major changes in the regulation of emissions from power plants. Recently, SO_2 lidar systems have been

[111] See http://www.nasa.gov/centers/langley/news/factsheets/LASE.html.

[112] H. Vogelmann and T. Trickl, 2008, "Wide-range sounding of free-tropospheric water vapor with a differential-absorption lidar (DIAL) at a high-altitude station," *Appl. Opt.* 47: 2116.

[113] A.O. Langford, 2012, "Differential Absorption Lidar (DIAL) for characterizing ozone distributions and transport," WESTAR Conference on Western Ozone Transport, Boulder, Colorado, October 11, http://www.westar.org/12percent20Tech percent20Conf/Presentations/Langford.pdf.

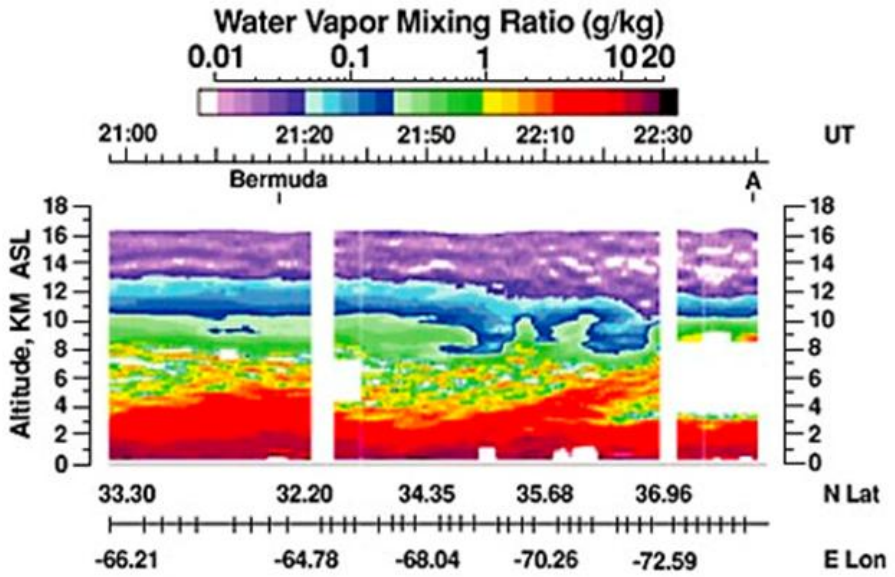

FIGURE 2-41 Measurement of water vapor by LASE onboard the ER-2 aircraft that flew from near Bermuda to Wallops Island, Virginia. The horizontal scales indicate the geographical coordinate of the measurement (at the Universal Time, UT) and the vertical scale indicates the vertical location. The color scale from pink (light detection) to black (heavy detection) indicates the concentration of water vapor in the atmosphere. White areas in the image indicate measurements below instrument threshold or a lack of data. SOURCE: NASA Langley Research Center, Lidar Atmospheric Sensing Experiment (LASE): Measuring Water Vapor, Aerosols and Clouds, http://www.nasa.gov/centers/langley/news/factsheets/LASE.html, accessed September 6, 2013.

developed in China to better understand the causes of severe smog events in cities such as Beijing, exacerbated by the Chinese use of coal-fired power plants. With a 10-mJ-energy, 10-ns source around 300 nm, reported sensitivities are 15 ppb for a 300-m averaging path, with a range extending to 3 km.[114]

Nitrogen dioxide and its related molecule NO are created by cars, trucks, and other combustion sources; reacting with hydrocarbons and sunlight, they lead to the formation of tropospheric ozone. The mix of gases forms smog, with the visible-wavelength absorption for NO_2, leading to the brownish atmospheric color evident in major smog events. DIAL systems have also been used to track the formation and dispersion of NO_2 and operate in the 450-nm wavelength region. With a 150-mJ source at 450 nm, sensitivities are 10 ppb for a 300-m path, and the range is up to 6 km.[115] NO can be sensed as well, but the detection wavelength, around 227 nm, has very limited transparency in a sea-level atmosphere, owing to the absorption of O_3 and other molecules, and the sensing range is limited to 0.5 km.[116,117]

[114] W. Staehr, W. Lahmann, and C. Weitkamp, 1985, "Range-resolved differential absorption lidar: Optimization of range and sensitivity," *Appl. Opt.* 24: 1950.

[115] Ibid.

[116] M. Aldén, H. Edner, and S. Svanberg, 1982, "Laser monitoring of atmospheric NO using ultraviolet differential-absorption techniques," *Opt. Lett.* **7:** 221.

[117] Z.G. Guan, P. Lundin, L. Mei, G. Somesfalean, and S. Svanberg, 2010, "Vertical lidar sounding of atomic mercury and nitric oxide in a major Chinese city," *Appl. Phys.* B101: 465.

Other entities measured by UV DIAL include air toxics such as atomic mercury at 254 nm,[118] generated primarily by volcanoes, coal-fired power plants, and waste incineration; Cl_2 gas around 300 nm,[119] often released by industrial accidents; and the aromatic hydrocarbons benzene, toluene, ethylbenzene, and the three isomers of xylene (collectively referred to as the BTEX compounds) around 220-270 nm.[120] Benzene has long been considered a carcinogen. Emission sources for BTEX compounds include refineries, gasoline service stations, and vehicle exhaust. In areas where mobile emissions are significant, ambient concentrations of BTEX are tens to perhaps 200 ppb C and represent the majority of total toxic organic compounds. Because, with the exception of benzene, these aromatics are moderately reactive in photochemical smog formation, their chemistry in the atmosphere is of great interest.

As noted above, mid-IR range-resolved DIAL systems cannot provide the range of shorter-wavelength systems, and their performance is more dependent on the variable concentrations of larger atmospheric aerosols. However, there is interest in sensing of the wide variety of hydrocarbons that have fundamental, intense vibrational transitions in the 3,300-3,400-nm wavelength region, both for science and for applications, which include not only pollution monitoring but also detection of gas leaks over pipelines and, possibly, exploration activities to discover new sources of oil and gas.[121] Probably the most utilized mid-IR DIAL system has been developed by the U.K. National Physical Laboratory (NPL).[122] The mid-IR source was used in combination with a UV source to make a variety of ground-based range-resolved DIAL measurements, and Table 2-3 summarizes the results. The technology employed in the NPL dates back to before 2000; for DIAL systems built now, a variety of improvements in sources and, for the mid-IR range, detectors, would allow longer ranges and, perhaps, higher detection sensitivities.

Many of the ground-based pollution measurements have been from mobile platforms that have allowed measurements at a variety of locations throughout the world. One of the most notable systems appear in Figure 2-42, from NPL, which has pioneered in the mid-IR sensing of hydrocarbons. Most recently, another such system was shipped to China, to measure Hg and NO in the city of Hangzhou.[123] In terms of systems engineering, these mobile sensors employ commercial lasers designed for operation in scientific laboratories; they are essentially like a large laboratory that has been given wheels, without any attempt at miniaturization or ruggedization. The helicopter-based aerosol sensor (Figure 2-34) is an example of a much more engineered sensor, with a source more powerful than those in the mobile lidar trucks.[124]

[118] H. Edner, G. W. Faris, A. Sunesson and S. Svanberg, 1989, "Atmospheric atomic mercury monitoring using differential absorption lidar techniques," *Appl. Opt.* 28: 921.

[119] H. Edner, K. Fredriksson, A. Sunesson, W. Wendt, 1987, "Monitoring Cl_2 using a differential absorption lidar system," *Appl. Opt.* 26: 3183.

[120] M.J.T. Milton, P.T. Woods, B.W. Jolliffe, N.R.W. Swann, and T.J. McIlveen, 1992, "Measurements of toluene and other aromatic hydrocarbons by differential-absorption LIDAR in the near-ultraviolet," *Appl. Phys.* B55: 41.

[121] U.S. Patent US 6509566 B1, "Oil and gas exploration system and method for detecting trace amounts of hydrocarbon gases in the atmosphere," Publication date Jan 21, 2003, and references cited therein.

[122] M.J.T. Milton, T.D. Gardiner, F. Molero, and J. Galech, 1997, "Injection-seeded optical parametric oscillator for range-resolved DIAL measurements of atmospheric methane," *Opt. Commun.* 142: 153.

[123] Z.G. Guan, P. Lundin, L. Mei, G. Somesfalean, and S. Svanberg, 2010, "Vertical lidar sounding of atomic mercury and nitric oxide in a major Chinese city," *Appl. Phys.* B101: 465.

[124] M. Sjöholm, P. Weibring, H. Edner, and S. Svanberg, 2004, "Atomic mercury flux monitoring using an optical parametric oscillator based lidar system," *Opt. Exp.* 12: 551.

TABLE 2-3 Detection Sensitivities and Maximum Ranges of NPL Range-Resolved DIAL System for Various Gases

Infrared DIAL System			UV/Visible DIAL System		
Species	Sensitivity	Max. Range	Species	Sensitivity	Max. Range
CH_4	50 ppb	1 km	NO	5 ppb	500 m
C_2H_2	40 ppb	800 m	NO_2	10 ppb	500 m
C_2H_4	10 ppb	800 m	SO_2	10 ppb	3 km
C_2H_6	20 ppb	800 m	O_3	5 ppb	2 km
higher alkanes	40 ppb	800 m	Hg	0.5 ppb	3 km
HCl	20 ppb	1 km	Benzene	10 ppb	800 m
N_2O	100 ppb	800 m	Toluene	10 ppb	800 m
CH_3OH	200 ppb	500 m	Xylene	20 ppb	500 m

NOTE: NB. The sensitivities apply at a range of 200 m for a 50 meter plume.
SOURCE: M.J.T. Milton, T.D. Gardiner, F. Molero and J. Galech, 1997, "Injection-seeded optical parametric oscillator for range-resolved DIAL measurements of atmospheric methane," *Opt. Commun.*142: 153.

FIGURE 2-42 NPL (U.K.) mobile lidar truck. SOURCE: Courtesy of NPL.

Range-resolved DIAL would be of great use for a number of defense applications, requirements, including the detection of chemical and biological weapons and buried explosives. Unfortunately, the signature absorption bands for chemical weapons fall in the 8-12 μm atmospheric window, where aerosol scattering is weak and highly variable. This means that detection is better done with path-averaged DIAL, employing tomographic techniques if spatial information is needed. Overtone transitions of chemical agents in the 3-5-μm window better suited for range-resolved DIAL tend to exhibit very broad and indistinguishable absorption features, making agent identification difficult. Biological agents, effectively very large molecules, do not exhibit identifiable absorption features. As noted above, the low vapor pressures of explosives[125]make detection even by path-averaged DIAL a challenge. The production of

[125] J. Steinfeld, and J. Wormhoudt, 1998, "Explosives detection: A challenge for physical chemistry," *Annu. Rev. Phys. Chem.* 49(1): 203.

HMEs might generate NH_3 but the most intense absorption bands are at around 10 μm, and the weaker overtone transitions overlap with common atmospheric components such as water vapor and CO_2.

DIAL systems have been considered for some law enforcement applications, such as the detection of drug production facilities. As an example, kerosene, diethyl ether, and acetone are all used in the production of cocaine, and a range-resolved DIAL system design might be able to locate production operations normally hidden in a jungle or in some industrial areas. Methamphetamine production involves a number of volatile chemicals, likely detectable with DIAL techniques, although the number of different production techniques suggests there may not be one chemical that could be used as a clear indicator.

Conclusion 2-8: Operation of differential absorption lidar at long ranges is easiest with the use of ultraviolet and visible wavelengths, which provide the strong aerosol and Rayleigh backscatter needed. An important, but limited, set of gases can be sensed at these wavelengths.

Conclusion 2-9: The range of differential absorption lidar systems is reduced at longer wavelengths and is severely limited out in the 3-5 and 8-12 μm atmospheric window regions, though improvements in sources and detectors are countering this.

Conclusion 2-10: Range-resolved differential absorption lidar is an important tool for scientific studies and will find further deployment on airborne and space-based platforms.

RAMAN SENSING

The Raman effect is the inelastic (non-wavelength-preserving) scattering of photons from a species, through excitation of electronic or vibrational states. A Raman lidar system (shown schematically in Figure 2-43) senses the inelastic scattered photons in a receiver tuned to a different wavelength, or wavelengths, from that of the transmitter; it can develop range-resolved data (gases "a' and "b" in Figure 2-43) by analysis of the time-resolved return signals at different wavelengths.

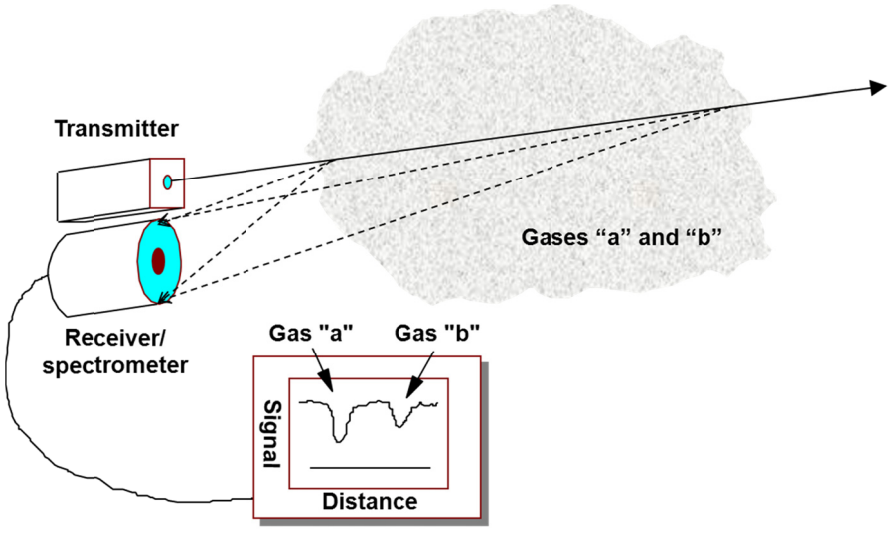

FIGURE 2-43 Schematic of Raman lidar system.

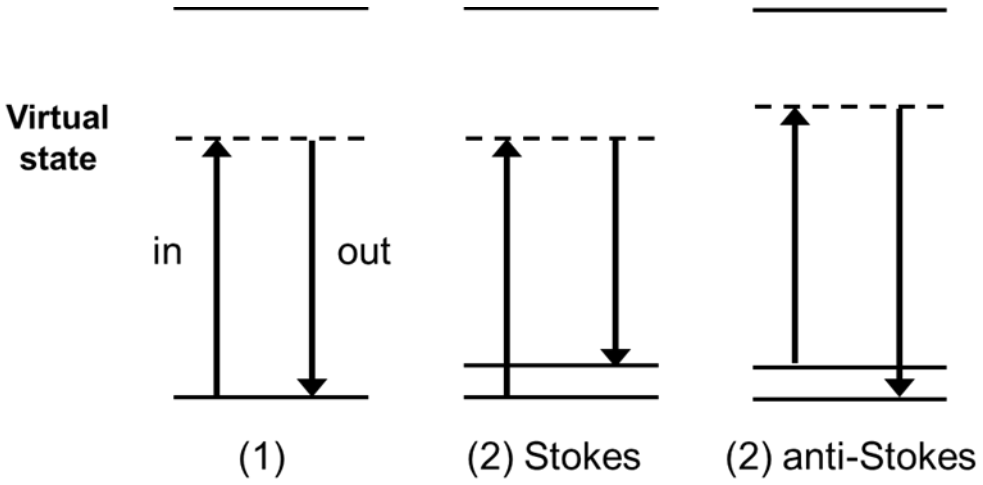

FIGURE 2-44 Comparison of scattering processes from atmospheric species: (1) elastic, (2) conventional (Stokes) Raman, and (3) anti-Stokes Raman.

Figure 2-44 shows the conventional Rayleigh scattering process, which leaves the photon going out with essentially the same energy (wavelength) as the photon coming in, but with a different propagating direction (i.e., momentum). From a quantum-mechanical treatment, the incoming light induces a transition from the ground state to a "virtual" electronic state in an atom or molecule of the species, which very rapidly (on the order of 10^{-16} s) reemits the light with the same wavelength, based on the same ground electronic state. The Raman process (2) differs in that the reemitted light terminates on a different ground state, most typically another vibrational or rotational level for a molecular species, and is lower in energy and thus longer in wavelength. The probability of a Raman scattering processing is typically less than 10^{-3} that of Rayleigh scattering,[126] with same $(1/\lambda)^4$-dependence, favoring the use of short wavelengths for detection. When the photon scatters off a molecule that is already vibrating, the anti-Stokes process (3) yields a higher energy scattered photon with a shorter wavelength. At the higher energy of vibrational levels, very few molecular vibrations are already excited, so the fraction of anti-Stokes scattering is negligible. For rotational levels, however, the energies are small enough to allow a reasonable level of anti-Stokes scattering.

When the excitation wavelength approaches the energy of the upper electronic level, the probability of the Raman process greatly increases (resonance Raman), much more than the $(1/\lambda)^4$ dependence, although for many species the wavelengths needed to observe a significant resonant enhancement are too short to be transmitted through the atmosphere.

The details of the spectral nature of the Raman scattering process depend on the nature of the species. This discussion considers only the sensing of molecules. Simple diatomic molecules like N_2 and O_2 at normal atmospheric temperatures have a single strong Raman line from the vibrational level with weak sidebands that result from the addition of rotational levels to the interaction, as well as purely rotational Raman scattering. More complex molecules with a number of vibrational modes have number of Raman lines, again with rotational structure except for spherically symmetric molecules. As with absorption spectra, molecules have unique Raman spectra, or fingerprints, that allow identification. For complex molecules some vibrations lead to the absorption of light while others only show up in Raman scattering.

A typical Raman lidar has a fixed-wavelength transmitter, with a receiver that may be (1) tuned to a specific Raman-shifted longer wavelength when only one species is to be sensed or, (2) capable of

[126] See http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730009019_1973009019.pdf.

spectrally resolving the returned signal with, say, a detector array, when multiple species are to be detected.

Advantages of Raman lidar include the following:

- No atmospheric scattering by entities other than the one to be sensed is needed for measurement.
- Each species produces a distinct Raman line (or lines) displaced in wavelength from the source, and one lidar system with only one wavelength laser can sense multiple species.
- Since the expected return signal is shifted in wavelength from the source laser, the appropriate filters can make the detection system immune to (elastic) backscatter of the transmitted light.
- The lidar source does not need to be tunable.
- Because Raman scattering has a weak probability, there is no appreciable attenuation of the source beam by the species. Thus large, concentrations can be measured over a long distance.

Problems with Raman lidar include these:

- Except for resonance Raman, the scattering from species is a weak process, much weaker than Rayleigh.
- Compared to DIAL the sensitivity is many orders of magnitude lower.
- When short wavelengths are employed for detection, fluorescence (see the next section) provides interference from species where the short-wavelength light creates real rather than virtual electronic transitions. This is especially true with resonance Raman, which employs wavelengths close to, if not actually overlapping with, real electronic transitions.

The last problem can be mitigated by the use of multiple transmitted wavelengths, since the fluorescence spectrum is generally unchanged with small changes in excitation wavelength, whereas the Raman signal shifts linearly with the transmitter wavelength.

The weakness of the Raman process generally limits Raman lidar to the detection of atmospheric species that have high concentrations, and deployed systems are typically used for scientific applications in atmospheric probing, such as detection of N_2, O_2, and water vapor as a function of altitude. For each of the purely vibrational (Q-branch) Raman lines there are weaker lines from vibrational-rotational Raman scattering, and a determination of the relative intensities of the lines allows the determination of the molecular temperature. In addition to the gas Raman lines, there is Raman scattering from liquid- and ice-phase water. Clearly, a Raman lidar is able to measure not only gas density and temperature, but also the phase state of atmospheric water.

An example of a Raman lidar used for atmospheric measurements is a facility run by the U.S. Department of Energy (Figure 2-45) at its Lamont, Oklahoma site; another, similar facility is in operation at Darwin, Australia. "The lidar uses a commercial lamp-pumped, tripled Nd:YAG laser, operating at 30 Hz with 300-400 mJ pulses, to transmit light at 355 nm. A 61-cm diameter telescope collects the light backscattered by molecules and aerosols at the laser wavelength and the Raman scattered light from water vapor (408 nm) and nitrogen (387 nm) molecules."[127] "These signals are detected by photomultiplier tubes and recorded using photon counting with a vertical resolution of 7.5 m."[128] A range-resolved data product on water vapor mixing for the vertically directed lidar appears in Figure 2-46.

[127] R.K. Newsom, 2009, Raman Lidar (RL) Handbook, March, DOE/SC-ARM/TR-038.
[128] Ibid.

FIGURE 2-45 The Raman lidar (RL) at Lamont, OK, operated by the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) Climate Research Facility. SOURCE: U.S. DOE Atmospheric Radiation Measurement Program.

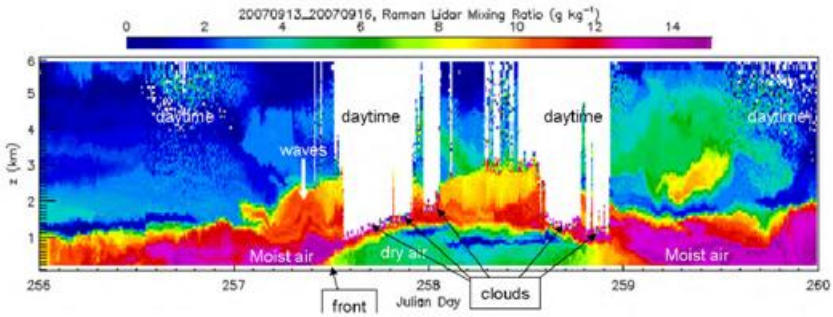

FIGURE 2-46 Data product as a function of time over 4 days showing the water mixing ratio in the atmosphere as a function of altitude. SOURCE: US DOE Atmospheric Radiation Measurement Program.

Similar systems are in use worldwide. To name a few, NASA Goddard has built a mobile-trailer based system, the Scanning Raman Lidar (SRL), and deployed it at a variety of locations: the Max Planck Institute for Meteorology in Hamburg, Germany, has a system in operation in Barbados; the Atmospheric Science Programme office of the Department of Space in India has built a system that is in operation in India; and Chinese-built Raman lidars have been reported to be operating in Beijing, Xi'an, and Wuhan.

For defense applications, other than providing localized meteorological data, Raman lidar systems do not provide the sensitivities required for the standoff atmospheric detection of chemical, biological and explosives at other than trivial ranges.

A variation on remote Raman detection is designed to sense the presence of an entity on a surface rather than in the atmosphere. One application that has been investigated is detection of surface-deposited chemical agents by 266-nm, UV illumination of surfaces at ranges up to 500 m,[129] and there is continuing work on employing short-range Raman sensors for surface contamination measurements, for both defense and homeland security uses. However, it appears that to detect low levels of chemical agents the system sensitivity is not sufficient to work beyond meter-level ranges.

Since the discovery of surface enhanced raman scattering (SERS),[130] where the Raman scattering signal for an entity on a particular metallic surface may be increased by a factor of 10^6-10^{14} with near- to visible-wavelength laser excitation, there have been applied studies of the technique to detect very small levels of entities on surfaces, especially chemical and biological agents and explosives. To date, most of the work has been for very short range detection, but in the future the detection range could be extended with suitable sources. Considerable challenges remain, since the enhancement varies over many orders of magnitude depending on the metal and its surface condition.

Conclusion 2-11: Raman detection provides data on range-resolved concentration of gases, but the low sensitivity of the technique limits its use to major atmospheric constituents such as nitrogen and water vapor.

Conclusion 2-12: Applications of Raman sensing technology are likely to be limited to scientific research.

[129] See http://www.ecd.bnl.gov/pubs/BNL69444AB.pdf.

[130] M. Fleischmann, P.J. Hendra, A.J. McQuillan, 1974, "Raman spectra of pyridine adsorbed at a silver electrode," *Chem. Phys. Lett.* 26: 163.

LASER-INDUCED FLUORESCENCE

Laser-induced fluorescence (LIF) lidars (see Figure 2-47) are similar to Raman lidars in that they require only that the entity to be detected be present in the atmosphere or on the surface of a target of interest. Unlike Raman lidars, however, LIF systems seek to create real transitions in the entity through excitation to higher electronic levels. In some cases, the entity returns to the ground electronic level by, in part, emitting light (fluorescing), and a LIF lidar system employs a receiver tuned to the wavelength region of the fluorescence. Through time analysis of the fluorescence, it is possible to localize the detected entity in the atmosphere, thus provided range-resolved data, and by employing a tunable receiver, or multiple receivers, one can detect multiple entities ("a" and "b") as a function of range. It is also possible to detect fluorescence from entities on a surface. Typical systems employ pulsed UV sources for the required excitation of the electronic levels as well as range/species resolution. The decay time of the returned signal, as well as the fluorescence spectrum itself, can provide some ability to distinguish different entities.

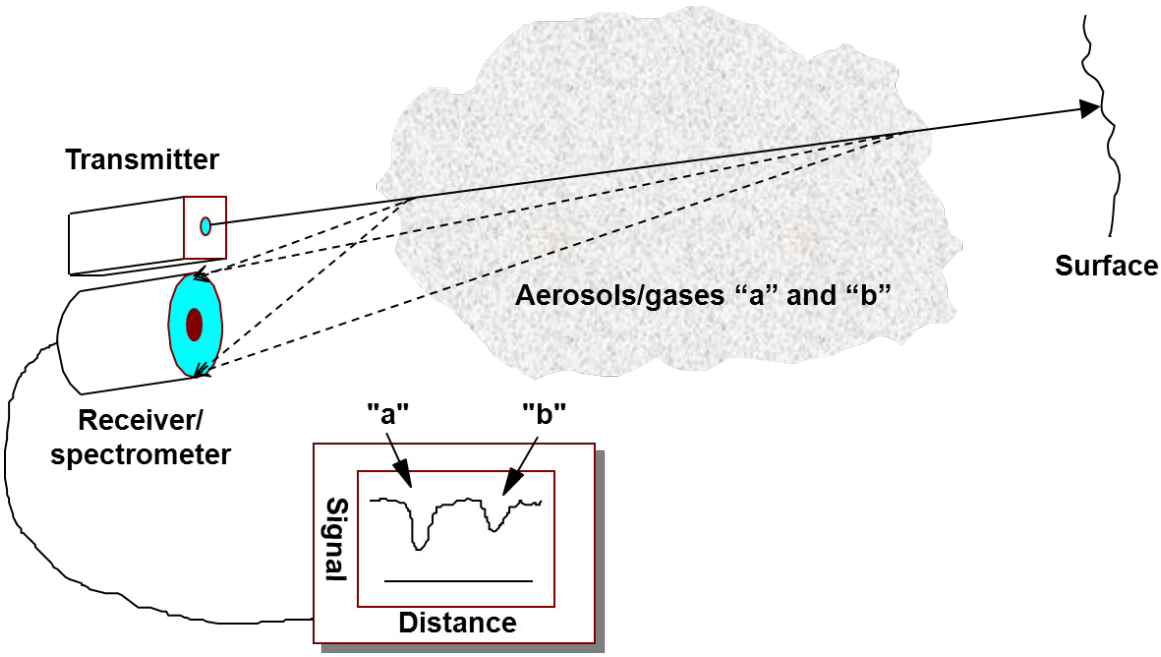

FIGURE 2-47 Schematic of LIF lidar.

Advantages of LIF systems include the following:

- Like Raman lidar no need for scattering by atmospheric constituents.
- The ability to detect some entities where no other long-range methodology is available, including biological compounds, oil films on water, and the condition of different forms of vegetation.
- The ability to use spectral and temporal resolution to distinguish multiple species with overlapping fluorescence-excitation bands.

Disadvantages include these:

- Complex analysis when the fluorescence decay time is long, confounding range resolution.

- The typical "quenching" or low quantum efficiency of fluorescence for many species, reducing the return signals and the ultimate sensitivity. DIAL may be more sensitive in many cases.
- Typical fluorescence emission spectra are broad, requiring a matching broad detector response, leaving the system susceptible to background light, especially from the sun during daytime.
- Interference by Raman scattering from major atmospheric constituents.

Ways to mitigate some of the disadvantages of fluorescence lidar include the use of multiple or tunable excitation wavelengths, which can provide additional means to discriminate among multiple species as well as the strong Raman signals from the atmosphere. In addition, one can detect the fluorescence signal with an array of wavelength-dispersed detectors, to generate a single-shot spectrum for further analysis. In fact, with a tunable source and detector array, one can generate a hypercube of data to aid in unraveling the returns.

To enhance the signal strength from a given entity, it may be possible to inject fluorescent taggants into the atmosphere—basically, chemicals that selectively bind to the entity to be detected—and greatly enhance the return. For detection of entities on surfaces, many materials in solid or liquid form have much higher fluorescent quantum efficiencies than in gas form, since their fluorescence is not quenched by transferring energy to atmospheric molecules. In contrast to Raman scattering (with the exception of SERS), fluorescence-based sensing of entities on surfaces may provide a sensitive means for detection of small quantities of the desired species.

LIF lidar is possible for the gaseous forms of numerous aromatic hydrocarbons, which all have broad fluorescence spectra in the region of 250 to 400 nm, peaking around 290 nm. Atmospheric plumes (a few meters thick) of benzene, toluene, xylenes, naphthalene, and other aromatics have been remotely detected with LIF at less than ppm-m levels at significant ranges.[131] The BTEX compounds, which have subtle distinguishing fluorescence features, were identifiable individually, but some mixes (e.g., *p*-xylene, *o*-xylene, and *m*-xylene) are not readily quantitatively separable by remote sensing at a single wavelength. However, the more complex aromatics remotely detected (e.g., chlorobenzene, benzoyl chlorine, and piperdine) have prominent fluorescence features, simplifying discrimination. Models show that concentrations of multicompound mixtures could be determined at the 1-ppm-m sensitivity over ranges of 1-2 km for laser pulse energies on the order of 10 mJ.

Fluorescence of hydrocarbons in water is used in situ to detect low levels of contamination, such as from oil spills, and can be done remotely as well. One commercial system, the FLS Lidarsupplied by Laser Diagnostic Instruments in Estonia, has been adapted for airborne, downward-looking applications, particularly for mapping organic pollution (including oil and petroleum products) in water and on soil at altitudes between 50 and 500 meters.[132] The system employs a 308-nm XeClexcimer laser; and Figure 2-48 shows typical spectra from several different measurements over water.[133] Recent work has looked at the use of onshore or ship-based UV-LIF systems for a lower-cost approach to spill detection.[134]

[131] W. Eichenger, private communication on results from unpublished work in 1992 at Los Alamos National Laboratory is support of a DEA remote sensing for a drug interdiction demonstration.

[132] S. Babichenko, 2008, "Laser remote sensing of the European marine environment: LIF technology and applications," *Remote Sensing of the European Seas*, 189, Springer.

[133] See https://spie.org/x43265.xml?ArticleID=x43265.

[134] R. Karpicz, A. Dementjev, Z. Kuprionis, S. Pakalnis, R. Westphal, R. Reuter, and V. Gulbinas, 2006, "Oil spill fluorosensing lidar for inclined onshore or shipboard operation," *Appl. Opt.* 45: 6620.

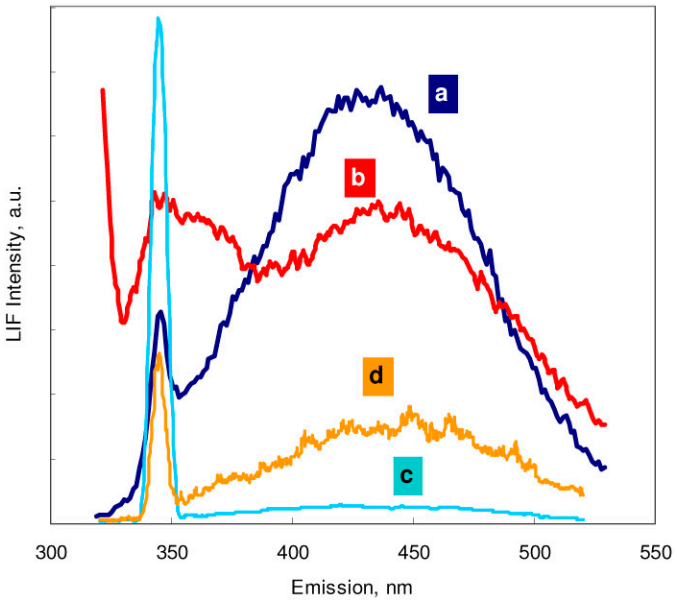

FIGURE 2-48 Sample fluorescence spectra generated by the airborne FLS (in arbitrary units, a.u.). Curves (a) and (b) represent clean Baltic Sea water and a sample polluted by ship motor oil (single laser shot, flight altitude 200 m). Curves (c) and (d) represent clean Atlantic Ocean water and a sample polluted by fuel oil (five combined laser shots, flight altitude 500 m). The peak at 344 nm is from Raman scattering by water. SOURCE: https://spie.org/x43265.xml?ArticleID=x43265.

LIF from surfaces has been applied to the study of vegetation to determine the presence of chlorophyll fluorescence from plankton, via 532-nm excitation[135] and, in another experiment, the physiological state (health) of trees in a forest through excitation of chlorophyll fluorescence with a 397-nm source and detection of fluorescence in the 450-800-nm region.[136]

Perhaps the most investigated LIF technology has been applied to the remote detection of biologically active aerosols, which could be deployed in biological warfare. As noted in the section "Aerosol Sensing," while aerosol lidar can detect the potential existence of a biologically active aerosol release, confirmation that it is truly a threat is best done through the fusion of data from multiple sensors. Biologically active molecules have inherent fluorescence (autofluorescence) resulting from several components: the amino acids tryptophan, tyrosine and phenylalanine; the reduced nicotinamide adenine dinucleotides (NADH, NADPH); and the flavin compounds.[137] Figure 2-49 shows the fluorescence spectra for each of the components, when driven with lasers at the appropriate peak excitation wavelength.[138] For local, or "point detectors" of biomolecules, one can employ multiple lasers to generate all of the fluorescing components and, by appropriate data analysis based on test aerosols, strive

[135] F.E. Hoge and R. N. Swift, 1981, "Airborne simultaneous spectroscopic detection of laser-induced water Raman backscatter and fluorescence from chlorophyll a and other naturally occurring pigments," *Appl. Opt.* 20: 3197.

[136] H. Edner, J. Johansson, S. Svanberg, and E. Wallinder, "Fluorescence lidar multicolor imaging of vegetation," *Appl. Opt.* 33, 2471 (1994).

[137] S.C. Hill, R.G. Pinnick, P. Nachman, G. Chen, R.K. Chang, M.W. Mayo, and G.L. Fernandez, "Aerosol-fluorescence spectrum analyzer: real-time measurement of emission spectra of airborne biological particles," *Appl. Opt.* 34, 7149 (1995).

[138] T.H. Jeys, W.D. Herzog, J.D. Hybl, R.N. Czerwinski, and A. Sanchez, "Advanced Trigger Development," *Lincoln Laboratory Journal* 17, 29 (2007).

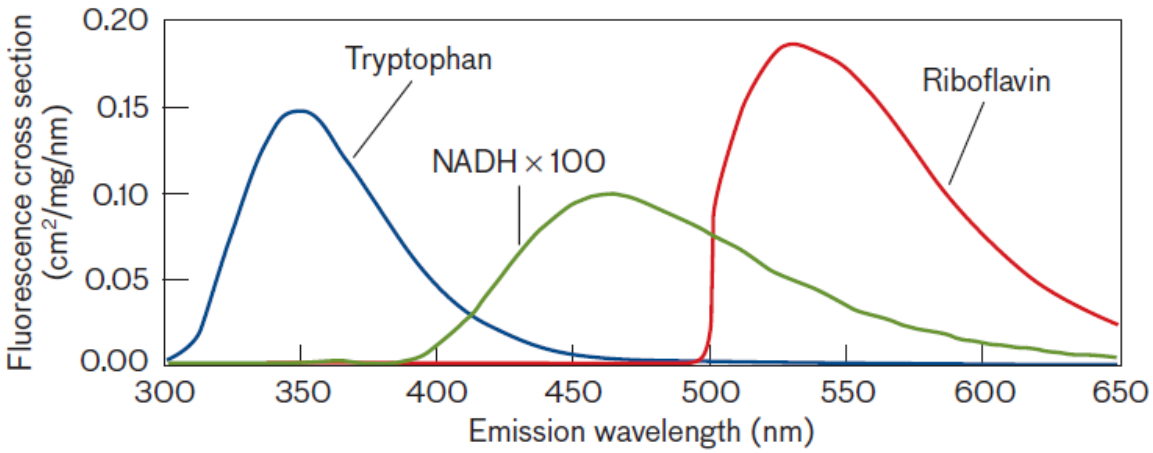

FIGURE 2-49 The fluorescence cross sections for tryptophan, NADH, and riboflavin are broad functions of emission wavelength for excitation at 280 nm, 340 nm, and 450 nm, respectively. SOURCE: T. Jeys, W. Herzog, J. Hybl, R. Czerwinski, and A. Sanchez, 2007, "Advanced trigger development", *Lincoln Laboratory Journal* 17(29).

to develop a sensor that can distinguish threats from benign biologically active or other species. Unfortunately, it has been found that there are wide variations in the intensity and spectra even for the same nominal bio-aerosol, and sensors may use other detection schemes in combination, such as LIBS, to provide a better detection probability.

For LIF lidar detection of bio-aerosols, tryptophan within the biological entities fluoresces and would seem to provide the best sensing. However, the excitation wavelength (<300 nm, and usually the fourth harmonic of Nd-doped lasers, around 260-266 nm) suffers significant atmospheric attenuation, especially under low-visibility conditions. In addition, as discussed above, use of the short wavelengths is likely to generate substantial interference signals from aromatic and other hydrocarbons that may be present in the atmosphere. NADH, while having weaker emission, can be excited by the Nd-doped laser third harmonic, which is better transmitted through the atmosphere and has been employed in several bio-aerosol LIF lidar systems. (While riboflavins might provide LIF, the signal from them does not appear to be significant in bio-aerosols, even if they have high fluorescence cross sections, and their emission spectra fall in a region where solar emission would produce major interference.)

Sandia National Laboratories (SNL) has developed the Ares lidar for bio-aerosol detection.[139] Figure 2-50 shows this system mounted in a van. The device employs a third-harmonic, pulsed, lamp-pumped Nd:YAG laser with a 10-ns pulsewidth. Schmitt et al. describes the specifics of

> Backscattered and fluorescent light collected by an 18.75-cm-diameter Maksutov telescope is collimated and directed to a long-pass dichroic beamsplitter, which reflects light with wavelengths shorter than ~360 nm, effectively separating the elastically backscattered laser light from the LIF light. The elastically scattered light is focused onto a photomultiplier tube (PMT), where it is detected and subsequently digitized to provide a record of the aerosol backscatter intensity as a function of range with 1.5 m resolution. Laser-induced fluorescence collected by the telescope passes through the long-pass filter and is focused onto the entrance pinhole of an imaging spectrometer. The LIF spectrum is detected by a gated, intensified charge coupled device (ICCD).

Figure 2-51 shows a data product from the ICCD detection channel with bio-aerosol fluorescence detected, along with atmospheric Raman components.[140]

[139] See http://www.sandia.gov/pcnsc/research/research-briefs/2004/Ares_Ulatraviolet_Laser_Induced_Fluorescence_(UV_LIF)_Standoff_System_Development_and_Testing_by_R._L._Schmitt,_et_al..pdf.

FIGURE 2-50 Photograph of the SNL-developed Ares aerosol and LIF lidar mounted in a van. SOURCE: http://www.sandia.gov/pcnsc/research/research-briefs/2004/Ares_Ulatraviolet_Laser_Induced_Fluorescence_(UV_LIF)_Standoff_System_Development_and_Testing_by_R._L._Schmitt,_et_al..pdf.

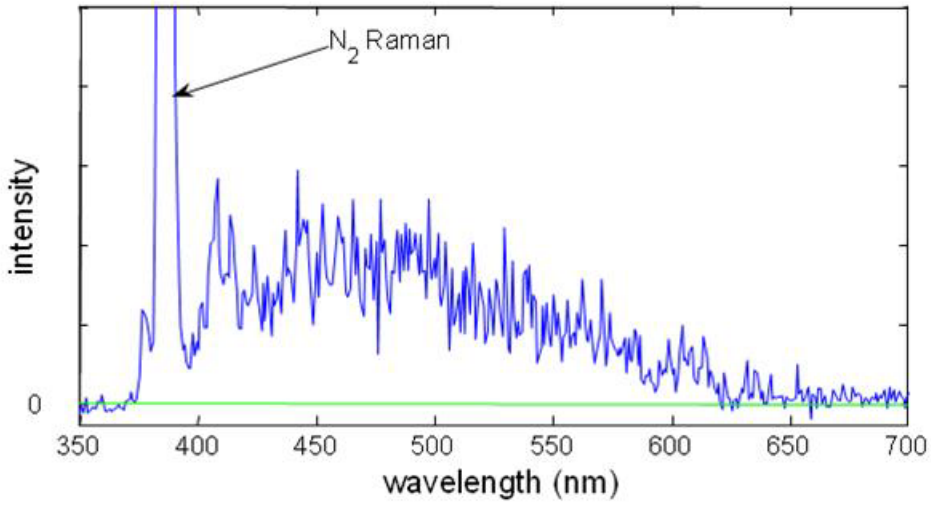

FIGURE 2-51 Bio-aerosol LIF signal detected by Ares LIF channel, with Raman signals from the atmosphere present as well. SOURCE: R. Schmitt and W. Seng, "UV Laser Induced Fluorescence Remote Sensing Technology and Applications," presentation to the committee, May 10, 2013.

[140] R. Schmitt and W. Seng, "UV Laser Induced Fluorescence Remote Sensing Technology and Applications," presentation to the committee, May 10, 2013.

FIGURE 2-52 Photograph of Canadian-developed, bio-aerosol sensor (SR-Biospectra) designed for 100-m ranges. SOURCE: http://www.ino.ca/media/134174/srbiospectra.pdf.

The Defence Research and Development Canada (DRDC) in Valcartier has also been developing LIF bio-aerosol detection systems,[141] including the Standoff Integrated Bio-aerosol Active Hyperspectral Detection (SINBAHD), a van-mounted system employing a 130-180 mJ/pulse, XeFexcimer laser at 351 nm with a 125-Hz pulse rate, 12-inch collecting optics, and a range-gated, intensified CCD covering 370-600 nm, with 5 nm spectral resolution. In addition, DRDC, working with INO in Quebec, has developed a short-range LIF lidar (SR-BioSpectra) designed to detect and classify bio-aerosol events in enclosed, semienclosed, and wide open spaces, over a 100-m range. A picture of that device appears in Figure 2-52, where the laser employed is a diode-pumped, frequency-tripled Nd:YAG laser.

Work on LIF-based bio-agent lidars in Europe is reported in Norway, the U.K., and Germany.[142] A French (CILAS)-led, European-funded consortium, Biological Optical Defence Experiment (BODE), developed a prototype short-range, 280- and 355-nm-based LIF lidar for civil applications.[143] A recent publication shows development, at least at the modeling stage, of a LIF-based lidar for biological warfare agent detection at the Beijing Institute of Technology in China,[144] while a recent Asian consortium employed a 355-nm LIF system to detect the presence of fluorescence from Asian dust and air-pollution aerosols transported from urban and industrial areas, indicating the scientific interest in LIF.[145]

[141] See http://cbdstconf.sainc.com/pdfs/Wednesday_5_0950_Simard.pdf.

[142] "Laser Based Stand-Off Detection of Biological Agents," NATO RTO Technical Report TR-SET-098, http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-SET-098///$$TR-SET-098-ALL.pdf.

[143] O. Meyer, C. Jacquelard, J.M. Melkonian, P. Chardard, P. Lanson, and D. Petitgas, 2010, "Stand-off biological detection by LIF (laser induced fluorescence) lidar," OPTRO 2010 – 4th International Symposium Optronics In Defence And Security, Paris, France, February 3-5.

[144] P. Liu, Y. Zhang, S. Chen, T. Lan, Y. Wang, Z. Qiu, W. Kong, and G. Ni, 2009, "Simulation of ultraviolet laser-induced fluorescence LIDAR for detecting bio-aerosol," *SPIE Proceedings* 7511, 2009 International Conference on Optical Instruments and Technology: Optoelectronic Measurement Technology and Systems, S. Ye, G. Zhang, and J. Ni, eds., 75111C.

[145] N. Sugimoto, Z. Huang, T. Nishizawa, I. Matsui, and B. Tatarov, 2012, "Fluorescence from atmospheric aerosols observed with a multi-channel lidar spectrometer," *Opt. Exp.* 20: 20800.

In Chapter 3 the use of femtosecond-duration, high-energy lasers to create atmospheric filaments for use in creating LIF from bio-aerosol clouds is discussed.

One variant of LIF lidar is photodissociation (PD)-LIF, followed by laser-induced fluorescence, which is being investigated for detection of trace amounts of solid explosives on surfaces.[146,147] The technique employs a UV laser, in the references cited operating at 236 nm, which dissociates the nitrogen-rich explosive materials on surfaces, generating NO molecules. These are then excited by the same laser pulse and generate fluorescence at a shorter wavelength than the exciting laser, in the 224-227-nm region. The short-wavelength LIF allows discrimination from long-wavelength fluorescence from other species as well as from atmospheric molecules, and the technique claims sensitivities high enough to be of practical use in, say, tracking sources of explosives. The main drawback for sensing is the poor atmospheric transmission at the wavelengths used, which may limit the working range to around 10 m.

In summary, operational LIF systems have been deployed to detect the presence of oil and other hydrocarbon pollutants on the oceans, as well as detect biologically active aerosols that could be deployed in biological-agent warfare. LIF can provide long-range detection and/or identification of some entities not easily sensed by any other techniques.

Conclusion 2-13: Laser-induced fluorescence can be used for range-resolved detection of a limited number of entities that exhibit fluorescence, and it is an important tool for detection of biologically active aerosols as well as small quantities of certain solids and liquids on surfaces.

WIND SENSING

Wind sensing makes use of Rayleigh-scattered returns from molecules, which become more intense as the wavelength decreases. If one employs a single-frequency source and examines the returned signal with GHz-level spectral resolution, the lidar returns at short wavelengths will show (Figure 2-53) a very narrow spectral line from the aerosol returns and a much broader linewidth from molecular Rayleigh scattering. In a normal atmosphere the broad linewidth is due to a mix of dephasing of the Rayleigh scattered light due to collisions with other molecules, leading to a Lorenztian lineshape, and the Doppler shift from molecular thermal motion, leading to a Gaussian lineshape. The Doppler shift is large due to the small mass of the molecules and their resultant high velocities. While aerosol particles are also in motion—primarily Brownian motion created by collisions with atmospheric molecules—the velocities are much slower and so is the resultant Doppler shift, leading to a spectrally narrow scattered return. As noted above, it is possible to employ narrow-linewidth spectral filters to distinguish aerosol from molecular scattering given the large difference in spectral character. One approach employs placing an etalon or a gas with a narrow absorption line in front of the detector to eliminate the aerosol signal.

The spectra shown in Figure 2-53 assume that there is no wind present in the volume probed by the laser. When there is, the bulk Doppler shift for both aerosols and molecules leads to a displacement in the center frequency of the returned signal, proportional to the Doppler shift along the direction of the beam, which can be either positive or negative depending on the wind direction. Lidar systems that can sense this shift can generate data on wind, a unique function available from no other remote-sensing system. Clearly, to generate full wind data the sensing beam must be scanned in different directions, since data from a single beam direction cannot separate out wind speed and wind direction.

Two types of lidar wind sensors can be built, one based on scattering from aerosols, the other from molecular species, and both are discussed in the subsections below.

[146] C.M. Wynn, S. Palmacci, R.R. Kunz, K. Clow, and M. Rothschild, 2008, "Detection of condensed-phase explosives via laser-induced vaporization, photodissociation, and resonant excitation," *Appl. Opt.* 47: 5767.

[147] C.M. Wynn, S. Palmacci, R.R. Kunz, and M. Rothschild, 2010, "Noncontact detection of homemade explosive constituents via photodissociation followed by laser-induced fluorescence," *Opt. Exp.* 18: 5399.

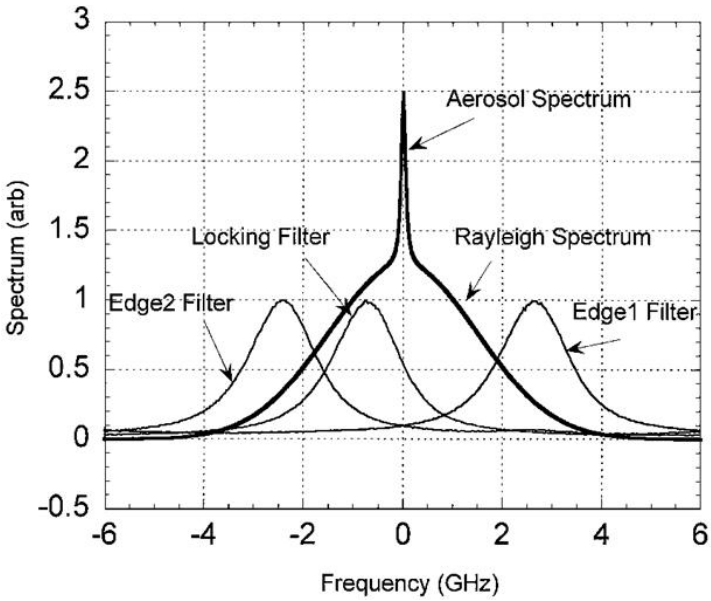

FIGURE 2-53 Spectra of the atmospheric backscattered Rayleigh and aerosol signals along with three etalon transmission functions used for incoherent wind-sensing lidars. SOURCE: B. M. Gentry, H. Chen and S.X. Li, 2000, "Wind measurements with 355-nm molecular Doppler lidar," *Opt. Lett*. 25: 1231.

Coherent Detection of Aerosol Scatter

In coherent wind-sensing (Doppler) lidar (Figure 2-54), narrow-spectral-bandwidth light is emitted from the transmitter laser and backscatters off aerosol particles in the atmosphere.

The relative speed of the particles with respect to the detector results in a Doppler shift, Δf, of the backscattered radiation compared to the transmitted radiation:

$$\Delta f = 2\, V_{rad} / \lambda_o$$

where V_{rad} is the radial component (along the line of site) of the velocity and λ_o is the transmitted wavelength. The backscattered radiation is mixed with a source (local oscillator) having a frequency near to that of the transmitted frequency. The mixing process leads to sum and difference frequencies, and for a Doppler lidar the difference (beat) frequency is detected and processed by RF electronics. As an example, for a 1,500-nm laser, the Doppler shift is 1.3 MHz for a 1 m/s velocity.

The majority of coherent wind sensors use pulsed lasers. A limit to the ability to resolve small velocities is the spectral width of the laser pulse. For a Gaussian-shape pulse the spectral width (full width at half maximum, FWHM) $\Delta\nu = 0.44/\Delta\tau$, where $\Delta\tau$ is the FWHM pulsewidth, independent of the laser wavelength. Thus, a 10-ns-duration pulse has a spectral width of 44 MHz and, for a 1,500-nm source it would be a challenge to resolve 1-m/s velocities from the signal returns. Typically, one would want to use pulses of several hundred nanoseconds to achieve this type of velocity resolution. There is also a trade-off between the laser pulsewidthand the range resolution Δr of the lidar via $\Delta r = \Delta\tau\, c\, /\, 2$, where c is the speed of light. For example, a 200-ns pulsewidth results in a 30-m range resolution. While a longer pulsewidth is indicated for more precise Doppler frequency measurement, increasing the pulsewidth proportionally decreases the range resolution. More sophisticated schemes can be used to

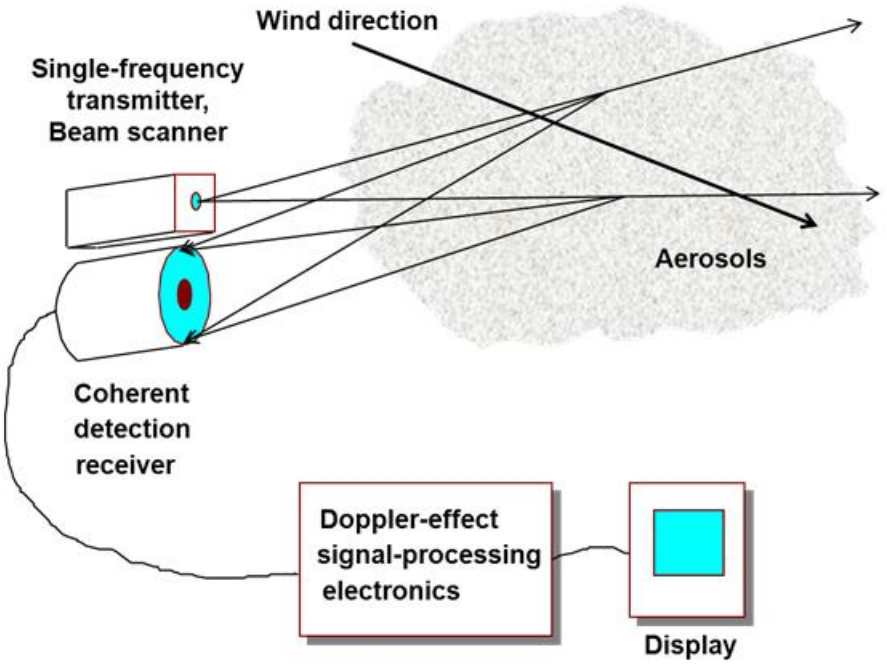

FIGURE 2-54 Schematic of coherent-detection wind-sensing (Doppler) lidar based on aerosol scattering.

accurately measure both position and velocity, such as using a series of laser pulses, called a pulse doublet or poly pulse measurement. They have a high time × bandwidth product, as will be discussed below.

To measure horizontal wind from a ground-based system, several strategies have been developed. In the velocity-azimuth Doppler (VAD) technique, the beam is directed at an angle from zenith and scanned 360 degrees in a cone. Returns are collected and the velocity components $\{u, v, w\}$ are calculated. Here, $\{u, v, w\}$ are the streamwise, cross-stream, and vertical components of the velocity. Another scan strategy is the range-height, or RHI, technique in which the beam is aligned with the average u vector and scanned in a small range of angles to probe the related range of heights downstream.

As is evident from the numbers, wind measurements require very-frequency-stable, narrow-linewidth laser sources. For a 1,500-nm laser, the center frequency is 2×10^{14} Hz, and a source with a 1-MHz linewidth must have a purity of 5×10^{-9}. Even if one locks the source laser to the local oscillator, the time delay in the returned signal places constraints on how much absolute drift of the operating frequency is allowable.

Historically, CO_2 gas lasers were first used to demonstrate coherent wind-sensing lidar systems, due to the relative ease of obtaining, for use in the local oscillator, stable, single-frequency operation from low-pressure, CW-discharge CO_2 lasers, given the narrow gain linewidth for the CO_2 vibrational-rotational transitions and the relatively minor perturbation of the laser resonator by the gas medium. The main challenge was to make stable optical cavities, which was accomplished through the use of good engineering design and the use of low-thermal-expansion materials such as Invar to set the cavity length. Unfortunately, as noted above, the aerosol backscatter level around the 10.6-μm wavelength of CO_2 lasers is low and varies widely with atmospheric conditions, requiring multi-Joule-level pulsed transmitters based on high-pressure, pulsed CO_2 lasers for even modest ranges.

Early solid state lasers were not good candidates for coherent lidars, owing to the considerably higher frequency instability from lamp-pumping-induced fluctuations in the optical path length of the laser medium compared to CO_2 lasers. With the development of diode-laser pump sources, it became

possible to make much more frequency-stable, single-frequency, low-power lasers used for the local-oscillator function, and through development of injection-seeding techniques based on the low-power lasers it became possible to make high-energy sources for use as the transmitter. Although the first solid state systems employed Nd:YAG lasers at 1,064 nm, concern over eye-safety has led to the recent development of sources based on Er-doped solid state lasers around 1,600 nm, and Tm and Ho-doped lasers for use around 2,000 nm. While fiber-format solid state lasers cannot yet generate the high-energy pulses needed for very long-range sensors, they have been employed for shorter-range systems and leverage the development of fiber components used for the telecom industry.

One of the most sought-after coherent wind sensors has been a space-based system that would provide global coverage and mapping of all atmospheric winds. Such a system would be of major value to science as well as of significant practical use for weather forecasting and commercial aviation routing, which could take advantage of favorable winds at different altitudes to speed up flights and save on fuel consumption. For military applications, such a system would provide greatly improved theater weather forecasts. It would also help with the determination of improved mission flying routes and support precision reconnaissance instruments and precise missile launch and reentry calculations, as well as improved ballistic wind estimates for conventional munitions.[148]

NASA began developing plans for space-based systems that would have required 20-J/pulse CO_2 lasers but in the 1990s moved to development of the 2,000-nm region with Ho-doped solid state lasers and now has system designs that can function from space with 250 mJ of energy.[149] At present, NASA is testing a diode-pumped, 10-Hz, 250-mJ, Tm, Ho:$LuLiF_4$ solid-state-laser-based system, the Doppler Aerosol Wind Lidar (DAWN) instrument, on aircraft platforms and has generated data over a 10-km vertical range from a DC-8, with a modest 15-cm-aperture receiver. Development of laser technology for space-based missions is in the planning stage. In the next section molecular-scattering-based wind sensing is discussed. This is also being developed for space-based sensors, generating data from atmospheric regions where aerosol densities are low.

Ground-based coherent wind sensors have been developed for a number of applications. Following several crashes of airplanes on takeoff or landing approaches, there was interest in fielding sensors to detect wind-shears around airports, particularly so-called "dry" wind shear not associated with rainstorms and thus undetectable by conventional radar systems. Other aviation applications have included detection of wake vortices generated by large aircraft on takeoff, a hazard to smaller aircraft in particular. Interest has lessened somewhat with improvements in air-traffic control procedures, but, in general, having a wind sensor as an adjunct to other meteorological sensors is a useful asset at an airport. (Attempts to put wind sensors directly on aircraft have met with resistance due to cost considerations, not the least of which is finding a location for the sensor.)

An example of a ground-based wind sensor for airport applications is the WindTracer system developed by Lockheed Martin Coherent Technologies. The device employs a pulsed, milliJoule-level solid state laser, which has evolved from a Tm-doped, 2,000-nm region laser to the present Er:YAG-based source at 1,617 nm. Specifications for the WindTracer product appear in Table 2-4, and a photograph of the shed-size unit is presented in Figure 2-55. Notable is the operating range of unit, which extends to as much as 33 km. According to the vendor, since the initial deployment in 2002, WindTracer successfully operates at Hong Kong, Tokyo, Osaka, London, New York, San Francisco, and Las Vegas airports, and is planned for installation at airports in Bangkok and Sicily. In addition, systems are in use at the Charles de Gaulle International Airport near Paris and at the Frankfurt International Airport, primarily for wake-vortex detection and ultimately determination of safe timing for departing aircraft. The goal is to attempt to speed up departure rates. A sample of the data product from the WindTracer system appears in Figure 2-56.

[148] S.B. Alejandro, "Application of LIDAR Against Soft Targets," presentation to the committee, March 5, 2013.

[149] U.N. Singh, "Emerging and enabling lidar technologies and techniques for NASA's space-based active optical remote sensing missions," presentation to the committee, April 17, 2013.

TABLE 2-4 Specifications for WindTracer Doppler Wind Sensor.

Measurement	
Typical range	400 m to 18 km
Maximum range	33 km
Radial wind velocity range	±35 m/s
Minimum range resolution	100 m
Scanner	
Azimuth range	0 to 360 degrees
Elevation range	-5 to 185 degrees
Resolution	0.001 degrees
Pointing accuracy	±0.1 degrees
Optical clear aperture	12 cm
Transceiver	
Laser wavelength	1,617 nm
Pulse energy	2.5 mJ ±0.5 mJ
Pulse duration	300 ns ±150 ns
Pulse repetition frequency	750 Hz
Beam diameter	9.6 cm (e^{-1} intensity width)
Shelter	
Environment	All weather
Weight	2600 kg
Dimensions	197 × 244 × 329 (h) cm (clearance with lightning rods)
Power specification	200-240 VAC single phase, 50 or 60 Hz (specified at time of purchase) 50 A service required.

SOURCE: See © 2013 Lockheed Martin Corporation. All Rights Reserved.

FIGURE 2-55 WindTracer system. SOURCE: © 2013 Lockheed Martin Corporation. All Rights Reserved.

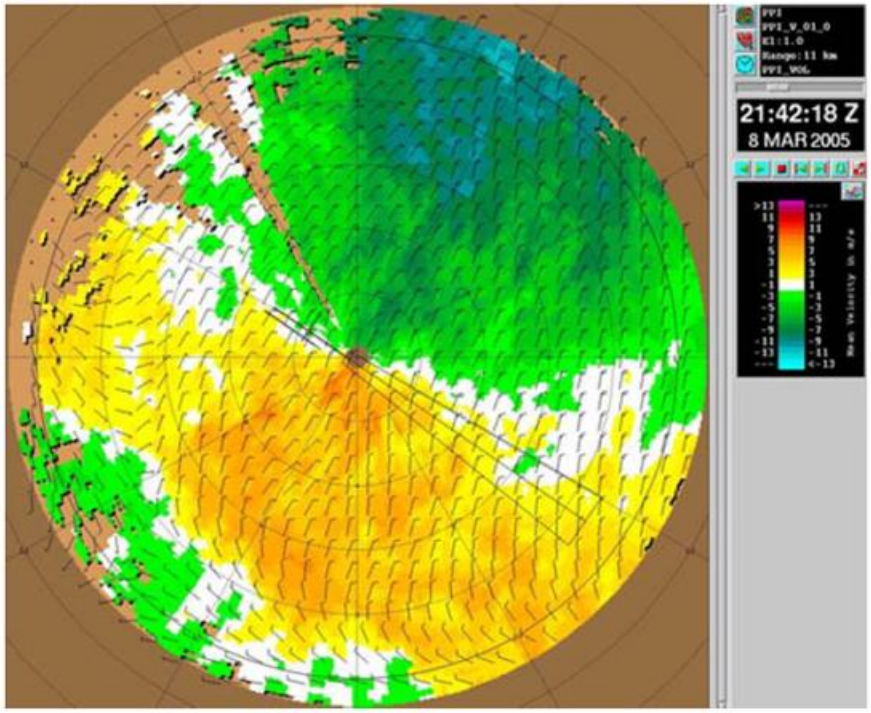

FIGURE 2-56 Sample WindTracer color-coded radial wind velocity map at a 1-degree elevation tilt and out to a range of ~11 km. Overlaid on the radial velocities are wind barbs depicting the vector velocity field. SOURCE: S. Hannon, K. Barr, J. Novotny A. Oliver, J. Bass, and M. Anderson, "Large scale wind resource mapping using a state-of-the-art 3-D scanning lidar," *WINDPOWER 2008*, Houston, June 1-4, 2008, Available at http://www.res-americas.com/media/918770/awea-2008_large-scale-wind-resource-mapping-using-a-state-of-the-art-3-D-scanning-lidar.pdf.

Mitsubishi Electric (Japan) claims manufacture of a similar-sized Doppler wind sensor, employing a 1,500-nm-region pulsed solid state laser[150] and lists a 20-km horizontal range. Based on a recent publication, the laser final stage may use a state-of-the-art, Yb, Er:glass planar waveguide amplifier, which has been shown capable of generating an output power of 3.3 kW with a pulse width of 580 ns, 1.9 mJ of pulse energy, a repetition rate of 4 kHz, and a high beam quality, with an M^2 factor of 1.3.[151] Applications of the device are listed as follows:

- Meteorological measurements (3-D profile of atmospheric wind, turbulence).
- Air traffic safety (low-level wind shear around airports, sudden gusts, and aircraft wake turbulence).
- Air environment consultation (urban heating, monitoring of substances released in emissions from factories, application in wind power generation).

[150] See http://www.mitsubishielectric.com/bu/lidar/products/coherent/index.html.

[151] S. Kameyama, T. Yanagisawa, T. Ando, T. Sakimura, H. Tanaka, M. Furuta, and Y. Hirano, 2013, "Development of wind sensing coherent Doppler lidar at Mitsubishi Electric Corporation from late 1990s to 2013," CLRC 2013, 17th Coherent Laser Radar Conference, Barcelona, Spain, June 17-20. Conference papers are available at http://www.tsc.upc.edu/clrc/.

(a) (b)

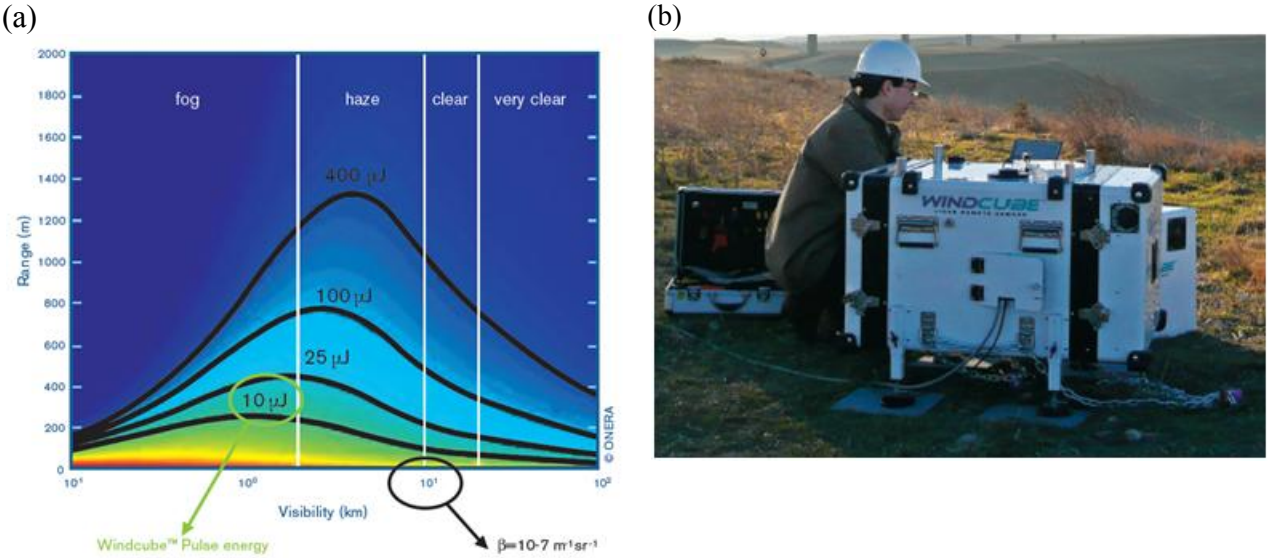

FIGURE 2-57 (a) Calculations of operating range for Doppler wind-sensing lidars as a function of visibility, with 1,550-nm laser pulse energy as a parameter. (b) Leosphere system in field installation at wind-turbine farm. SOURCE: (a) See http://esamultimedia.esa.int/docs/technologytransfer/Leosphere-Windcube.pdf. (b) See http://www.upwind.eu/media/576/D6.1.1.pdf.

The emergence of a significant number of wind-turbine power systems has led to the development of a number of commercial short-range wind sensors. The economics of first determining good locations for turbines and then developing sensors that provide look-ahead data for turbine control justifies expenditures for laser-based systems rather than simple anemometers. The devices tend to be based on 1,550-nm region, all-Er:fiber architectures, given that only 0.1- to 1-km ranges are needed, and systems are advertised from Mitsubishi Electric, HALO Photonics (U.K.), Leosphere (France), and ZephIR (U.K.).

Leosphere employs technologies initially developed at the Office National d'Etudes et RecherchesAérospatiales (Onera), the French national aerospace research center. Figure 2-57a shows data developed by Onera for a 1,550-nm based system with a 30-m range resolution, showing operating range under different atmospheric visibility conditions, with laser pulse energy as a parameter.[152] The LeosphereWindcube product for wind-turbine power generation facilities runs with 10-µJ (20-kHz-rate, 200-ns-duration) pulse energies and provides condition-dependent ranges of 40-200 m.[153] A photograph of the device appears in Figure 2-57b. The unit emits four separate beams at an angle to the vertical, spaced 90 degrees apart to develop wind-direction data.

The ZephIR wind sensor is unique since, in contrast to the sensors just discussed, it employs a CW, 1,550-nm laser provided by NKT Photonics (Denmark). The source is based on a distributed feedback (DFB) single-frequency Er:fiber laser, amplified to approximately a 1-W power level.[154] According to the ZephIR manufacturer,[155] the device measures wind speeds at five different heights at altitudes in the 10-200 m range. The beam is projected at 30 degrees to the vertical and rotated to form a cone. The beam is focused to different heights in that cone and measures 50 points of data every second at each height. Speed measurements fall in the <1 to 70 m/s range, with <0.5 percent speed accuracy

[152] See http://esamultimedia.esa.int/docs/technologytransfer/Leosphere-Windcube.pdf.
[153] Ibid.
[154] See http://esamultimedia.esa.int/docs/technologytransfer/Leosphere-Windcube.pdf.
[155] "Fibre lasers stretch from cells to wind farms," *Optics and Lasers Europe*, March 2009, 15, optics.org/ole.

FIGURE 2-58 The ZephIR unit in the field. SOURCE: ZephIR 300 installed, image courtesy of ZephIRLidar.

variation and <0.5 degrees of direction accuracy variation. The manufacturer claims that "ZephIR has also now amassed more than 3 million hours of operation across 650+ deployments globally spanning a decade of commercial experience."[156] A photograph of the trash-can-sized unit appears in Figure 2-58.

Conclusion 2-14: Wind sensor system complexity has been an issue for long-range systems but has been addressed and solved for short-range sensors, based on a market pull.

Conclusion 2-15: Any country that develops a space-based wind sensor could use it for commercial and military advantages.

Incoherent Detection of Molecular Scatter

Figure 2-59 shows a schematic of an incoherent-detection, wind-sensing lidar system. In contrast to the coherent-detection system, the device employs narrow-line filters and multiple detector channels to determine the shift in the center wavelength of the scattered light. Figure 2-55 shows the positioning and response of three band-pass filters (Edge1, Edge2, and Locking Filters) employed in the so-called double-edge detection technique, which has evolved from the original edge-detection concepts for incoherent-detection wind sensing.[157,158,159]According to Gentry, Chen, and Li,

> The separation of the two edge filters is chosen to be 5.1 GHz, so the sensitivity of the broader molecular signal is equal to that of the narrower aerosol signal. The locking etalon peak is located 1.7 GHz from the Edge1 filter peak, so the crossover point of the two edge-filter band-passes is aligned to the half-height point of the Locking Filter band-pass. Actively locking the laser and the etalon at this point on the Locking Filter ensures symmetry of the edge-filter channels for wind measurement.

---

[156]Fred Olson Ltd. Ten years at the top, measuring right to the top, with ZephIR Lidart. http://www.fredolsen.co.uk/news/ten-years-top-measuring-right-top-zephir-lidar. Accessed on March 14, 2014.

[157] C.L. Korb, B.M. Gentry, and C.Y. Weng," Edge technique: Theory and application to the lidar measurement of atmospheric wind," *Appl. Opt.* 31 4202 (1992).

[158] C.L. Korb, B.M. Gentry, S.X. Li, and C. Flesia," Theory of the double-edge technique for Doppler lidar wind measurement," *Appl. Opt.* 37 3097 (1998).

[159] B.M. Gentry, H. Chen, and S.X. Li, "Wind measurements with 355-nm molecular Doppler lidar," *Opt. Lett.* 25, 1231 (2000).

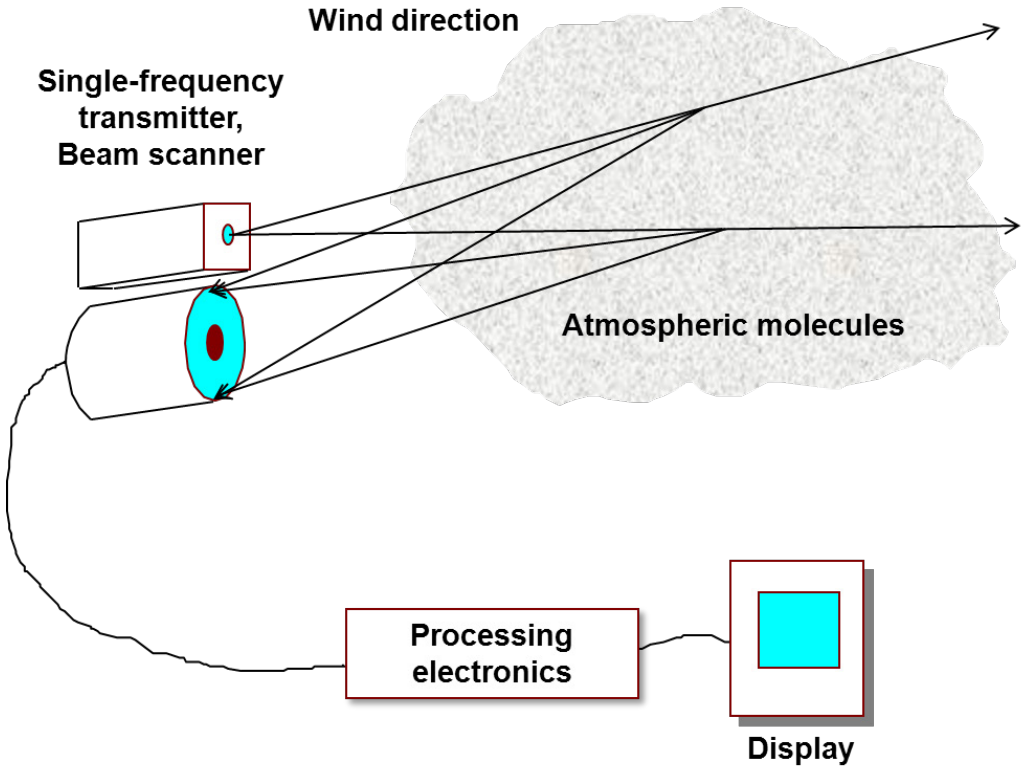

FIGURE 2-59 Schematic of incoherent-detection wind-sensing (Doppler) lidar based on molecular scattering.

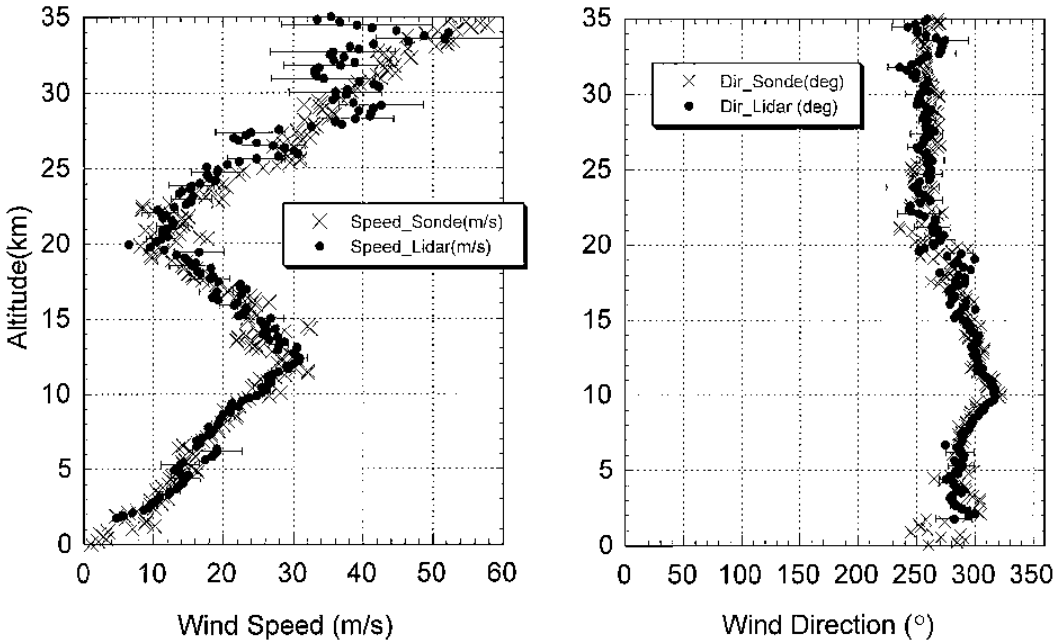

FIGURE 2-60 Profiles of wind speed and direction measured by the molecular lidar system compared with data from collocated balloons on November 17, 1999. SOURCE: B. M. Gentry, H. Chen, and S. X. Li, 2000, "Wind measurements with 355-nm molecular Doppler lidar," *Opt. Lett.* 25: 1231.

Differential comparison of the signals from the two-edge filters allows determination of the Doppler shift from the backscattered signal. Although, in principal, one could apply the edge technique to sense aerosol Doppler shifts, current efforts center on the use of the scheme for molecular backscatter, since it provides data for regions of the atmosphere having low aerosol content and thus not suited for coherent sensing.

Figure 2-60 shows validation data[160] taken by NASA from a ground-based, double-edge lidar that used a 45-degree elevation and four azimuthal angles for wind-direction determination. The system employed a single-frequency, tripled Nd:YAG laser that produced 15-ns pulses at a 10-Hz repetition rate, with 70 mJ of energy and a 80-MHz linewidth. The receiver telescope had a 45-cm aperture. Lidar data were compared with a conventional balloon-borne wind sensor and show excellent agreement in terms of both wind velocity and direction, with evident larger scatter at the highest altitudes.

Currently, NASA is fielding the TWiLiTE (Tropospheric Wind Lidar Technology Experiment) designed for airborne platforms, and flight tests have been carried out on DC-8, WB57, and ER-2 aircraft, with plans for long-duration tests on a Global Hawk UAV.[161]

The European Space Agency (ESA) has been funding development of the Atmospheric Dynamics Mission (ADM-Aeolus) effort, to employ an incoherent wind-sensing lidar (Aladin) to sense the lowermost 30 km of the atmosphere. The system was originally set for launch in October 2007 but experienced significant issues with performance of the 355-nm laser due to optics contamination and damage issues, and the launch date is now set for 2015.

An alternative approach to incoherent wind sensing is under NASA-funded development, the Optical Autocovariance Wind Lidar (OAWL),[162] which uses a double-edge sensor for the molecular-backscatter channel but is able to sense the Doppler shift of the aerosol channel through the use of a high-spectral-resolution interferometer in another detection channel. The advantage of the technology is that one sensor, at one wavelength, would be able to carry out the entire wind-sensing mission from space. Ground-based testing validation of the technology has been carried out, and aircraft-based measurements are in the planning stages.

COMMERCIAL LASER/LADAR PRODUCTS

As discussed in this chapter, there is a wide variety of uses for ladars, ranging from mapping and imaging to velocimetry.[163] While there is no rigid dividing line between scientific, military, and commercial applications, the first two domains are characterized by small numbers of high-performance devices, while the third is performance-constrained by marketplace affordability. Nonetheless, the compelling advantages of ladar in terms of distance and velocity measurement have created an active marketplace with a variety of products in two principal application areas: (1) speed detection for law enforcement and (2) distance sensing for positioning and obstacle avoidance.

Police use ladar because it precisely measures target distance and velocity within a small field of view. For example, the illuminator (laser) can be used to select a specific vehicle (even in traffic) from which reflections can be measured. The devices (one is illustrated in Figure 2-61) have measurement filters to overcome errors introduced by surface irregularities such as mirrors and may rely on features on license plates (e.g., retroreflective backing materials) intended to support their use. The police ladar

[160] B.M. Gentry, H. Chen, and S.X. Li, 2000, "Wind measurements with 355-nm molecular Doppler lidar," *Opt. Lett.* 25: 1231.

[161] See http://twilite.gsfc.nasa.gov/index.php.

[162] C.J. Grund, S. Tucker, R. Pierce, M. Ostasziewski, K. Kanizay, D. Demara, and J. Howell, 2010, "Development and demonstration of an optical autocovariance direct detection wind lidar," Paper B3P5, 2010 Earth Science Technology Forum (ESTF2010), June 22-24, Arlington, Virginia. http://esto.nasa.gov/conferences/estf2010/papers/Grund_Chris_ESTF2010.pdf.

[163] See http://www.lidar-uk.com/usag EOf-lidar/ for a listing.

devices are very affordable. Online auction sites have them for $1,500-2,000. Interestingly, the devices have stimulated active countermeasures by motorists, such as so-called "ladar jammers" which emit light in the same frequency range as the speed guns. The effectiveness of these countermeasures is unclear.

The marketplace for the second application area, distance sensing, also seems primarily oriented toward automotive applications. In this case, rather than law enforcement, the primary motivations are safety and navigation. The devices are characterized by multiple lasers and detectors organized in an array format to achieve a wide field of view; the array can be cyclically rotated to achieve 360 degree swaths of real-time information.

These devices can in principle be used to provide greater situational awareness or even driver assistance for a human driver. A more intriguing application with substantial commercial investment is underway that has a potentially much larger market than speed guns: driverless automobiles. In this application, the ladar serves as part of a package of sensory apparatus needed for the autonomous vehicle to avoid stationary obstacles as well as moving objects such as pedestrians and other vehicles.

The sensors used in this application (e.g., for the Stanley[164] self-driving car, which won the 2005 DARPA Grand Challenge) include those from SICK, examples of which are shown in Figure 2-62 were also used atop an autonomous vehicle that completed the DARPA Urban Challenge in 2007.

FIGURE 2-61 UltraLyte LTI 20-20 speed enforcement LiDAR. SOURCE: http://www.lasertech.com/UltraLyte-Laser-Speed-Guns.aspx.

[164] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, et al., 2006, "Winning the DARPA Grand Challenge," *Journal of Field Robotics*.

FIGURE 2-62 Ladars from SICK atop "Little Ben," an experimental self-driving car developed for the 2007 DARPA Urban Challenge. Credited to Lehigh University.

The ladar type highlighted by the ovals in Figure 2-62 is intended for short-range (e.g., 20-50 m) outdoor operation. It uses a 905 nm light source, has a 270 degree field of view, fog correction, a 25-50 Hz scanning frequency, and an operating range (depending on model) of 20-50 m, with an error of about 1 cm. Retail price ranges from $4,000 to $6,000.

Significant indicators would include dramatic reductions in cost and/or performance. The thousand dollar range of commercial ladar products listed above does not cover integration into a system vehicle. The integration is a barrier to entry but one that many are striving to overcome commercially. The achievable performance could be improved by signal processing methodologies such as algorithmic advances or advances in signal processing hardware such as a higher degree of parallelism or integration. Any advances that are applicable in the commercial lidar domain are equally useful in more specialized systems, such as those for defense applications. Also, improvements in size, weight, and power are generally indicative of manufacturing capabilities that are applicable across technology applications, including defense.

Incentives for the commercial lidar technologies include (for the producer) general advantages that stem from active and profitable instrument manufacturing capabilities, including the ability to allocate some profits to further research and development. Disincentives for the commercial ladar technologies include trade-offs that come from commercialization, such as trading away performance to hit an attractive price point, or overoptimizing for a particular problem domain such as speed detection. Improvements in laser range finder technologies will enhance advanced manufacturing capabilities such as autonomous, flexible manufacturing and sensor-actuator models.

The key technology improvements will increase performance (including measurement accuracies, scan rates, and system reliability) while decreasing cost. The likely means of achieving desirable new positions in the cost/performance trade space is integration (and the associated miniaturization) of components rather than a particular component technology such as a better laser, detector, or processor. Experience with both applications of the technologies and the processes used to manufacture them will present opportunities for technology advances.

There are significant foreign capabilities in commercial ladar. Vendors exist abroad, such as SICK in Germany and Hokuyo in Japan as well as in the United States (e.g., Velodyne).

Conclusion 2-16: Commercial applications for active EO sensing will become widespread, dramatically increasing the use of active EO sensing and significantly lowering the cost.

The performance and precision of active EO sensors are compelling, but there is currently a cost trade-off that may limit commercial penetration in very cost-sensitive settings. For example, ultrasonic sensors that aid in automobile parking exist and only cost a few dollars; an active EO sensor in this application could be considered overkill. However, the importance of the human lives involved in autonomous vehicles will put pressure on the sensor technologies for accuracy, and with sensor fusion, diversity and redundancy. The "overkill" of an active EO sensor could in fact be a desirable attribute if the sensor is multifunctional and can be made cheaply.

Conclusion 2-17: Robotics—for example, autonomously navigated vehicles—is likely to be a dominant commercial and military application for active EO sensing.

Conclusion 2-18: Active EO sensing is poised to significantly alter the balance in commercial, military, and intelligence operations, as radar has done over the past seven decades.

Conclusion 2-19: Most active EO sensors used for military and security application will fall into one of two categories: (1) adaptation of inexpensive commercial devices to military application or (2) development of unique ladar systems for the military. The first category is accessible to nonstate actors and resource-poor countries. The second category is only accessible to countries with significant resources.

Conclusion 2-20: Development programs for high-end active EO sensors will involve observable activities or indicators of the direction these programs are taking.

Conclusion 2-21: Widespread, commercially available active EO technology offers a low barrier to entry for asymmetric adversaries with limited resources as well as to those seeking to adapt the technology for large-volume applications.

Recommendation 2-1: Analysts assessing states of technological development and projected timelines for developments of military significance should pay close attention to activity in the commercial sector.

3

Emerging Electro-Optical Technologies

This chapter focuses on emerging active electro-optical (EO) technologies that are rapidly developing and whose implementation is still evolving. The focus is on several coherent systems such as temporal and spatial heterodyning, synthetic aperture ladar, multiple-input, multiple-output (MIMO) imaging, and speckle imaging. Also discussed are emerging approaches in femtosecond sources and quantum technologies. Although these technologies are not fully matured, they may have a significant impact on the field of active EO sensing in the next 5-10 years and beyond.

MULTIWAVELENGTH LADAR

Color can be a very useful discriminant. People experience this when they compare black and white pictures to color pictures. Color distinctions are based on the difference in reflectivity versus wavelength. Active multispectral EO can complement conventional ladars when viewing solid targets by adding additional surface material discrimination information not available with just 2-D or 3-D imaging. Active multispectral imaging for targets can have an advantage over passive imaging of targets because one can control the illumination source. Therefore even at night near-IR (NIR) wavelengths can be used, whereas there would not be enough signal to use passive multispectral at these wavelengths. An active EO multispectral sensor will combine the benefits of conventional ladar and multispectral wavelengths in a band that has significant color variation in its reflectance. Conventional ladars, and some passive imaging sensors, utilize the shape of a target for detection and/or identification. Two different approaches have been used to deploy active multispectral sensors against hard targets. One is to use the laser wavelengths that are easy to generate, for example 1.064 μm and around 1.5 μm, and take whatever recognition benefit one can gain. The second approach is to determine what wavelengths offer the best active multispectral recognition probabilities and make the lasers appropriate to these discriminants. In the second case, a class separability metric is formulated and optimal wavelengths are selected.[1] Figure 3-1 shows reflectivity versus wavelength and angle for six different materials. Figure 3-2 shows reflectivity versus wavelength for leaves, showing significant spectral reflectivity changes versus wavelength in the NIR.

Multispectral and hyperspectral sensing depend on variation in reflectivity versus wavelength for surface materials of the object being viewed. The main fundamental limit is that reflectance, or absorption, reflects only the surface properties of a solid object being viewed.

The key technologies for active multispectral imaging are the illuminator, a multispectral laser to illuminate an object at more than one wavelength, and the associated detector technologies.

To achieve active multispectral imaging, one needs laser sources that cover all of the wavelengths of interest. Therefore, developing active multispectral sensing requires that laser sources either have multiple laser lines or be very broadband.

[1] M. Vaidyanathan, T.P. Grayson, R.C. Hardie, L.E. Myers, and P.F. McManamon, 1997, "Multispectral laser radar development and target characterization," *Proc. SPIE*, 3065: 255.

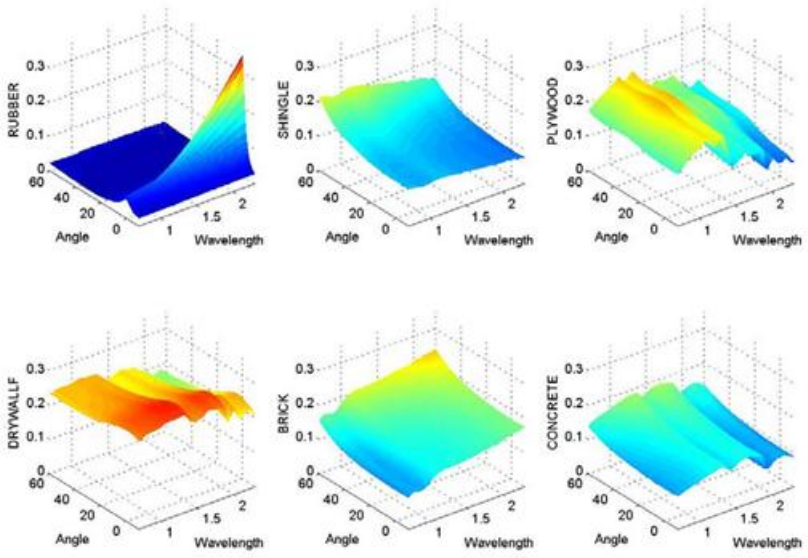

FIGURE 3-1 Reflectivity versus wavelength and angle for six different materials. SOURCE: D.G. Jones, D.H. Goldstein, and J.C. Spaulding, 2006, "Reflective and polarimetric characteristics of urban materials," *AFRL Tech Report*, AFRL-MN-EG-TP-2006-7413.

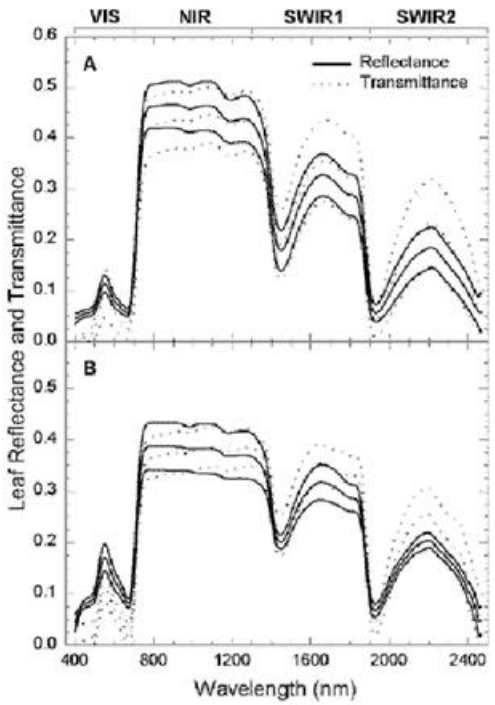

FIGURE 3-2 Leaves from trees have different spectral qualities. Plot shows leaf reflectivity and transmission. SOURCE: Reprinted from Remote Sensing of Environment, 64/3, G.P. Asner, Biophysical and Biochemical Sources of Variability in Canopy Reflectance, 234-253, 1998, with permission from Elsevier.

As an active system, detection of NIR spectral bands is made possible even at night, and NIR bands have a significant amount of reflectivity variation, or color. This sensor type will be especially useful in high clutter situations or situations where the target is partially obscured.

Active multispectral sensing does not require precision angle/angle resolution because the surface material discrimination is based on the ratio of reflectance at various wavelengths, not on the shape of an object. Therefore, to the first order, active multispectral sensing is not dependent on the diffraction limit, so it can be a useful long-range discriminant. It is also not dependent on exact object shape, which explains its usefulness when a target is partially obscured. The only angular size effect is based on color mixing over a pixel when pixels become larger.

One disadvantage is that active multispectral requires laser illumination at all wavelengths being used. Three bands would require three lasers, or the ability to divide a single laser into sources at each of the three wavelengths, or else sequencing through the bands. An optical parametric oscillator (OPO) can be used to shift the wavelengths transmitted over time, as long as the object being viewed is stationary over that time period. Alternatively, a broadband laser source containing many wavelengths can be used. A broadband laser has the disadvantage of spreading the laser light over that broadband, reducing the available light at any particular wavelength.

Active multispectral EO sensing containing multiple lasers or a single laser shifted to multiple wavelengths, has a low scientific barrier to entry. Many countries have excellent laser technology, which is the first requirement to manufacture multiline or broadband lasers. As a result, the comparative state of the art in this area is not very meaningful. Once it is decided to provide laser sources at multiple wavelengths, it is relatively straightforward to make an active multispectral sensor. That said, development of this technology also requires investment in the associated receivers and their integration into a single unit that meets constraints of size, weight, and power (SWaP) and cost.

TEMPORAL HETERODYNE DETECTION: STRONG LOCAL OSCILLATOR

As described in Chapter 2, laser radar systems using direct detection (as in 3-D flash imaging) can be limited by the noise in the detector. One method for reducing the effect of detector noise was taken from the radio frequency (RF) community. Heterodyne detection is a method originally developed in the RF domain as a way to convert a received signal to a fixed intermediate frequency that can be more conveniently—and more noiselessly—processed than the original carrier frequency. This practice has been similarly adapted to optics as a way to detect weak signals.

In optical heterodyne detection, a weak input laser signal is mixed with a local oscillator (LO) laser wave by simultaneously illuminating a detector with both signals. For temporal heterodyne it is very important to match the illumination angles of both the LO light and the return signal light across the detector, or else spatial fringes develop. High spatial frequency fringes smaller than a detector can average the interfering signal to zero across the detector. Figure 3-3 illustrates the arrangement of a simple optical heterodyne receiver. A laser transmits a coherent waveform of light toward a target. The reflected light beam is mixed with the reference laser (local oscillator) beam at a beam splitter in the receive optical path and the beams are superimposed on the detector. The resulting photocurrent is proportional to the total optical intensity, which is the square of the total electric field amplitude. If the LO power is increased above all other noise sources, the signal-to-noise ratio becomes limited only by shot noise of the return signal. For temporal heterodyne detection, the reference laser frequency is offset from the laser source by $\omega_{if.}$ and the resulting optical intensity has fluctuations at the difference and sum frequencies of the two fields and at double the frequency of each of the fields. The LO frequency is offset so it is possible to determine whether any velocity is toward or away from the sensor. The coherent

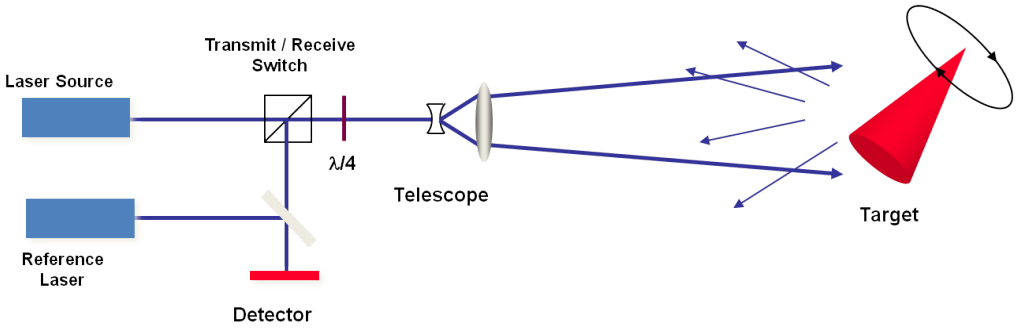

FIGURE 3-3 Simple heterodyne (coherent) laser radar configuration.

receiver is usually designed to isolate the difference frequency component from fluctuations and noise at other frequencies.[2]

For traditional heterodyne detection, the LO field strength must be much higher than the return signal strength in order to mitigate the effects of various noise sources. However, if the detector is sufficiently sensitive, there is little need to mitigate detector noise sources by using a strong LO. If the field strengths are similar, the noise mitigation benefit of a strong LO is lost, but the frequency comparison and narrowband filtering features are still met.

TEMPORAL HETERODYNE DETECTION: WEAK LOCAL OSCILLATOR

Traditional heterodyne detection with a strong LO has an added challenge if arrays of detectors are desired. While a high-resistance receiver, such as a focal plane array with relatively low bandwidth, can operate with low LO power, other standard high-bandwidth IR detectors used in heterodyne detection systems, such as a linear GHz-bandwidth photodiode, can require as much as 1 mW of LO power or more to reach the shot noise limit. This may result in unacceptable heat loads of greater than 10 W for large arrays of 10,000 elements.[3] A 256 × 256 array would require even larger LO powers. Using a photon counting detector allows use of a low-power LO and potentially the ability to use the same detector for both coherent and direct detection. In order to use photon counting detectors such as Geiger-mode (GM) avalanche photodiodes (APDs), the LO strength must be matched to the signal strength. In this case, the noise mitigation effects from the strong LO are lost, but the ability to detect frequency shifts is maintained.

The block diagram for weak LO heterodyne detection is the same as that shown in Figure 3-3, with the detector being replaced by a GM-APD array. A laser transmits a coherent waveform toward a target, the reflected beam is mixed with the LO at a beam splitter in the receive optical path, and the beat signal is detected by the receiver. The object is imaged onto the detector array, but in this case the readout is simply the photon arrival time. The size of the angle/angle resolution element depends on the detector angular subtense and the size of the focused optical spot. To detect more than one photon per angle/angle resolution element per reset of the detector, the receive optics can be constructed so that each focused optical spot is spread across a group of pixels (called a macropixel).[4] Each macropixel then acts as a photon-number-resolving detector whose dynamic range is equal to the number of pixels contained in the macropixel. For example, a 32 × 32 GM-APD can be broken up into an 8 × 8 array of photon-

[2] P. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51(6): 060901.

[3] L. Jiang and J. Luu, "Heterodyne detection with a weak local oscillator," *Applied Optics* 47(10), 1486-1503, (2008).

[4] L. Jiang et al., 2007, "Photon-number resolving detector with 10-bits resolution," *Phys. Rev. A*, 75(6): 062325.

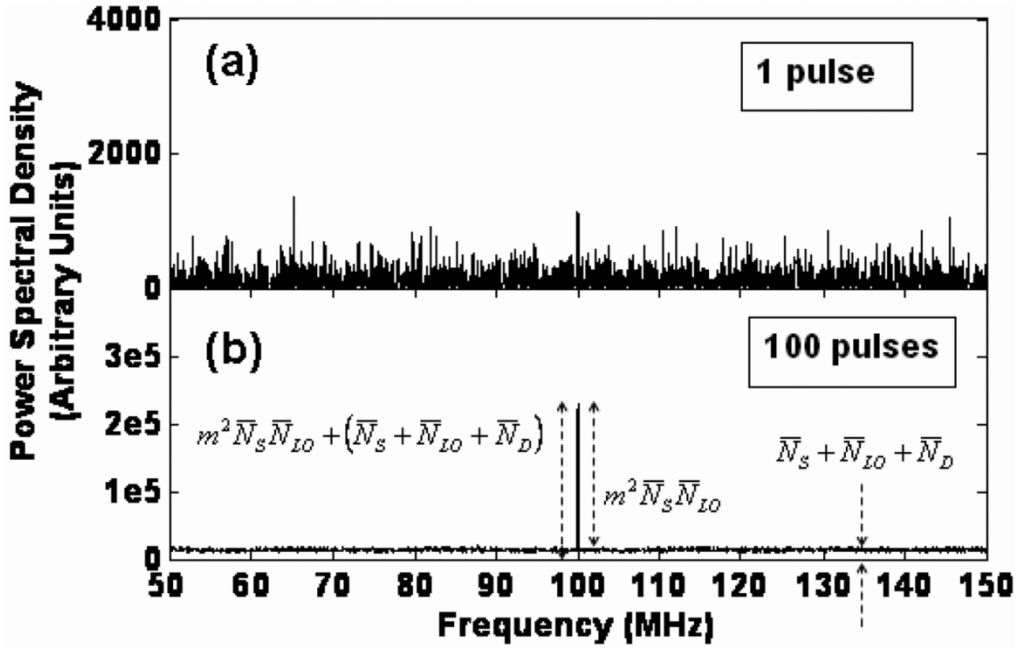

FIGURE 3-4 Power spectral density (PSD) of the detected current for a temporal heterodyne laser radar receiver, shown for (a) 1 pulse and (b) 100 pulses of incoherent averaging. SOURCE: L. Jiang, E. Dauler, and J. Chang, 2007, "Photon-number-resolving detector with 10 bits of resolution," *Physical Review A* 75 (6):062325.

number-resolving detector macropixels. Each macropixel can be a 4 × 4 array of pixels, or 16 pixels, resulting in a 4-bit dynamic range. When plotted as a function of time for a single macropixel, the photon arrival times map out the frequency of the beat signal. The detector readout rate must be high enough to sample the beat frequency between the LO and the return signal, or something must be done to mitigate aliasing. For a single pixel detector this means the beat frequency cannot be sampled if it is higher than half the array read out rate, but the macropixel allows detector bandwidth-limited sampling of the time of arrival of photons. With large enough macropixels, beat frequencies can be sampled up to the detector bandwidth rather than being limited by the detector array frame rate.

In experiments, the return signal for each pulse is coherently integrated—that is, a fast Fourier transform of the photon arrival times for each pulse is performed—then incoherently averaged over multiple pulses—that is, the power spectral densities (PSDs) are summed over all collected pulses. If the LO and transmit signals are offset in frequency by f_{offset}, the beat signal is located at f_{offset}. If the target has a velocity component along the ladar's line of sight, the PSD should show a peak at the sum of the offset frequency and the target's Doppler frequency. The PSD of several return pulses can be averaged to smooth out the curve, as shown in Figure 3-4.

There are important limitations to this technique that limit its applicability and the measurement concept of operations (CONOPS). Principal among these are the dynamic range and visibility constraints and requirements on LO power control, as well as the timescale constraints that link photocount rate, beat frequency, and vibrational frequency. These limitations were discussed in more detail in the section on laser vibrometry in Chapter 2.

As indicated in the preceding section, it is critical to maintain coherence between the wavefronts of the signal and the LO. While it may be possible to minimize the effects on the system side, the transmission medium may ultimately determine the performance of heterodyne detection. In addition to large fluctuations of attenuation produced by things like fog, rain, and smoke, inhomogeneities in the atmosphere itself may produce wavefront distortion. As a result of this wavefront distortion caused by air turbulence, a large portion of the light power can be converted into higher order modes and makes it

difficult to match the LO and return signal patterns. Therefore, heterodyne detection over large distances through the atmosphere is inherently difficult.

Unlike direct detection receivers, the dominant noise source in heterodyne or coherent receivers is the shot noise generated by the local oscillator beam. For a matched filter receiver, that effective noise is equal to one detected photon per resolution element (in both time and space).[5] In order to efficiently contend with this inherent noise, the coherent detection system design is most efficient by ensuring that on the order of 1 (or a few) signal photon(s) is detected per angle/angle/range resolution cell per pulse. Below one detected signal photon per resolution element per pulse, the required transmitter power scales as the square root of the pulse repetition frequency (PRF). Therefore, if a 10 W, 100 Hz transmitter is the optical coherent design (giving ~1 photon per resolution element), then 100 W would be required for a higher PRF, 1 KHz system. This can lead to higher pulse energies and lower PRFs being the optimal energy efficiency solution for many measurement problems, which may not be as feasible as other designs, whether technologically or where low SWaP and/or high reliability are required. Direct detection receivers do not have this fundamental noise constraint and can have much lower than one-detected-noise-photon per resolution element per pulse.[6]

Another fundamental limit of heterodyne detection is the effect of speckle present in highly coherent light. As discussed, the LO and signal must be temporally coherent. They also need to be spatially coherent across the face of the detector and avoid having an additional linear phase difference, or they will produce spatial fringes across a detector, destroying the signal. In many usage scenarios the signal is reflected from optically rough surfaces, producing randomized phases and a "salt and pepper" intensity modulation on the image known as speckle,[7] as also occurs in microwave synthetic-aperture radar, if the illumination source is sufficiently temporally coherent, as for a laser. Techniques such as averaging independent speckle realizations can be performed to reduce the speckle.[8]

Coherent detection techniques require narrow linewidth lasers with coherent illumination. The coherence length must be larger than the two-way round trip time of the pulse, or a sample of the outgoing signal must be stored until the signal return and used to develop the LO. In the second case the coherence length of the illuminating laser must still be longer than twice the depth of the target. A method of shifting the LO frequency is required for knowledge of the direction of the velocity. This is usually accomplished with an acousto-optical modulator. Larger, high bandwidth arrays are also important. Finally, adaptive optics may help reduce the effects of turbulence.

Published literature on the development of laser systems described above would be an indicator. Indicators of progress in heterodyne detector systems may also be found in literature and/or research about some of the applications of coherent ladar (vibrometry, spectroscopy, synthetic aperture ladar, etc.). Additional indicators may be noticed by work on high bandwidth arrays. For heterodyne approaches that do not have sensitive detectors, alternating current (AC) couple array work would also be an indicator.

There are a number of advantages to using heterodyne detection rather than direct detection. First, as mentioned above, a weak signal can be amplified by a strong LO and the signal-to-noise ratio with heterodyne detection depends only on the signal strength, the detector quantum efficiency, and the signal bandwidth. These results are true of both temporal and spatial heterodyne detection. This means that high gain detectors like those used in the direct detect 3-D ladars are not necessarily needed, and shot-noise-limited detection is possible with low-tech photodiodes.

Second, the heterodyne receiver provides high discrimination against background light and other radiation. Unlike direct detection, where background light causes problems when it is of the same order of magnitude as the signal power, in heterodyne detection the background light must be comparable to the LO power, which in many cases is made quite high and can be set to dominate the background light.

---

[5] P. Gatt and S. Henderson, 2001, " Laser radar detection statistics: A comparison of coherent and direct detection receivers," *Proc. SPIE,* 4377: 251.

[6] Personal communication from Sammy Henderson, President, Beyond Photonics, April 20, 2013.

[7] C. Dainty, ed., 1984, *Laser Speckle and Related Phenomena*, Springer Verlag.

[8] J.W. Goodman, 2007, *Speckle Phenomena in Optic: Theory and Applications*, Roberts & Co.

Furthermore, the coherent detection bandwidth can be controlled by a postdetection electronic filter that can be as narrow as desired.[9] Heterodyne detection usually has much narrower receiver bandwidths than direct detection.

In temporal heterodyne detection, a third feature of coherent detection takes advantage of the fact that the amplified output occurs at the difference frequency between the LO and signal beams. This sensitivity to frequency difference makes it possible to measure the phase or frequency shift of the signals and hence obtain Doppler measurements for moving targets. This type of measurement is not directly possible with direct detect systems, which only measure intensity. It takes multiple range measurements to indirectly measure velocity using direct detection.

While heterodyne detection offers the potential for highly sensitive measurements, there are a number of practical limitations to the scheme. Coherent ladars are essentially interferometers. If the phases of the two beams are not well matched, the fringes will oscillate back and forth and the signal will be washed out. This feature lowers efficiency in temporal heterodyne. For heterodyne detection, the two fields from the transmitter and the LO must be spatially locked in phase at the detector. The two beams must be coincident and, to provide maximum signal-to-noise ratio, their diameters must be equal. The beams must propagate in the same direction and the wavefronts must have the same curvature. For spatial heterodyne the LO and return signal propagate in slightly different directions, causing spatial fringes to develop. Finally, for temporal heterodyne, the beams must be identically polarized, so that their electric vectors will be coincident.[10] These requirements are called "coherent superposition," and failure to meet them can cause a loss of signal reception.

A good way to deal with some of these constraints is to use a single laser for both the LO and signal and, for temporal heterodyne, coupling with an acousto-optic modulator to create the frequency difference between the two. In this configuration, the relative phase of the two beams is fairly stable even if the source does not exhibit low phase noise. In addition to the superposition requirements, the laser must be coherent and the coherent length must either be longer than the round trip distance or at least longer than the depth of the target as long as master oscillator drift is compensated by some technique, such as delaying a sample of the master oscillator to use as the local oscillator.

Heterodyne (or coherent) detection can be used in a number of applications such as coherent Doppler ladar measurements, vibrometry, spectroscopy, and very high resolution imaging techniques such as synthetic aperture ladar and inverse synthetic aperture ladar. Several of these applications will be discussed in more detail in later sections. Using a photon counting detector allows use of a low power local oscillator and potentially the ability to use the same detector for both coherent and direct detection.

Conclusion 3-1: Advantages of shot-noise-limited detection, high background discrimination, and measurement of phase or frequency shifts in addition to intensity make heterodyne detection a compelling and promising technology.

Conclusion 3-2: Heterodyne detection can be used with a weak local oscillator if detectors are already sensitive enough so that a strong local oscillator is not required as a method of increasing the receiver sensitivity.

SYNTHETIC-APERTURE LADAR

According to Voxtel "Conventional optical imagers, including imaging. ladars, are limited in angle/angle spatial resolution by the diffraction limit of the telescope aperture. As the aperture size increases, the angle/angle resolution improves; as the range increases, spatial resolution degrades. Thus, high-resolution, real-beam imaging at long ranges requires large telescope

[9] S. Jacobs, 1988, "Optical heterodyne (coherent) detection," *Am. J. Phys.* 56(3): 235.
[10] O.E. DeLange, 1968, "Optical heterodyne detection," *IEEE Spectrum* 5(10): 77.

diameters. Imaging resolution is further dependent on wavelength, with longer wavelengths producing coarser angle/angle resolution. Thus, the limitations of diffraction are most apparent in the radio-frequency domain (as opposed to the optical domain)."[11]

Buell et al. describes "A technique known as synthetic-aperture radar (SAR) was invented in the 1950s to overcome this limitation: In simple terms, a large radar aperture is synthesized by processing the pulses emitted at different locations from a radar aperture as it moves, typically on an airplane or a satellite. The resulting image resolution is characteristic of significantly larger apertures. For example, the Canadian Radar Sat–II, which flies at an altitude of about 800 km, has an antenna size of 15 × 1.5 meters and operates at a wavelength of 5.6 cm. Its real-aperture resolution is on the order of 1 kilometer, while its synthetic-aperture resolution (with a transmission bandwidth of 100 MHz) is as fine as 3 m. This resolution enhancement is made possible by keeping track of the phase history of the radar signal as it travels to the target and returns from various scattering centers in the scene. The final synthetic-aperture radar image is reconstructed from many pulses transmitted and received during a synthetic-aperture evolution time using sophisticated signal processing techniques.[12]"

An alternative description of what is happening to create a high resolution synthetic- aperture ladar (SAL) image is that at a given instant a laser waveform is transmitted from either a monostatic or bistatic aperture. The laser light reflects off the target, and the return field is captured using either spatial or temporal heterodyne. To date, almost all SAL work has been done using temporal heterodyne, but the key issue is capturing a sample of the pupil plane field as large as the real receive aperture. At another location shortly later the same thing is done, and then again and again as the transmit and receive apertures move. If motion issues can be compensated, then a physically large representation of the pupil plane field is captured, which can then be Fourier transformed to form a high-resolution image. Since for monostatic operation both the transmitter and the receiver move, the synthesized pupil plane image is almost twice the distance flown.

In recent years, researchers have investigated ways to apply the techniques and processing tools of RF SARs to optical laser radars. According to Buell et al. "There are several motivations for developing such an approach in the optical or visible domain. The first is simply that humans are used to seeing the world at optical wavelengths. Optical SAL would potentially be easier than microwave radar for humans to interpret, even without specialized training. Second, optical wavelengths are around 20,000 times shorter than RF wavelengths, and can therefore provide much finer spatial resolution and/or much faster imaging times."[13] A typical synthetic aperture motion distance will be many kilometers for SAR but only meters for SAL, assuming the same target resolution requirement. Over time, new applications may arise for which additional resolution requirements impose longer motion distances on SAL.

The SAL concept is illustrated in Figure 3-5. This paragraph is drawn from Beck et al. "A platform with a transmitter-receiver module moves with velocity v while it illuminates a target with light of mean wavelength λ and receives the scattered light. The imaging angular resolution in the direction of travel of the platform is given approximately by $\delta x = \lambda/2\Delta\theta$, where the change in azimuth angle $\Delta\theta$ as seen by an observer at the target at range R is $\Delta\theta = D_{SA}/R$ and the synthetic-aperture length developed during flight time T is $D_{SA} = v \times T$".[14] The range resolution, in the orthogonal direction, is determined by the bandwidth, B, of the transmitted waveform range resolution $\delta y = c/2B$, so long as the receiver can measure the returned bandwidth. Coherent (heterodyne) detection is used to measure the phase history of the returned ladar signals throughout the synthetic-aperture formation time. According to Beck et al. "Two of the main types of synthetic aperture active sensors are (1) spotlight mode (Figure 3-5), in which

[11] Voxtel, http://www.virtualacquisitionshowcase.com/document/602/brochure. Accessed on March 14, 2014.

[12] W.F. Buell, N.J. Marechal, J.R. Buck, R.P. Dickinson, D. Kozlowski, T.J. Wright, and S.M. Beck, 2004 "Synthetic Aperture Imaging Ladar" *Crosslink* (Summer): 45-49.

[13] Ibid.

[14] S. Beck, J. Buck, W. Buell, R. Dickinson, D. Kozlowski, N. Marechal, and T. Wright, 2005 "Synthetic-aperture imaging laser radar: Laboratory demonstration and signal processing," *Appl. Opt.* 44(35): 7621-7629.

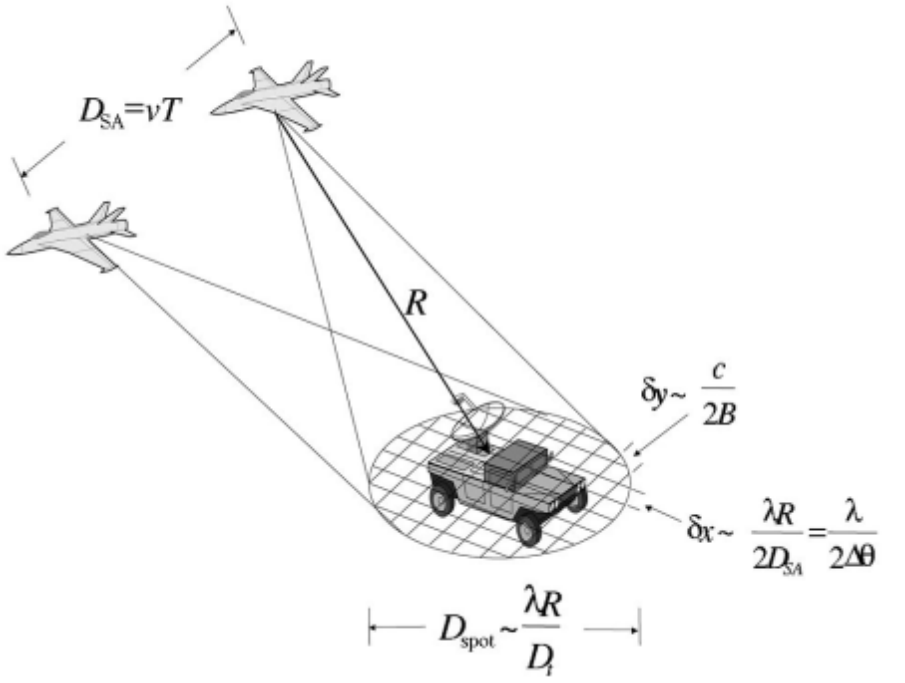

FIGURE 3-5 Spotlight synthetic-aperture ladar (SAL). The illuminating spot size D_{spot}, at the target is determined by the diffraction limit of the transceiver optic, with diameter D_t, corresponding to the imaging resolution of a conventional imager with the same aperture. The resolution in the direction of travel (azimuthal, δx) is determined by the wavelength and evolved aperture length. (For strip-map SAL, this length is limited by the spot illuminating spot size at the target.) The resolution in the orthogonal direction (range, δy) is determined by the transmitted waveform bandwidth, *B*. The angle $\Delta\theta$ is the angle subtended by the synthetic aperture as viewed from an image element at the target. To obtain the resolution in the ground plane, a simple rotation from the slant plane to the ground plane is performed. SOURCE: S.M. Beck, J.R. Buck, W.F. Buell, R.P. Dickinson, D.A. Kozlowski, N.J. Marechal, and T.J. Wright, 2005, "Synthetic aperture imaging laser radar: laboratory demonstration and signal processing," *Appl. Opt.* 44(35): 7621.

the transmitted beam is held at one position on the target for the coherent dwell period and then moved to another spot, and (2) strip mode, in which the transmitted beam is continuously scanned across a target. Most of this discussion applies to either case. In strip mode, the aperture synthesis time is limited by the beamwidth of the sensor and the velocity of the platform (the time during which the target is illuminated). Smaller real apertures result in larger illuminating spots and concomitantly longer synthetic apertures, which leads to the nonintuitive (from a conventional imaging perspective) result that the azimuthal resolution in strip-mode SAL is half of the real-aperture diameter"—the smaller the transmitter aperture the better the resolution.[15] Bistatic configurations, where the transmit and receive apertures are not collocated, can also be considered.

Figure 3-6 shows an early embodiment of a laboratory SAL demonstration system developed at The Aerospace Corporation based on wide-bandwidth frequency-modulated continuous-wave (FMCW) waveforms. According to Beck et al. "The components employed are all common, off-the-shelf, telecommunication fiber-based devices allowing a compact system to be assembled that can easily be isolated from environmental effects: The source is split into five paths, for target illumination, target–LO, reference, reference–LO, and wavelength reference. A circulator is used to recover the return pulse, which

[15] Ibid.

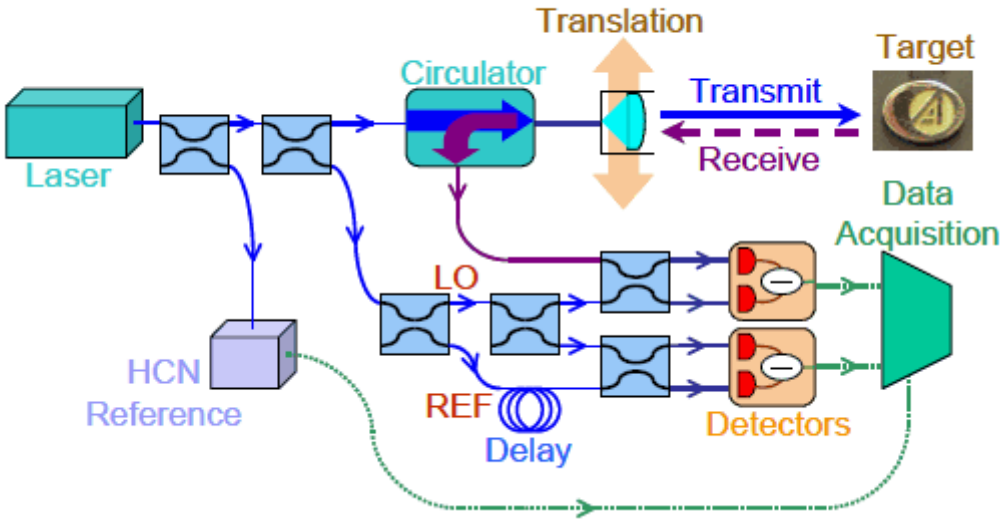

FIGURE 3-6 Component layout for the fiber-based SAIL system. The components employed are all common, off-the-shelf, telecom, fiber-based devices, allowing a very compact system to be assembled that can be easily isolated from environmental effects. The source is split into five paths for the target illumination, target-local oscillator (LO), reference (REF), reference-local oscillator, and wavelength reference. A circulator is used to recover the return pulse, which is mixed with the target-local oscillator in a balanced heterodyne detector. The reference channel is delayed by a fiber loop and then mixed with the reference-local oscillator in a similar manner. The synthetic aperture is created by using a translation stage to scan the aperture across the target. A molecular wavelength reference cell (hydrogen cyanide - HCN) provides a pulse to pulse frequency absolute reference. SOURCE: from Walter Buell (variation of Figure 3 in S.M. Beck, J.R. Buck, W.F. Buell, R.P. Dickinson, D.A. Kozlowski, N.J. Marechal, and T.J. Wright, 2005, "Synthetic aperture imaging ladar: laboratory demonstration and signal processing," *Appl. Opt.* 44(35): 7621).

is mixed with the target–LO in a balanced heterodyne detector. The reference channel is delayed by a fiber loop and then mixed with the reference–LO in a similar manner. The synthetic aperture is created by use of a translation stage to scan the aperture across the target."[16] Figure 3-7 shows a typical laboratory image from the system of Figure 3-6 (see caption for details).

In 2003, DARPA initiated the Synthetic Aperture Lidar Tactical Imaging (SALTI) program with the aim of achieving high-resolution synthetic aperture lidar imagery from an airborne platform at tactical ranges, moving SAL from the laboratory to operationally relevant environments. The performance characteristics of the SALTI program are classified, but the system did achieve synthetic aperture resolution exceeding the real-aperture diffraction-limited resolution of the system. The program progressed through DARPA Phase 3 before being terminated in 2007.

Lockheed Martin Coherent Technologies (LMCT) also pursued an airborne SAL system[17] and presented ground and airborne results in the Laser Sensing and Communications, (LS&C) meeting in Toronto in 2011 (Figure 3-8). While the RSAS SALTI system employed a fiber-based linear FMCW system like the Aerospace system, the LMCT system used a wide-bandwidth (7 GHz, 2 cm resolution), pulse-coded approach.

[16] Ibid.

[17] B. Krause, J. Buck, C. Ryan, D. Hwang, P. Kondratko, A. Malm, A. Gleason, and S. Ashby, 2011, "Synthetic aperture ladar flight demonstration," in CLEO:2011, Laser Applications to Photonic Applications, *OSA Technical Digest* (CD) (Optical Society of America, 2011), PDPB7.

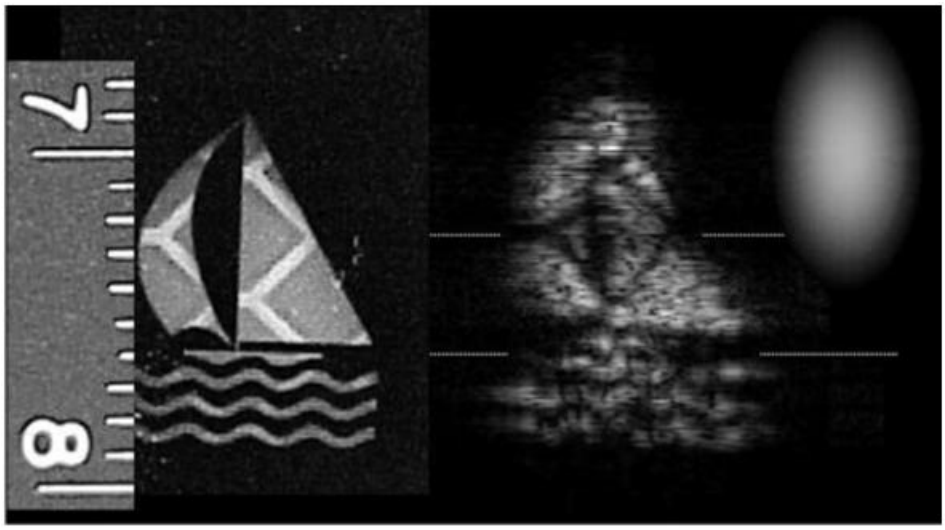

FIGURE 3-7 SAL-boat target and mosaicked SAL image results. The real-aperture diffraction-limited illuminating spot size is represented at the right. A picture of the target is shown at the left. This target consists of the same retroreflective material used for the triangle images, placed behind a transparency containing the negative of the sailboat image. The image was formed by scanning the target in overlapping strips and then pasting these images together to form a larger image. Some degradation is present due to the phase-screening effects of the transparency film; however, the pattern of the retroreflective material is clearly visible. The range to target in this example was ~2 m, with range diversity achieved by placing the target at a 45 degree angle with respect to the incident light. SOURCE: S.M. Beck et al. op. cit.

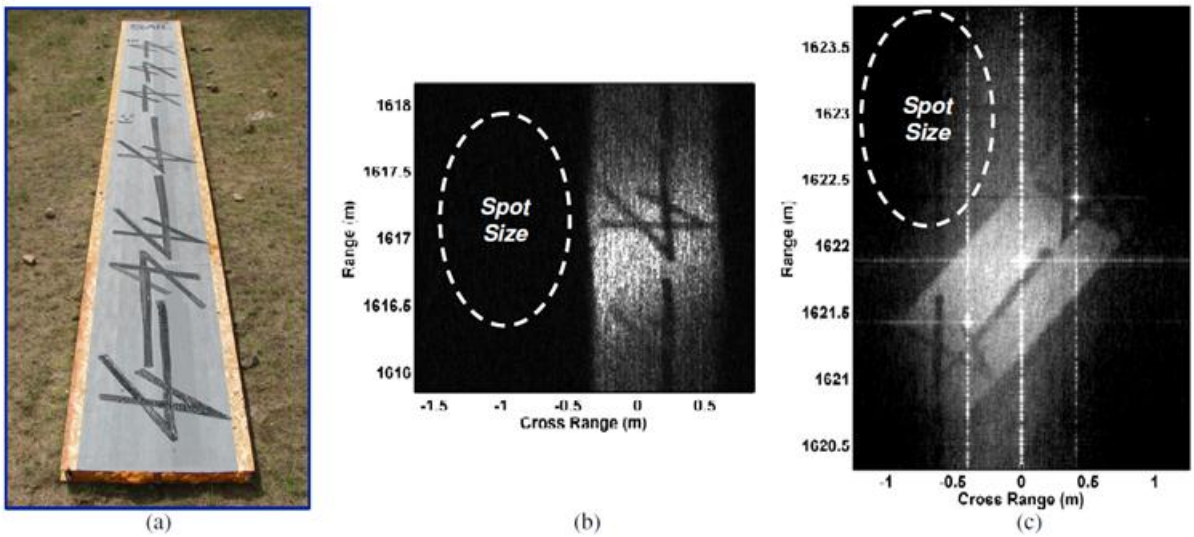

FIGURE 3-8 SAL demonstration images. (a) Photograph of the target. (b) SAL image, no corner cube glints. Cross range resolution = 3.3 cm, 30× improvement over the spot size. Total synthetic aperture = 1.7 m, divided into 10 cm subapertures and incoherently averaged to reduce speckle noise. (c) SAL image with corner cube glint references for clean phase error measurement. Cross range resolution = 2.5 cm, 40× improvement over the spot size. Total synthetic aperture = 5.3 m, divided into 10 cm subapertures and incoherently averaged to reduce speckle noise. SOURCE: B. Krause, J. Buck, C. Ryan, D. Hwang, P. Kondratko, A. Malm, A. Gleason, and S. Ashby, 2011, "Synthetic aperture ladar flight demonstration," in CLEO:2011 - Laser Applications to Photonic Applications, *OSA Technical Digest* (CD) (Optical Society of America, 2011), PDPB7.

The range resolution of any ladar, including SAL, is limited by the transmitted waveform bandwidth, assuming the receiver can capture the returned signal at this bandwidth. This limits performance from technologies that cannot achieve significant transmitter bandwidth (such as CO_2 lasers as employed in the Northrop-Grumman (NGES) approach to the SALTI program). In the absence of bandwidth, angular resolution in the cross-motion direction is limited by the size of the real aperture, usually in elevation. The diffraction limit in this dimension can be expanded by a factor of almost two using techniques described in the MIMO section. The along-track angular resolution is limited by the size of the synthetic aperture:

$$\vartheta = \frac{\lambda}{D + 2L}$$

where D is the real aperture diameter in the along-track direction and L is the distance moved in the along-track direction. In microwave SAR the size of the real aperture is neglected.

The next major limitation is on range performance (how far away the objected viewed can be, not the range resolution), which is constrained by both atmospheric transmission and the modest aperture sizes that aperture synthesis enables. The effect of the atmosphere is discussed further in a later section.

For example, consider the problem of achieving high spatial resolution at low elevation angle through long atmospheric slant paths. At shorter wavelengths, the diffraction-limited resolution is very good, but the propagation through the atmosphere, both from molecular scattering and aerosols, is quite limited. Range can be increased at longer wavelengths where the scattering attenuation is reduced, but at the expense of spatial resolution, until aperture sizes become impractical. Synthetic aperture techniques offer a solution to this dilemma by enabling operation at longer wavelengths, using aperture synthesis to achieve high resolution in the along-track direction with modest aperture sizes.

Of course the SWaP advantages of modest aperture sizes come at the price of not collecting as many photons on receive from each image resolution element, making synthetic aperture ladar a rather power-hungry technique, placing a premium on laser and detector efficiency, and driving laser power requirements. This is a fundamental consideration that whenever resolution is increased, it means photons are scattered from a smaller resolution cell. With a monolithic real aperture, the smaller resolution cell is accompanied by a proportional increase in aperture receiver collection area. In this case, the increase in resolution is not accompanied by a proportional increase in the aperture receiver collection area.

An engineering challenge for SAL is the ability to do motion compensation. The full sample of the field in the large pupil plane synthetic aperture is only available to extent that any movement while the pupil plane field is being collected has been removed. The quality of the motion compensation can limit the SAL resolution.

Atmospheric turbulence is expected to be a limiting phenomenon for SAL imaging at long range. A detailed model of the impact of refractive turbulence on image formation has been presented by Karr[18] based on earlier treatments by Fried.[19] A limited discussion of the effects of atmospheric turbulence on ladars is given in Chapter 4.

As a coherent ladar technique, the technology burdens on the receiver are relatively modest. Deramp-on-receive systems—linear frequency modulated (FM) systems with a chirped LO—have modest detector bandwidth requirements. They do not require the full bandwidth of the transmitted waveform, but just enough bandwidth to accommodate the range depth of the target. This is sometimes called stretch processing, with the LO chirping along with the return signal to reduce required detector bandwidth. Because the image formation is through synthetic aperture processing of the received signal, in principle only a single detector is required. If however only one detector is used, the area imaged would be limited to the diffraction limit of the real beam. This is not a problem in SAR, since real beam microwave resolution is so poor, but it would severely limit the area covered for SAL. In practical systems covering

[18] T.J. Karr, 2003, "Synthetic aperture ladar resolution through turbulence," *Proc. SPIE* 4976: 22.

[19] D.L. Fried, 1966, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *Journal of the Optical Society of America,* 56 (10): 1372.

TABLE 3-1 Required SAL Laser Power for Two Sets of Modeling Assumptions: Airborne System and Spaceborne System

	Airborne	Spaceborne
SNR	10	10
D (m)	0.05	0.5
L	15	15
H (km)	20	350
DF	0.5	0.5
B (GHz)	3.0	1.5
P (W)	90	2500

NOTE: *D*, real aperture diameter in along-track direction; *L*, distance moved in along-track direction; *H*, Height of the platform; *DF*, duty factor; *B*, bandwidth; *P*, Power.
SOURCE: W. Buell, N. Marechal, D. Kozlowski, R. Dickinson, S. Beck, 2002, "SAIL: Synthetic aperture imaging ladar," Meeting of the MSS Specialty Group on Active E-O Systems, Vol. I, C15.

useful areas, typically one uses a modest array of photodetectors to build up multiple real-aperture spots to form an image of a moderate size area. The Raytheon SALTI system used a 1-D array of p-doped-intrinsic-n-doped (p-i-n) photodiodes and scanned that 1-D array. Synthetic aperture processing was performed on each detector separately. Although with sufficient LO power, coherent detection can reach the shot noise limit, there is still a premium on low-noise receivers in order to reduce the required LO power, especially for large arrays. Besides the detectors themselves, the optics of a coherent ladar system, particularly ones with an array of detector elements, can be quite challenging, since the wavefront overlap must be optimized for good heterodyne efficiency.

The requirements on the transmitter are more challenging. As noted above, the modest aperture size that SAL enables means that a smaller fraction of the return light is captured, which must be made up by increased transmitter power. Assuming resolution requirements and aperture diameter are held fixed, the required laser power for a SAL system scales as R^3. Some example link budgets for notional SAL systems[20] are reproduced in Table 3-1. In addition to the transmitter power requirements, SAL also places stringent requirements on local oscillator phase stability. The LO must remain phase coherent over the round trip of the range to target, unless a sample of the transmitted waveform is stored for use in beating against the return signal. One storage method to reduce coherence length requirements is to input a sample of the master oscillator into a long fiber delay and use the output of that fiber as an LO rather than creating the LO from the master oscillator as it exists when the signal is returned from the target. Finally, although less stringent than the LO phase stability requirements, the transmitter waveform quality can be challenging. For a linear FM system, the chirp must be linear to within, say, 20 degrees of phase error over the chirp. This requirement need not be levied entirely on the hardware, however. If transmit waveform phase errors can be monitored, they can be compensated for in the SAL processing. This is the approach taken in the early work at Aerospace. Later work by Bridger Photonics and the University of Montana[21] employing similarly wide bandwidth chirps has achieved the required linearity through sophisticated laser frequency and phase stabilization techniques, simplifying the required processing. The

[20] W. Buell, N. Marechal, D. Kozlowski, R. Dickinson, and S. Beck, 2002, "SAIL: Synthetic aperture imaging ladar," Meeting of the MSS Specialty Group on Active E-O Systems, Vol. I, C15.
[21] S. Crouch and Z.W. Barber, 2012, "Laboratory demonstrations of advanced synthetic aperture ladar techniques," *Optics Express*, 20(22): 24237.

Bridger system has demonstrated resolutions of a few microns in laboratory settings and centimeter resolution in long-range field tests.

Pointing control on SAL systems is not overly stressing, as the transmitter illuminates an area large compared to the SAL spatial resolution. Pointing requirements are simply that the illumination covers the desired target area. Pointing knowledge is a much tighter requirement to a fraction of resolution element.

As noted above, several contractors have developed and demonstrated synthetic aperture ladar systems using a variety of technological approaches and with a significant range of performance parameters. The other country that has demonstrated significant interest in SAL imaging has been China, with a series of papers since 2004 addressing both hardware demonstrations[22] (based largely on U.S. published results) and theoretical analyses, including an advanced algorithm for atmospheric compensation.[23]

It is reasonable to expect that the advantages of SAL and related advanced coherent active imaging techniques will drive the research, development, and deployment of such systems in a variety of countries. This will drive development of high power coherent laser systems capable of achieving wide bandwidth waveforms, as well as long coherence length LOs. Advances in modest size low-noise coherent receiver arrays and techniques for improving heterodyne mixing efficiency over array detectors can also be expected.

A closely related technology to SAL is inverse synthetic aperture ladar (ISAL), in which the transceiver is stationary and the target moves and/or changes aspect relative to the transceiver. This technology is of significant interest for ground-based imaging of space objects, including GEO satellites. ISAL has been the subject of research programs in the United States, such as the DARPA LongView program[24] and research at AFRL.[25] It has also been the subject of research in China.[26,27] It is natural to expect that organizations researching SAL would also be researching ISAL. From the information available in the open literature, the technological developments in China along these lines are not as advanced as those in the United States, but there is ample evidence of interest in continued development.

Conclusion 3-3: Synthetic aperture ladar enables high-resolution active imaging at long range with modest size receiver optics. A synthetic aperture ladar has the potential to provide long-range, easily interpretable (person friendly) imagery because optical systems tend to use mostly diffuse scattering from the viewed object.

Conclusion 3-4: Significant foreign interest in synthetic aperture ladar technology has been demonstrated.

[22] W. Jin, 2010, "Matched filter in synthetic aperture ladar imaging," *Acta Optica Sinica* 2010-07.

[23] L. Guo, M. Xing, Y. Tang, and J. Dan, 2008, "A novel modified omega-k algorithm for synthetic aperture imaging lidar through the atmosphere," *Sensors* 8: 3056.

[24] http://www.globalsecurity.org/space/systems/long-view.htm.

[25] C.J. Pellizzari, C.L. Matson, and R. Gudimetla, 2011, "Inverse synthetic aperture LADAR for geosynchronous space objects–Signal-to-noise analysis," Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference, Wailea, Maui, Hawaii, September 13-16, S. Ryan, ed.

[26] X. Zhao, X. Zeng, C. Cao, Z. Feng, and C. Fu, 2009, "Research on inverse synthetic aperture ladar," *Proc. SPIE*, 7382, International Symposium on Photoelectronic Detection and Imaging 2009: Laser Sensing and Imaging, August 27.

[27] G. Liang, 2009, "Study on experiment and algorithm of synthetic aperture imaging lidar," Ph.D. dissertation, Xi'an University of Electronic Science and Technology.

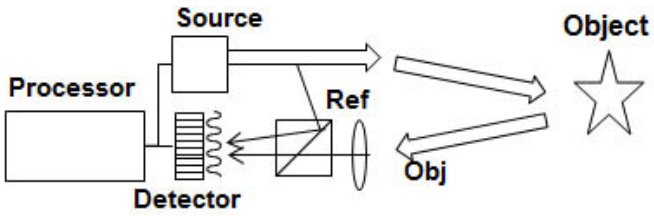

FIGURE 3-9 Diagram of off-axis holography experiment.
SOURCE: J. Marron, Raytheon, Space and Airborne Systems, "3D Holographic Laser Radar," presentation to the committee, March 5, 2013.

In summary, synthetic aperture ladar is becoming an increasingly mature coherent ladar technology that offers considerable system-level and performance advantages over real-aperture imaging ladar systems. The technology base continues to evolve, enabling improved angular resolution and increased standoff range. It is reasonable to expect that this technology will proliferate to other countries and initial indicators of this can be seen.

DIGITAL HOLOGRAPHY/SPATIAL HETERODYNE

Digital holography, also referred to as spatial heterodyne detection, utilizes two light beams with a spatial/angular difference (as opposed to frequency difference) between them. In digital holography, the return signal from an illuminated object is coherently interfered with a reference beam. The reference beam could be a glint/retroreflector in close physical proximity to the object of interest or a local oscillator signal that transverses a similar optical path length to the return from the object in order to maintain coherence properties between the two optical beams. Figure 3-9 depicts a typical holographic arrangement using a reference or local oscillator beam. The laser is split into two beam paths; the transmitter path illuminates the target and the reflection from the target is interfered using a beam splitter with the reference beam on a detector focal plane array (FPA). The FPA for spatial heterodyne can be a framing array; it does not have to have high bandwidth, because only the fringes across the array are being detected, not any high bandwidth signals.

The interference between the object and reference beams is recorded on the charge-coupled device (CCD) detector. Although detectors record only intensity and do not directly preserve the phase profile of the electric field, digital holography provides a means to extract the spatial phase variation across an optical aperture using the spatial beat frequency between the signal and the LO. As previously discussed in the section on synthetic aperture ladar, having access to both the amplitude and phase of the optical field enables capabilities not readily possible with intensity-only imaging; with digital post-processing, the exact electric field at any point can be calculated from spatial heterodyne phase extraction. This can allow for digital refocusing and 3-D imaging.

When the scenario or requirements allow a retroreflector to be placed in proximity to the target, the coherence requirements on the system may be lessened. In addition, aberrations imparted by turbulence in the atmosphere are nearly identical if the retroreflector is within the same isoplanatic patch as the target, meaning that the effects of turbulence will be minimized. Early long-range holography experiments were performed at up to 12 km.[28]

In heterodyne detection, the image term is proportional to the strength of the LO component multiplied by the strength of the object component; an extremely weak image signal can be magnified by

[28] J.W. Goodman, D.W. Jackson, M. Lehmann, and J. Knotts, 1969,"Experiments in long-distance holographic imagery," *Appl Optics* 8: 1581.

a strong reference or LO signal. This can be strongly advantageous when extracting the image component of interest. An important feature of a strong LO arrangement is that the object's autocorrelation term will be extremely weak and negligible compared to the strength of the image term, allowing the use of lower spatial frequencies while still achieving separation between image terms.[29] The maximum strength of the local oscillator will be limited by the electron well capacity of the detector pixel. Detector saturation occurs when the well capacity is exceeded.

Techniques using this technology have several different names. In additional to digital holography, spatial heterodyne detection and lensless imaging, scientists and engineers in the field also refer to this technology as "holographic aperture ladar"[30] and "spatially processed image detection and ranging (SPIDAR)."[31]

In addition to performing single-wavelength digital holography, multiple-wavelength holography is a subset that allows for fine-resolution 3-D imaging. Figure 3-10 shows results using multiple-wavelength digital holography. As shown in Figure 3-10 (b) and (c), active imaging provides intensity and phase information about the object, giving both shading and, with multiple wavelengths, depth information.

By utilizing two wavelengths, a contour map of the object surface can be generated by recording two holograms at two wavelengths; a fringe pattern is generated by reconstructing and superimposing the holograms.[32] Alternatively, using a tunable laser source, a series of holograms are recorded for a set of laser frequencies.[33] The collected dataset has of coordinates of (spatial frequency, spatial frequency, laser frequency). By Fourier transforming the hologram data, the resulting image has coordinates of (angle, angle, range). The range resolution ΔR_{res} of the system is given by:[34]

$$\Delta R_{res} = \frac{c}{2\Delta \nu_{tot}}$$

where $\Delta \nu_{tot}$ is the total frequency bandwidth over which the source is tuned. Additionally, the range ambiguity interval is given by $\Delta R_{unamb} = c/(2\ \Delta \nu_{inc})$ where $\Delta \nu_{inc}$ is the frequency sampling increment for the laser. Using multiple-wavelength holography, it is possible to measure a large range with high depth resolution without 2π ambiguities of the phase difference that can occur with the two-wavelength version.

Digital holographic EO sensing has military applications in intelligence, surveillance, and reconnaissance (ISR), target tracking, target identification, and directed energy. Beyond these specific applications, digital holographic EO sensing is more broadly used in biomedical imaging/microscopy, imaging through scattering media, horizontal path imaging, and 3-D imaging for commercial and entertainment purposes. A broad overview of the latest advancements in digital holography was published recently[35] and articles therein discuss research topics across a wide variety of applications.

Digital holography is a sensing technology. However, the outcomes of spatial heterodyne detection may influence the overall system architecture for a larger system, such as a directed energy system, which needs to transmit a laser beam. Using digital holography, phase aberrations present in the atmosphere between the object and sensor can be estimated. Electro-optic phase modulators, liquid crystal spatial light modulators, or piezo mirrors can pre-distort the outgoing illumination beam to compensate

[29] J.C. Marron, 2009, "Photon noise in digital holographic detection," AFRL-RD-PS-TP-2009-1006.

[30] B.D. Duncan and M.P. Dierking, 2009,"Holographic aperture ladar," *Appl. Opt.* 48: 1168.

[31] J.C. Marron, 2008, "Spatially processed image detection and ranging (SPIDAR)," in *IEEE LEOS Meeting* (IEEE 2008), 509.

[32] B. Hildebrand and K. Haines, 1967, "Multiple-wavelength and multiple-source holography applied to contour generation," *J. Opt. Soc. Am.* 57: 155.

[33] A. Wada, M. Kato, and Y. Ishii, 2008, "Multiple-wavelength digital holographic interferometry using tunable laser diodes," *Appl. Optics* 47: 2053.

[34] J.C. Marron and K.S. Schroeder, 1993, "Holographic laser-radar," *Opt. Lett.* 18: 385.

[35] M.K. Kim, Y. Hayasaki, P. Picart, and J. Rosen, 2013, "Digital holography and 3D imaging: introduction to feature issue," *Appl Optics* 52: Dh1-Dh1.

(a) (b) (c)

FIGURE 3-10 (a) Passive broadband image of a mannequin taken at a 100 m range using 1.0 to 1.65 micron light; (b) active image of same object using 1.6 μm illumination; (c) 3-D phase difference acquired through active illumination. SOURCE: J.C. Marron, R.L. Kendrick, S.T. Thurman, N.L. Seldomridge, T.D. Grow, C.W. Embry, and A.T. Bratcher, 2010, "Extended-range digital holographic imaging," *Proc. SPIE*, Laser Radar Technology and Applications XV, 7684, 76841J.

for these aberrations. As a result, digital holography may influence the entire system architecture, from how outgoing beams are transmitted to how return beams are received.

For the current horizontal path imaging, published results show an extended range of 1.5 km and voxel dimensions of 3 cm. Work was performed at 1.6 μm using an Er:YAG laser. The detector array was a commercial InGaAs array with 640 × 512 pixels. Images were taken of a truck, a mannequin, a resolution chart, a model missile, and some calibration blocks.

Spatial heterodyning is an active, emerging technology. Work has been performed in research laboratories as well as field test demonstrations.[36] Trends in this area have recently focused on variations of aperture and focal plane array arrangements, discussed more in the next section on multiple-input, multiple output (MIMO) receiver and transmitter geometries. Digital holography was first performed using a single focal plane array. In order to increase the aperture size, current work is geared toward optimizing this technology for multiple apertures (with gaps between adjacent apertures)[37,38] and synthetic apertures (combining multiple apertures together to form a zero-gap, larger full aperture).[39,40,41,42] Proper phasing between individual arrays increases the complexity of these multiaperture apertures compared to a single FPA; this is a current area of research.[43]

The maximum angular resolution is set by the full effective FPA width when the image is sampled in the pupil plane. Therefore, a larger effective FPA size will result in finer angular resolution for a given wavelength and imaging distance. Utilizing synthetic aperture techniques with spatial heterodyning can further increase the effective array size. Optical magnification can also improve the

[36] J.C. Marron, R.L. Kendrick, N. Seldomridge, T.D. Grow, and T.A. Hoft, 2009, "Atmospheric turbulence correction using digital holographic detection: Experimental results," *Opt. Exp.* 17: 11638.

[37] J.W. Haus, N.J. Miller, P. McManamon, and D. Shemano, 2011, "Digital holography for coherent imaging for multi-aperture laser radar," Conference paper, *Digital Holography and Three-Dimensional Imaging*, Tokyo Japan, May 9-11, Optical Society of America, p. DMA3.

[38] R.L. Kendrick, J.C. Marron, and R. Benson, 2009, "Anisoplanatic wavefront error estimation using coherent imaging," in *Coherent Laser Radar Conference* (Toulouse, France), 205.

[39] A.E. Tippie, A. Kumar, and J.R. Fienup, 2011, "High-resolution synthetic-aperture digital holography with digital phase and pupil correction," *Opt. Exp.* 19: 12027.

[40] J.H. Massig, 2002, "Digital off-axis holography with a synthetic aperture," *Opt. Lett.* 27: 2179.

[41] D. Claus, 2010, "High resolution digital holographic synthetic aperture applied to deformation measurement and extended depth of field method," *Appl. Opt.* 49: 3187.

[42] R. Binet, J. Colineau, and J.C. Lehureau, 2002, "Short-range synthetic aperture imaging at 633 nm by digital holography," *Appl. Opt.* 41: 4775.

[43] B.K. Gunturk, N.J. Miller, and E.A. Watson, 2012, "Camera phasing in multi-aperture coherent imaging," *Opt Express* 20: 11796.

angular resolution by effectively expanding the size of the FPA. The field of view of the system will be determined by the pixel pitch of the detector elements. The smaller the pixel pitch, the larger the field of view for pupil plane based imaging. As with many optical systems, the push is still toward larger and larger arrays and smaller and smaller pixels. The roles of pixel pitch and FPA arrays size are reversed when the FPA is placed in the image plane.

Laser requirements are an important consideration for digital holographic imaging. This technology requires coherent illumination with narrow linewidths in order to record the interference between the local oscillator and object return signal. Narrow linewidths and coherence length requirements are inherent issues for all types of coherent imaging; the section on synthetic aperture ladar previously discussed provides details regarding these issues. The narrow linewidth required may be considered a constraint of the technology, as the current cost of these lasers is significant.

Another fundamental drawback of spatial heterodyning is the speckle effects inherently present in highly coherent light. Speckle effectively reduces the angular resolution from the theoretical limit;[44] techniques such as averaging independent speckle realizations can be performed to reduce the speckle.[45] Independent speckle realizations require multiple exposures and/or dividing the full aperture into subapertures, thus increasing the acquisition and processing time.

Atmospheric turbulence may ultimately be an external fundamental limit affecting this technology. Although digital holography allows for phase aberration correction, it may be limited over a small isoplanatic patch if the turbulence is severe enough and evenly distributed throughout the imaging path. [46]

As an emerging technology, significant indicators of technology development will most likely be seen through research funding and publications in universities, research institutions, government laboratories, or private industry. Demonstration projects are most likely; production scaling would indicate significant progress and acceptance of digital holography. The further development of digital holography for commercial applications such as 3-D technology for entertainment could drive this technology forward as well.

Spatial heterodyne has two main advantages over both passive imaging systems and many other forms of ladar systems. First, spatial heterodyning is potentially a “lensless” imaging technique. Unlike conventional imaging techniques that use lenses and other optics to form the conjugate of the image on the detector plane, spatially heterodyne systems can map the pupil plane directly onto the detector. The image is then digitally converted from the recorded detection. From a practical standpoint, SWaP may be reduced compared to these other conventional systems since the weight and volume of the optics/lenses can be removed. As the desire to create larger and larger focal planes to obtain higher imaging angular resolution increases, the weight and volume of the corresponding optics for a focal plane imaging system also increases significantly. Using a spatial heterodyne imaging modality, the focal plane aperture area can continue to increase with less burden on SWaP requirements.

If the receive aperture is not a 1:1 ratio with the FPA, a telescope is required to adjust the size of the pupil plane. However, in synthetic aperture digital holography, multiple smaller telescopes may be used, still allowing for a reduction in volume and weight compared to a single (longer) focal length telescope for the entire FPA.

Atmospheric turbulence can severely degrade both the outgoing illumination beam as well as the return signal, resulting in the need for higher laser power from the increased beam divergence and reduced angular resolution and degraded imagery. Even as the aperture size of the imaging system is increased for higher angular resolution capabilities, atmospheric turbulence effects can undermine and limit the achievable angular resolution. In astronomy good seeing conditions for telescopes, defined by the coherence area, the area over which the incoming light is considered to be spatially correlated, is

[44] A. Kozma, and C.R. Christensen, 1976, “Effects of speckle on resolution,” *J Opt Soc Am* 66: 1257.

[45] J.W. Goodman, 2007, *Speckle Phenomena in Optics: Theory and Applications*, Roberts & Co.

[46] A.E. Tippie and J.R. Fienup, 2010, “Multiple-plane anisoplanatic phase correction in a laboratory digital holography experiment,” *Opt. Lett.* 35: 3291.

typically on the order of 10 cm. For ground-to-ground or horizontal path imaging scenarios, the coherence area is even smaller. Adaptive optics using wavefront sensors and deformable mirrors are one way astronomers combat the turbulence problem while at the same time adding weight and complexity to the complete imaging system. Digital holographic techniques can address some of these atmospheric turbulence effects directly; initial results are presented in Marron et al.[47] and Tippie and Fienup.[48]

The fundamental limits associated with digital holography—laser coherence requirements, speckle effects, and degradation in imaging due to atmospheric turbulence—would be factors that may be seen as downsides to implementing this technology.

Since pupil-plane spatial heterodyne does not directly acquire focal plane images on the detector array, processing is required to extract the desired information. In the case of optimizing for aberration correction and/or proper phasing of multiapertures, additional computation is required. As the focal plane size continues to increase, the computation burden will scale as well. Dedicated hardware and parallelization of processes will be required to reduce the processing time as much as possible. Parallel implementation or use of graphics cards may reduce the computation time for the image reconstruction process. As developments in processing continue, the burden of postprocessing for image correction will lessen.

As an emerging technology, there is strong possibility for digital holography to grow. As mentioned previously, digital holography has applicability in intelligence, surveillance and reconnaissance (ISR), target tracking, target identification, and directed energy. As the technology continues to advance, digital holographic systems will likely be designed with these applications in mind. Advancements in narrow laser linewidths and laser power will have direct impact on the ranging capabilities of this technology, as longer and longer ranges are desired.

If digital holography is implemented as "lensless imaging," one possible future capability would be the use of digital holography with conformal FPAs. An array of conformal focal planes could match the shape of the surface of the designated platform (vehicle, plane, etc.), enabling more flexibility in design, as well as the collection and capture of returned light. However, many serious technical issues such as the conformal focal planes technology itself, as well as analysis and implementation of local oscillator illumination of a curved surface, would need to be researched and proven before such designs could be considered.

In the broad field of digital holography, researchers worldwide are actively involved in this field of research. Key countries outside the United States include France, Germany, Israel, Japan, and China.[49] While the United States may be considered a leader in digital holography for remote sensing applications, researchers in Pacific Rim countries continue to make significant progress in holography for commercial applications for 3-D imaging and display.

In summary, current trends in spatial heterodyning have focused on variations of aperture and FPA arrangements using multiple apertures (with gaps between adjacent apertures) and using motion to form synthetic apertures (combining multiple apertures sampled at different times to form a zero-gap, larger full aperture).

Conclusion 3-5: Digital holography/spatial heterodyne is a growing segment of active EO sensing, evidenced by conferences being held across the world that cover the diverse applications of this technical area.

[47] J.C. Marron, R.L. Kendrick, N. Seldomridge, T.D. Grow, and T.A. Hoft, 2009, "Atmospheric turbulence correction using digital holographic detection: Experimental results," *Opt. Exp.* 17: 11638.

[48] A.E. Tippie and J.R. Fienup, 2010, "Multiple-plane anisoplanatic phase correction in a laboratory digital holography experiment," *Opt. Lett.* 35: 3291.

[49] See, for example, H. Luo, X.H. Yuan, and Y. Zeng, 2013, "Range accuracy of photon heterodyne detection with laser pulse based on Geiger-mode APD," *Optics Express* 21(16): 18983.

MULTIPLE INPUT, MULTIPLE OUTPUT ACTIVE ELECTRO-OPTICAL SENSING

Using multiple transmitter apertures and multiple receive apertures allows design flexibility not present in monolithic apertures. SAR is comprised of a single moving aperture, so the transmit and receive apertures both move.[50,51,52,53] A SAL also uses motion of the single aperture to synthesize a larger effective aperture. For both SAR and SAL, it is as though the flown distance is a synthetic aperture almost twice as large as the actual flown distance, because the angle of incidence is equal to the angle of reflection. This has been experimentally demonstrated over decades in the microwave region and, recently in the optical regime, when SAL was demonstrated.[54,55] Duncan[56] provides a crossrange resolution equation for spotlight-mode synthetic aperture ladar:

$$\delta = \frac{R\lambda}{2L + D}$$

where D is the size of the aperture, R is the distance to the object, λ is the wavelength, and L is the distance moved. As a result, the effective aperture from a diffraction point of view is

$$D_{eff} = 2 * L + D_{real}$$

For RF systems, the real aperture is tiny (on the order of meters) compared to the distance moved in a SAR (likely multiple kilometers), so the size of the real aperture is neglected, making the effective aperture twice as large as a monolithic aperture of width equal to the distance flown. In an EO system, the real aperture can be a significant fraction of the size of the moved aperture, so the D_{real} term is retained.

Instead of using motion to synthesize a larger effective aperture, MIMO active EO sensing uses multiple physical subapertures to create a larger effective aperture. An array of receive-only subapertures can synthesize an effective aperture as large as the receive array if the field across the array can be measured or estimated. With an array of transmit and receive subapertures, even more flexibility is obtained, so long as it is possible on receive to identify which transmitter each photon initially came from.

One effect of multiple transmit and receive subapertures is increased angular resolution similar to the angular resolution increase from motion based synthetic aperture sensors. Instead of motion, an array of n subapertures that both transmit and receive can be used. For nine subapertures in a row, the array will have a diffraction limit consistent with a monolithic aperture that is 1.89 times as large in diameter as the array. This is because eight subapertures are equivalent to the distance moved, L, while one subaperture is equivalent to the real aperture, D_{real}. If transmission occurs from one subaperture in the middle of an array, the receive aperture array is effectively in its normal location. If the transmit beam is moved up one subaperture, it is as though the receive aperture were moved down one subaperture. If this process is continued, the result is something like what is shown in Figure 3-11, where the lighter color linear arrays indicate the perceived location of the linear receive array, depending on which transmit subaperture is used. The dark color column shows the actual location of the arrays. The full extent of the linear arrays is

[50] M.I. Skolnik, 1990, *Radar Handbook* (2nd ed.) McGraw-Hill, New York. Chapter 17, by Roger Sullivan Eq. 17.1 and 17.2, Figure 17.2.

[51] M. Soumekh, 1999, *Synthetic Aperture Radar Signal Processing with Matlab Algorithms*, Wiley, New York, Section 2.6, Cross Range Resolution.

[52] M.A. Richards, 2005, *Fundamentals of Radar Signal Processing*, McGraw-Hill, New York. Chapter 8.

[53] M.I. Skolnik, 1980, *Introduction to Radar Systems* (2nd ed.), McGraw-Hill, New York. Chapter 14.

[54] B. Krause et al., 2011, "Synthetic aperture ladar flight demonstration," Conference on Lasers and Electro-Optics, PDPB7.

[55] S.M. Beck et al., 2005, "Synthetic-aperture imaging laser radar: Laboratory demonstration and signal processing," *Appl. Opt.* 44: 7621.

[56] B.D. Duncan and M.P. Dierking, 2009, "Holographic aperture ladar," *Applied Optics* 48(6): 1168.

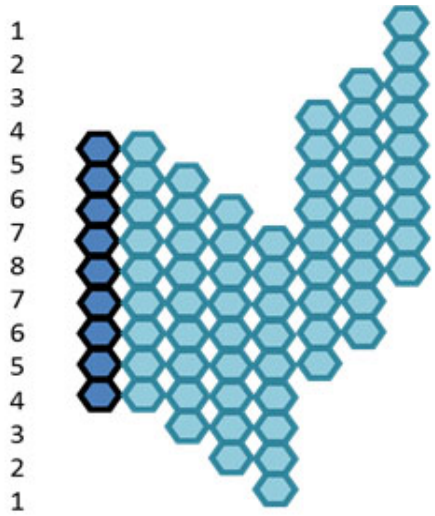

FIGURE 3-11 Effective receiver aperture placement based on transmit subaperture utilized. Dark color shows actual location of the arrays. Light color shows perceived location. The number on the left shows how many receive subapertures are perceived to be at that location.

1.89 times larger, but it is sampled more near the middle of the effective aperture, represented by eight samples in the middle down to one on either end. This is similar to an apodized aperture.

The potential increased angular resolution from the arrays of transmit and receive subapertures is just one facet provided by MIMO imaging. If tagged transmitters are spaced at distances less than the size of a receive subaperture, it provides sampling that can allow closed-form solution to differences in atmospheric path across the subaperture array.[57] An example of the resulting effective receiver pupil using illuminator diversity is shown in Figure 3-12. Because of the overlap, one can solve in a closed-form manner for the phase between subapertures, allowing its use to compensate for atmospheric phase disturbances in the pupil plane.

MIMO imaging systems are very new, and all of their uses have probably not yet been discovered. Multiple transmitters of course can allow more rapid compensation for speckle because more realizations of speckle can be gathered rapidly.

MIMO techniques as described here can be implemented using either temporal or spatial heterodyne (digital holography) techniques. It will, however, be much easier to tag the emitted transmitter signals, allowing simultaneous transmission, if high bandwidth temporal heterodyne is used, since high bandwidth tagging schemes can then be used. RF MIMO techniques that use multiple simultaneous phase centers have been developed by Coutts et al.[58] Tyler talks about using transmitter diversity to phase up laser beams on transmit.[59] This could be used for an illuminator.

The angular resolution of an array of subapertures can be almost twice the resolution of the diffraction limit for a monolithic aperture. In addition, atmospheric turbulence between the object imaged and the sensor can be very quickly and accurately compensated. This compensation is relatively straightforward for turbulence in the pupil plane but will be more difficult for volume turbulence. These are narrow band sensors, but using multiple transmitters allows speckle mitigation. A significant

[57] J.R. Fienup, 2000, "Phase error correction for synthetic-aperture phased-array imaging systems," *Proc. SPIE* 4123-06: 47, *Image Reconstruction from Incomplete Data*, San Diego, Calif.

[58] S. Coutts, K. Cuomo, J. McHarg, F. Robey, and D. Weikle, 2006, "Distributed coherent aperture measurements for next generation BMD radar," Fourth IEEE Workshop on Sensor Array and Multichannel Processing.

[59] G. Tyler, "Accommodation of speckle in object-based phasing," 2012, *J. Opt. Soc. Am. A* 29(4).

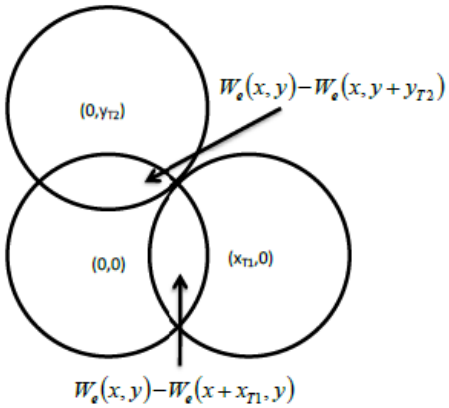

FIGURE 3-12 Multiple subaperture overlap using three transmitters in a pattern about half a receive subaperture diameter apart. When images are adjusted to account for illuminator location, the receiver subaperture pupils overlap as shown. SOURCE: D.J. Rabb, J.W. Stafford, and D.F. Jameson, 2011 "Non-iterative aberration correction of a multiple transmitter system," *Optics Express* 19(25): 25048D.

limitation is that the received signal is captured in a smaller receive aperture area. The required laser power will increase by the ratio of the area of the monolithic aperture to the received aperture array area. Also, if the angular resolution is greater than a monolithic subaperture, less laser return will come from each image voxel.

Arrays of high temporal bandwidth detectors will be very helpful in implementing MIMO techniques for EO imaging. The papers cited sequenced through the multiple transmitters because they implemented MIMO using a digital holography/spatial heterodyne approach to imaging, using low bandwidth framing detector arrays. If high-bandwidth detector arrays and a temporal heterodyne approach to imaging are used, it should be possible to simultaneously emit multiple tagged transmitter beams and to have each receiver be able to distinguish the transmit aperture any photon came from. High-bandwidth detectors will allow high-bandwidth modulations to be imposed on each transmitted beam, and sorted on receive. Temporal heterodyne arrays will need to be AC-coupled, or have high dynamic range, or have high sensitivity such that temporal heterodyne can be implemented using a relatively weak LO.

A second technical hurdle to overcome is volume turbulence. Techniques for calculating and compensating for volume turbulence still need to be developed.

Published work in multiple aperture array active sensors systems using transmitter as well as receiver diversity would be one indicator of active interest in this area. Work in high-bandwidth detector arrays suitable for use in temporal heterodyne sensors would be another indicator.

MIMO technology will allow imaging with high angular resolution using much lighter and more compact aperture arrays than a monolithic aperture. An array of small subapertures can be much thinner and lighter than a monolithic aperture. Also, a MIMO approach can achieve almost twice the diffraction limited angular resolution of a monolithic aperture. Speckle averaging using multiple transmitters will be another advantage.

To be really useful in freezing the atmosphere, MIMO should be implemented using high-bandwidth detector arrays, which still need to be further developed. There will be a digital implementation requirement to conduct the required calculations. There will also be a need to have many different optical trains, complicating the optical system. Also, narrow line lasers will need to be used to

do either spatial or temporal heterodyne. Higher power lasers will be required to image a given area using a MIMO array compared with imaging with a monolithic aperture

This technology is well suited for long-range imaging applications from air or space. MIMO could be used in the cross-range dimension along with motion-based synthetic aperture imaging.

The United States appears to have a lead in this technology so far, but the developments to date have not been developments requiring large investments or significant infrastructure, so this lead could evaporate quickly. The United States also has a lead in high-bandwidth FPAs that can be used in MIMO applications. High-bandwidth FPAs are a more enduring lead in terms of infrastructure required to produce them, so that can help preserve the U.S. lead in this area.

In summary, MIMO approaches for active EO sensing can, at a minimum, increase the effective diameter of an aperture array by a factor of almost two and can allow multiple subapertures on receive to be phased using a closed-loop calculation of the phase difference between subapertures. This can compensate for atmospheric turbulence at least at some locations between the sensor and the imaged object.

Conclusion 3-6: The multiple input, multiple output approach is a very promising active EO research area. At a minimum, it will be very valuable for longer-range imaging sensors, and is also likely to become valuable for many other applications.

SPECKLE IMAGING

Speckle is inherent in laser radar measurements. When a beam of light from a laser transmitter illuminates a target, phase irregularities occur in the backscattered light coming from the backscattered regions. This is due to surface roughness on the scale of a wavelength of light. Interference between the various contributions to the optical field produces a speckle pattern of bright and dark intensity regions in the receiver. In many optical applications, speckle is considered a nuisance—it degrades target images obtained when conventional microwave-radar imaging techniques are applied to laser radar. However, researchers since the 1960s have known that the speckle pattern carries information about the physical properties of a target and have looked for ways to take advantage of the information contained in the speckle patterns.[60]

The shape, size, and distribution of speckle depends on a number of different target-related parameters, including the surface roughness, the angular tilt and movement of the surface, as well as sensor-related characteristics such as the laser wavelength and bandwidth, size of the patch being illuminated, angle of illumination, angle of detection, illumination intensity, and distance between the focal plane and speckle image plane. Speckle imaging techniques take advantage of the correlations between changes in speckle patterns due to small changes in one or more of these parameters.

A variety of laser speckle techniques have been used in biomedical applications. Some of these techniques use the statistics of the speckle patterns to get information about surface roughness or the nonlocalization properties for vision testing—for example, to measure the refraction of the eye. Speckle photography has been used to measure displacements, vibrations and strains.

Speckle techniques such as electronic speckle-pattern interferometry (ESPI) overcame this problem by storing speckle patterns electronically and using image processing techniques to compare the images in near real time. This method has been widely used for displacement measuring and vibration analysis. It takes advantage of movement of the target "surface" and produces correlation fringes that correspond to the object's local surface displacements between two subsequent exposures, with or without

[60] L. Shirley et al., 1992, "Advanced Techniques for Target Discrimination Using Laser Speckle," *Lincoln Lab. J.* 5(3), 367-440.

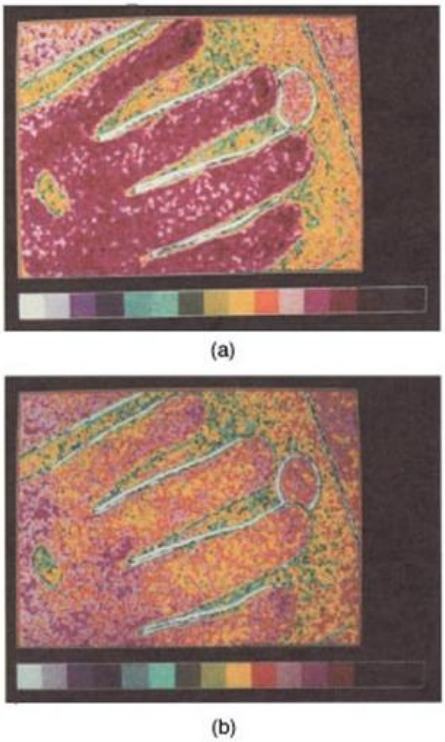

FIGURE 3-13 LASCA images of a hand, showing the effect of occluded blood flow: (a) hand under normal conditions and (b) same hand when a blood-pressure cuff is inflated to reduce blood flow to the hand. SOURCE: J. Briers and S. Webster, 1996, "Laser speckle contrast analysis (LASCA): a nonscanning, full-field technique for monitoring capillary blood flow," *J. Biomedical Optics* 1(2): 174.

an additional reference beam.[61] The addition of a local oscillator phase and frequency modulation using an optical phase-locked loop to the ESPI system (OPL-ESPI) allows for the generation of Doppler speckle contours of a vibrating surface even from unstable sensor platforms.[62]

Laser speckle contrast analysis (LASCA) is another method that utilizes the movement of the "target" in that it measures flow velocity. Instead of using the temporal statistics of speckle like many other techniques, this method requires only a single exposure and uses the spatial statistics of the laser speckle to map the flow velocity at a particular instant in time without requiring scanning.[63] With LASCA, a time-varying speckle pattern is captured by an imaging device with a finite integration time, causing some of the speckle fluctuations to be averaged out, or blurred. LASCA takes advantage of the fact that the ratio of the standard deviation to the mean intensity of the speckle patterns can provide a measure of the contrast of the pattern. If there is a lot of movement, the blurring will increase and the standard deviation of the intensity will decrease, resulting in a lower contrast. Conversely, if there is little

[61] J. Garcia et al., 2008, "Projection of speckle patterns for 3D sensing," *J. Phys.,* Conf. Ser. 139-12026.

[62] S. Moran et. al., 1987, "Optically phase-locked electronic speckle pattern interferometer," *Appl. Optics* 26(3): 475.

[63] J. Briers and S. Webster, 1996, "Laser speckle contrast analysis (LASCA): A nonscanning, full-field technique for monitoring capillary blood flow," *J. Biomedical Optics* 1(2): 174.

movement, the blurring will decrease and the standard deviation will increase, resulting in a higher contrast. The mean intensity will remain unchanged.[64] If the right wavelength is chosen, these noncontact velocity-measuring methods are able to see through skin and provide a velocity map of capillary blood flow in biomedical applications. Figure 3-13 shows an example of a blood flow map obtained by the LASCA technique.

Several techniques that take advantage of the information contained in speckle patterns have been investigated for the purpose of imaging earth-orbiting satellites with high resolution from ground based imaging systems. Imaging correlography and sheared-beam imaging (SBI) (also called sheared coherent interferometric photography (SCIP)) are able to produce 2-D intensity images of the targets of interest at potentially long ranges using lasers with reduced temporal coherence requirements, no local oscillator, and simple array detectors. A more detailed treatment of these techniques can be found in Voelz et al., 2002[65] and the references therein. The first of these techniques, imaging correlography,[66,67] which uses a single laser and takes array measurements of intensity speckle patterns over multiple frames (with the assumption that the target or the array is moving relative to the other) has been demonstrated at ground-to-space distances on satellites.[68] Sheared Coherent Imaging Photometry (SCIP) or Sheared Beam Imaging (SBI) uses three laser beams, which are displaced spatially and have temporally shifted frequencies, and processes the backscattered speckle patterns to form the images. The common-path geometry of the returned light from the three lasers makes this technique well suited to situations where turbulence is present near the receive aperture. A third imaging technique, Fourier telescopy, is also described in Voelz et al., 2002.[69] While this technique is similar to SCIP in that it involves projecting triplets of laser beams and requires only light-bucket detection on the ground, it uses the information from the fringe pattern formed by the separated transmitters rather than the speckle return and so is not a speckle imaging technique.

Note that speckle photography or interferometry direct techniques can measure the distortion of an object but not its actual 3-D shape. Several techniques take advantage of other "sensor-side" changes to obtain information about the 3-D shape of the object. These techniques may involve taking measurements from multiple aspect angles, at different distances, with different illumination region sizes or at different wavelengths. In each case, it is important to make sure that the parameter being varied does not vary so much that the speckle images completely decorrelate, which ultimately requires a very application-specific system.

Wavelength diversity has been used to gather information about the size, shape, surface and rotation characteristics of objects based on the wavelength dependency of speckle intensity patterns. This can be done statistically (called wavelength decorrelation) to obtain the laser radar cross section of an object, which has been demonstrated with submillimeter range resolution in the laboratory. The decorrelation technique can be implemented using either direct detection, which requires a great deal of processing, or a form of coherent detection using a reference plane or point.[70]

---

[64] A.B. Perimed, "Laser Speckle Contrast Analysis," http://www.perimed-instruments.com/support/theory/laser-speckle-theory.

[65] D.G. Voelz et al., 2002, "Ground-to-space laser imaging: review 2001," *Proc. SPIE* 4489.

[66] P.S. Idell, J.R. Fienup, and R.S. Goodman, 1987, "Image synthesis from nonimaged laser speckle patterns," *Opt. Lett.* 12(11): 858.

[67] J.R. Fienup and P.S. Idell, 1988, "Imaging correlography with sparse arrays of detectors," *Opt. Engr.* 27: 778.

[68] D.G. Voelz et al., 1984, "High-resolution imagery of a space object using an unconventional, laser illumination, imaging technique," *Proc. SPIE* 2312: 202.

[69] Ibid.

[70] L. Shirley and G. Hallerman, 1996, "Applications of tunable lasers to laser radar and 3D imaging," MIT Lincoln Lab., TR-1025.

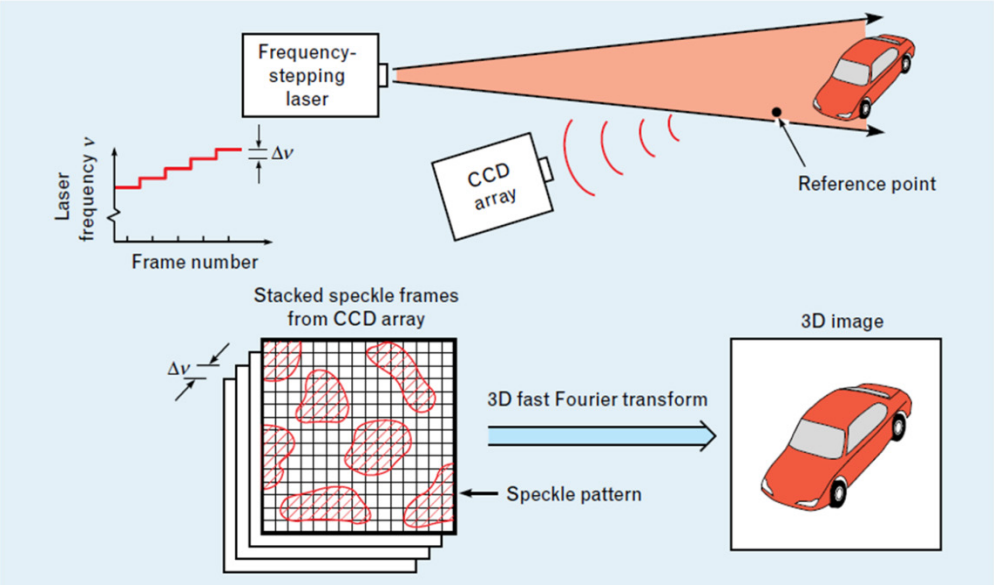

FIGURE 3-14 3-D imaging concept based on speckle-pattern sampling. SOURCE: L. Shirley and G. Hallerman, 1996, "Nonconventional 3-D imaging using wavelength dependent speckle," *Lincoln Lab. J.* 9(2): 153. Reprinted with permission of MIT Lincoln Laboratory, Lexington, Massachusetts.

Wavelength-diversity-based techniques can also be done deterministically. An example of this, called spectral pattern sampling (SPS), is based on the concept of sampling 3-D Fourier space of the scattering object. In the remainder of this section, the focus is on using speckle to obtain 3-D images or contours of a target. This application serves to highlight the key components and limitations of speckle imaging.

Figure 3-14 illustrates the SPS concept. This concept can be thought of as a 3-D extension of the imaging correlography technique mentioned earlier. An object is flood illuminated by a tunable, coherent laser beam and the radiation is sampled at a series of laser frequencies by a detector array. The resulting speckle pattern is measured with a detector array at equally spaced laser frequencies. Individual speckle frames are stacked to form a 3-D data array representing the 3-D Fourier transform squared of the image, and its 3-D Fourier transform yields the 3-D autocorrelation function of the 3-D image of the object.[71] "The reflective reference point near the scattering object causes bright voxels to appear in certain regions of the 3-D array that represent the location in space of scattering cells on the surface of the object" this is similar to digital holography, with a glint near the object acting as the local oscillator.[72] The 3-D image is "formed by recording the location of these bright voxels."[73] The speckle frames are stacked to form a 3-D array. According to Shirley "This array carries information about the location and complex amplitude of the scattering cells located on the surface. If both the amplitude and phase of the speckle pattern were known (as is possible with digital holography), the 3-D Fourier transform of the speckle's complex amplitude would provide the target's 3-D image. However, because phase information is lost in the direct-detection process, the spatial autocorrelation function of the 3-D image is obtained instead."[74]

[71] Ibid.
[72] Ibid.
[73] Ibid.
[74] L.G. Shirley, Reconstruction of 3D Target Images from Wavelength-Dependent Speckle Intensity. Available at http://www.physics.uci.edu/~isis/Yountville/Shirley.pdf. Accessed on March 14, 2014.

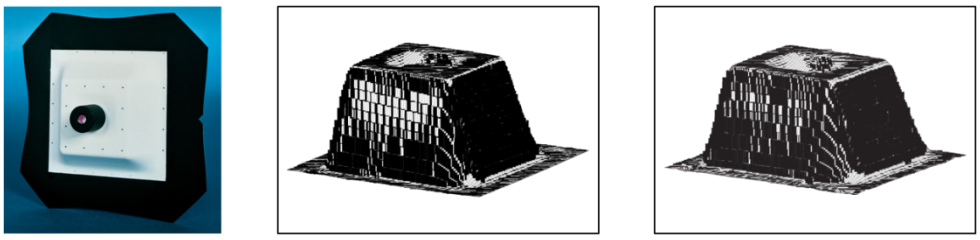

FIGURE 3-15 Comparison between contact measurement and SPS measurement. (a) stamped sheet-metal test object (100 × 100 × 20 mm), (b) coordinate-measuring-machine (CMM) contact measurement of surface contour of test object, and (c) SPS measurement of surface contour of test object. SOURCE: L. Shirley and G. Hallerman, "Nonconventional 3D Imaging Using Wavelength Dependent Speckle," *Lincoln Lab. J.* 9(2), 153-186 (1996). Reprinted with permission of MIT Lincoln Laboratory, Lexington, Massachusetts.

Computational approaches have been developed to reconstruct the 3-D image from the autocorrelation function. Techniques for such a reconstruction have been published.[75,76]

This type of frequency-based imaging technique requires a setup that is designed for the application and target class of interest. The angular resolution depends on both the scan length and detector array size. The larger the volume of the Fourier space sampled, the better the achievable angular resolution.[77] The *x* and *y* resolutions are limited by the detector array size (w_x, w_y) and the distance between the image plane and the detector plane (z_d) and are given by

$$\Delta x = \frac{\lambda_0 z_d}{w_x}, \quad \Delta y = \frac{\lambda_0 z_d}{w_y}$$

The raw range resolution is inversely proportional to the frequency tuning range (B) used and is given by $\Delta z = \frac{c}{2B}$.

For example, for a 10 mm × 10 mm detector array, a distance of 1 m, a wavelength of $\lambda = 0.8$ μm, and a frequency scan of 5 THz, one can achieve resolutions of $\Delta x = \Delta y = 80$ μm and range resolution of $\Delta z = 30$ μm. This method compares well with contact measurements, as shown in Figure 3-15, but does not require contact with the test object and can be collected in a fraction of the time (30 min versus 10 hr for the test in Figure 3-15).

As mentioned above, the speckle imaging system must be specifically designed and tuned for the application of interest. The frequency span must be chosen on the order of the depth of the features of interest. The frequency steps for the 3-D imaging method described above must also be chosen with the range extent of the target in mind in order to avoid aliasing. The larger the range extent of the target, the smaller the frequency step must be.

Laser requirements are an important consideration for all speckle imaging techniques. This technology requires coherent illumination with narrow linewidths in order to record the speckle patterns and avoid decorrelation. The narrow linewidth and tunability required may be considered constraints of the technology, as the current cost of these lasers is significant. However, the coherence constraints of the laser for speckle imaging are much less stringent than for optical coherent detection. In this case, the

[75] J. Fienup et al., 1999, "3-D imaging correlography and coherent image reconstruction," *Proc. SPIE* 3815, *Digital Image Recovery and Synthesis IV*, 60, September 30.

[76] L. Shirley and G. Hallerman, 1996, "Nonconventional 3D imaging using wavelength dependent speckle," *Lincoln Lab. J.* 9(2): 153.

[77] Ibid.

coherence length must be only twice as long as the range extent of the target, and coherence between the laser frequency steps is not required.

Aberrations also impact the performance of speckle imaging. Wavelength aberrations, detector-plane distortions, and depth-of-field aberrations can also degrade the image return. Compensation for these aberrations can be obtained by digital compensation or altering the sensing configuration; in some cases, however, custom optics may be required. Because the speckle return can be small, sensitive cameras with low noise and high frame rates are desired.

The speckle technique above requires a reference beam in order to form the image. Other methods that allow for imaging without the reference beam have been described,[78] but they are limited to the depth of field of conventional imaging. Additionally, significant research has been carried out to obtain 3-D images from the 3-D autocorrelation functions, which do not require a reference point.[79,80,81] Lensless coherent imaging is another alternative method in which a support constraint for a low-resolution image (instead of using a triple intersection of the autocorrelation support) in combination with intensity measurements over a large aperture, can be used to reconstruct fine-resolution images.[82, 83]

Any technique involving speckle imaging requires narrow linewidth, coherent lasers. For applications taking advantage of multiple wavelengths, tunable lasers are also highly desirable. Since the 3-D imaging technique described above requires stepping through multiple wavelengths during the scan, rapidly tunable lasers or multiwavelength lasers will reduce the scan time necessary.

Speckle imaging techniques such as LASCA and ESPI are widely described in the open literature and new developments in these areas would also likely be widely published. Speckle imaging techniques like the one detailed above are still relatively immature and further developments in this research area would also be published. Published literature on the development of laser systems described above would be an indicator. In addition, research on imagery from autocorrelation functions that can enable improved performance may be a sign of work to further develop this area.

This technology allows noninvasive, very-high-resolution imaging. Movements can be visualized on a very small scale (blood flow, etc.) without the need to scan. Speckle imaging can also be used to visualize shear, stress, and strain measurement; vibration mode analysis; and nondestructive testing**.** It can measure complex and irregular shapes having discontinuities and steep sloped surfaces. Formation of 3-D images is possible with simple receivers. Speckle imaging does not require LOs, although they may be used in some applications. Furthermore, although a coherent laser source is required, the required coherence length depends only on the range extent of the target rather than the full round-trip travel time, and coherence between frequency steps is not required.

The 3-D technique detailed above has not been demonstrated at long ranges, although the 2-D imaging correlography technique mentioned earlier has. Turbulence may limit the correlation between measurements; however this technique is less sensitive to atmospheric turbulence than direct imaging. As mentioned previously, these measurements typically require a very controlled setting for optimal performance, especially at high resolution. Another drawback is the high cost of fast-scanning tunable lasers if frequency variation is the phenomenon being exploited. Furthermore, the processing requirements for the sensing mode described above are very stringent, although real-time processing will become feasible as processer technology continues to evolve at a rapid pace. It should be noted that 2-D

[78] L. Shirley and G. Hallerman, 1996, "Nonconventional 3D imaging using wavelength dependent speckle," *Lincoln Lab. J.* 9(2): 153.

[79] Fienup et. al. 1982, "Reconstruction of the support of an object from the support of its autocorrelation," *J. Opt. Soc. Am.* 72(5): 610.

[80] Paxman et al., 1994, "Use of an opacity constraint in three-dimensional imaging," *Proc. SPIE* 2241: 116.

[81] J.M. Fini, 1997, "Three dimensional image reconstruction from Fourier magnitude measurements," BS/MS Thesis, Dept. EECS, MIT, Cambridge, Mass.

[82] J.R. Fienup and A.M. Kowalczyk, 1990 "Phase retrieval for a complex-valued object by using a low resolution image," *J. Opt. Soc. Am. A* 7(3): 450.

[83] H.N. Chapman et al., 2006, "High-resolution ab initio three-dimensional x-ray diffraction microscopy," *J. Opt. Soc. Am. A* 23: 1179.

techniques with a monochromatic laser have been proven experimentally for imaging satellites from ground-based telescopes.[84,85]

Speckle technology is currently used in biomedical applications and for nondestructive testing for stress/strain, vibration mode analysis, etc. There are a number of potential applications of wavelength-dependent speckle in advanced manufacturing, machine vision, industrial inspection, robotics, and dimensional metrology. These methods can also be used to obtain laser radar cross sections, surface slope maps, or even 2-D and 3-D images of targets.

In theory, because the range resolution does not degrade with range, it could be attractive for applications involving long-range imaging.

Conclusion 3-7: Speckle imaging can be used to measure 3-D shape, surface roughness, displacements, vibrations, and strains.

LADAR USING FEMTOSECOND SOURCES

As discussed in more detail in Chapter 4, broad-gain-bandwidth lasers such as those based on the solid state material Ti:sapphire can generate (ultrashort) pulses with femtosecond duration, through the process of mode-locking. Amplification of the pulses to high energies is possible through the technique of chirped-pulse amplification (CPA), and current technology has made it possible to generate pulse energies in the Joule range, with durations on the order of 20 fs. The first ultrashort-pulse lasers were based on liquid-dye gain media, required considerable efforts to maintain operation, and were totally unsuited for use outside the laboratory. With the development of Ti:sapphire and other solid-state-laser-based ultrafast lasers, it became possible to build systems that could be used for ladar and lidar systems, and deployed in mobile laboratories for field measurements. Further developments now under way of fiber-format solid state lasers promise even more possibilities for field-deployable ultrafast lasers.

The simplest application for an ultrafast source would be in single-pulse, time-of-flight ranging ladar systems, from one to three dimensions. Since a 20-fs pulse in air has a spatial dimension of 6 μm, it could provide range resolution, with appropriate processing, at the multimicron level. Unfortunately, sensitive detectors with the necessary response time do not yet exist. And even without appropriate detectors, there are two major limitations to long-range, time-of-flight sensing with ultrafast sources. First, the spectral dispersion of the atmosphere, normally not a concern for nanosecond-duration pulses, cannot be ignored with femtosecond pulses and leads to significant pulse stretching for light at short wavelengths. This can be countered, in principle, by adding frequency chirp to the transmitted pulse that is undone by the atmosphere in the round-trip path but that presumes a good knowledge of the distance to the target and a full understanding of the exact atmospheric conditions along the path. Another limitation is in the pulse energy, since nonlinear atmospheric effects, such as the nonlinear refractive index of the atmosphere, stimulated Raman scattering, and self-focusing, can distort the pulse in a nontrivial way. This is discussed in more detail in the next section, where the use of the nonlinear effects is a benefit rather than a limitation.

Femtosecond lasers can be used for precision range measurements, though other techniques are less subject to their limitations for simple, single-pulse, time-of-flight measurements. Here the frequency-spectrum characteristics of the mode-locked laser are employed. One approach uses the femtosecond source with a conventional swept-frequency (FM) ladar system as a means to accurately determine the frequency of the ladar source as a function of time, and thus provide higher accuracy in the lidar measurement. The femtosecond laser spectrum consists of a "comb" of frequencies centered about the

[84] D.G. Voelz, S.D. O'Keefe, J.D. Gonglewski, D.B. Rider, and K.J. Schulze, 1994, "High-resolution imagery of a space object using an unconventional, laser illumination, imaging technique," *Proc. SPIE* 2312.

[85] D.G. Voelz, J.F. Belsher, L. Ulibarri, and V. Gamiz, 2002, "Ground-to-space laser imaging: Review 2001," *Proc. SPIE* 4489.

laser peak wavelength. The frequency spacing can be locked to an RF standard with 10^{-8} accuracy. By heterodyning a portion of the FM ladar transmitter output against the comb of the femtosecond source, one gets a precise sample of the sweep at frequency intervals set by the comb frequency spacing, a better and more accurate measurement than, say, using an optical interferometer as a sampling means. Operation of an FM lidar around 1,560 nm, based on an external-cavity diode laser with 1-msec-duration, 1-THz-span sweeps is claimed to result in a bandwidth-limited 130-μm range resolution and 100 nm accuracy, with 1-msec update rates.[86]

Clearly, use of a continuous wave (CW) comb as the source for a ladar, while lowering the peak power and greatly updating the acquisition rate, would suffer from the range ambiguity set by the pulse rate, 1.5 m for a typical 100-MHz-rate comb. One approach to precision ranging with a longer ambiguity region employs one femtosecond laser as the transmitter and a second femtosecond laser as the LO in a heterodyne-detection ladar.[87]

> One comb serves as the "signal" source and samples a distance path defined by reflections off a target and reference plane. The second comb serves as a broadband local oscillator (LO), and recovers range information in an approach equivalent to linear optical sampling (that is, a heterodyne cross-correlation between the signal and LO). The heterodyne detection provides shot-noise limited performance so that even weak return signals can be detected and the information in the carrier phase is retained.

The pulse rate of the LO is slightly offset from that of the transmitter. Figure 3-16 shows the time-domain picture, for nominal 100-MHz-rate pulses, with a 5.2 kHz difference in the rates. Provided that the two combs are coherent, locked to the same optical standard and with the RF rates held to very high precision, the system can, through just the time-of-flight measurement, provide range precision of 3 μm over the 1.5-m ambiguity range with a 200-μs measurement time, and with determination of the optical phase of the return, provides a 5-nm precision over a 60 ms averaging time.

Most significantly, if the transmit and LO combs are interchanged, it is possible, by calculating the difference in the ranges measured, to determine the integer multiple of the number of pulses between the transmitter and target and thereby increase the range ambiguity to the pulse propagation velocity divided by twice the 5-kHz difference frequency, extending the ambiguity range, for the case shown, to 30 km.

In a real atmosphere rather than the vacuum of space, the precisions predicted would not be possible due to atmospheric fluctuations, which limit measurement precisions to 1 part in 10^7.[88] Simpler dual-comb ladar systems, based on free-running mode-locked Er:fiber lasers, can achieve a ranging precision of 2 μm in 140-μs acquisition time, increasing to 0.2 nm with a 20-ms time,[89] with a 1-m range ambiguity. The latter can be eliminated through, say, an adjunct, nanosecond-pulse time-of-flight ranging system run over the same sensing path. Another approach to a simplified dual-comb system encodes the transmitted pulse with a pseudo-random binary modulation, and in one demonstration a pair of 100-MHz combs with variable timing between one comb and the other were employed, and the coding allowed expansion of the ambiguity range to 190.5 m.[90]

---

[86] E. Baumann, F.R. Giorgetta, I. Coddington, L.C. Sinclair, K. Knabe, W.C. Swann, and N.R. Newbury, 2013, "Comb-calibrated frequency-modulated continuous-wave ladar for absolute distance measurements," *Opt. Lett.* 38: 2026.

[87] I. Coddington, W.C. Swann, L. Nenadovic, and N.R. Newbury, 2009, "Rapid and precise absolute distance measurements at long range," *Nature Photonics* 3: 351.

[88] N. Bobroff, 1993, "Recent advances in displacement measuring interferometry," *Meas. Sci. Technol.* 4: 907.

[89] T.-A. Liu, N.R. Newbury, and I. Coddington, 2011, "Sub-micron absolute distance measurements in sub-millisecond times with dual free-running femtosecond Er fiber-lasers," *Opt. Exp.* 19: 18501.

[90] M. Godbout, J.-D. Deschenes, and J. Genest, 2004, "Spectrally resolved laser ranging with frequency combs," *Opt. Exp.* 18: 15981.

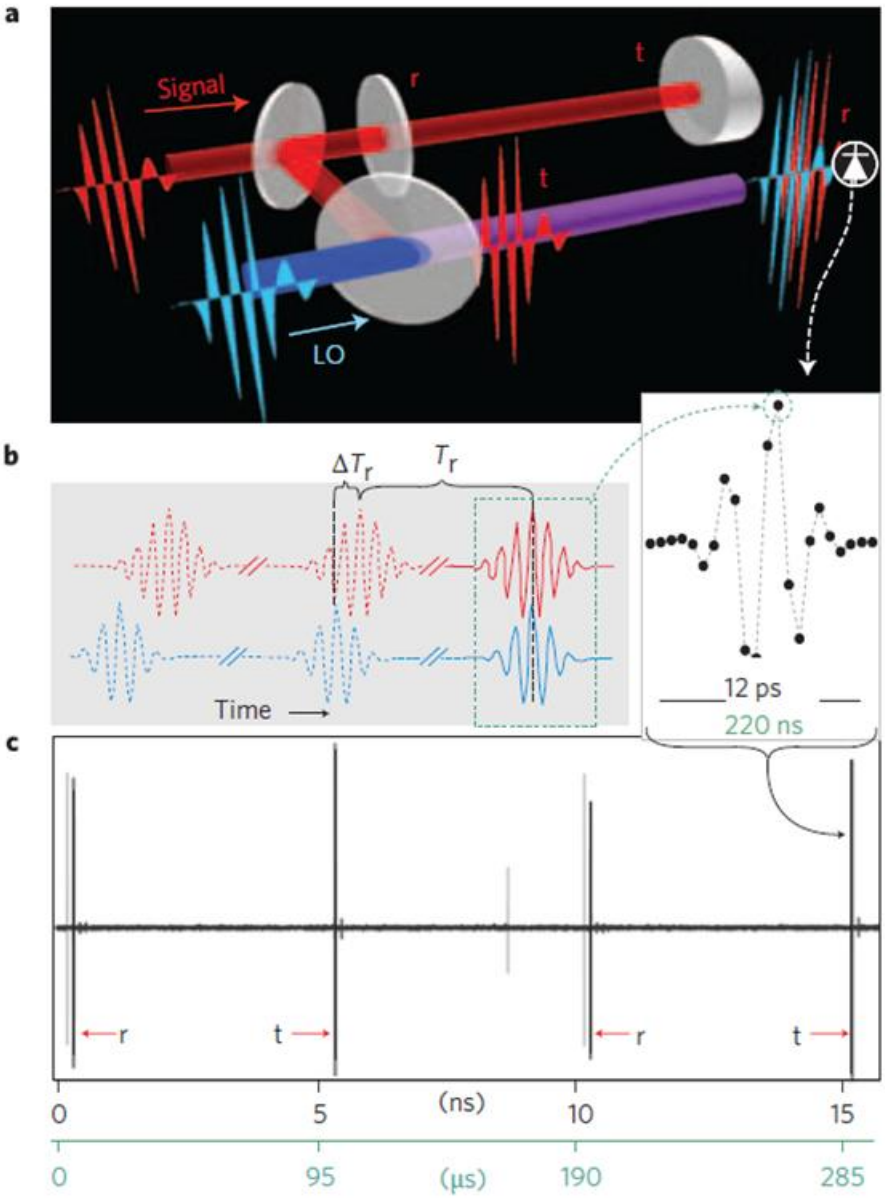

FIGURE 3-16 "Dual-comb ranging concept. (a) A high-repetition-rate "signal" source transmits pulses that are reflected from two partially reflecting planes (glass plates): the reference (r) and the target (t). The reference is a flat plate and yields two reflections, the first of which is ignored. Distance is measured as the time delay between the reflections from the back surface of the reference flat and the front of the target. (b) The signal pulses are detected through linear optical sampling against an LO. The LO generates pulses at a slightly different repetition rate. Every repetition period (T_r), the LO pulse "slips" by ΔT_r relative to the signal pulses and samples a slightly different portion of the signal. Consecutive samples of the overlap between the signal and LO yield a high-range-resolution measurement of the returning target and reference pulses. Actual data are shown on the right side, where each discrete point corresponds to a single digitized sample and only the immediate overlap region is shown. (c) The measured voltage out of the detector in both real time (lower scale) and effective time (upper scale) for a target and reference plane separated by 0.76 m. A full "scan" of the LO pulse across the signal pulse is accomplished every ~200 ms in real time and every ~10 ns in effective time. Two such scans are shown to illustrate the fast, repetitive nature of the measurement. Also seen are two peaks in grey which are spurious reflections of the launch optics." SOURCE: Reprinted with permission from Macmillan Publishers Ltd :Nature Photonics. I. Coddington, W. Swann, L. Nenadovic, and N. Newbury, 2009, "Rapid and precise absolute distance measurements at long range," *Nature Photonics* 3(6).

Conclusion 3-8: While the short pulses that can be generated with femtosecond lasers could allow the measurement of range to be improved by 5 orders of magnitude compared to conventional nanosecond-based sensors, dispersion and fluctuations in the atmosphere prevent such a dramatic result, and nonlinearities in the atmosphere limit the pulse energy.

Conclusion 3-9: In space, use of the precision frequency combs generated by stabilized femtosecond lasers would allow nanometer-precision range measurements with 30-km-level range ambiguities.

Conclusion 3-10: Femtosecond-laser-based ranging systems would enable deployment of satellite arrays with optical-wavelength-level measurement and control of spacings, enabling distributed-aperture sensors.

Recommendation 3-1: Further development of femtosecond-laser ranging systems is of importance to the development of space-based sensors with significantly improved performance and should be encouraged in the United States and monitored in other countries.

As is discussed in more detail below, if the single-pulse peak intensity from a femtosecond laser exceeds a certain level in the atmosphere, self-focusing due to the nonlinearity of the molecules in the atmosphere leads to the formation of a "filament" of light, where the beam size remains constant over a wider range of distance than allowed by diffraction. Laser-induced breakdown spectroscopy (LIBS), discussed in Chapter 2, can obtain an extension in range to several hundred meters through the use of filaments, and work is ongoing to determine the efficacy of the technique, compared to conventional, nanosecond-duration-excited LIBS.[91, 92]

In Chapter 4 the fundamentals of the development of supercontinuum (SC) sources based on optical fibers are discussed. The sources provide very broad, nearly structureless optical spectra similar to that from incandescent (blackbody-like) sources, but with many orders of magnitude brighter. For path-averaged differential absorption lidar (DIAL) systems, SC sources provide more data than limited-tuning-range lasers, and especially in complex atmospheric environments, provide much more data that can help to identify and quantify different species. At present, sources based on silica fibers are readily available, generating wavelengths extending to 2,450 nm, but there is active development of sources based on other types of glass fibers to generate wavelengths extending in to the long-wave IR region.

Figures 3-17 and 3-18 show the experimental setup and atmospheric data generated, respectively, for one SC-based, path-averaged system, where the source is based on a Q-switched microchip laser and commercial photonic-crystal fibers and detection is done through use of a commercial optical spectrum analyzer (OSA) to determine the path transmission as a function of wavelength. Through the maximum likelihood estimation (MLE) method, the calculated concentration levels of species are found to converge rapidly to a narrow error level with an increase in the number of matched individual absorption lines from

[91] Ph. Rohwetter, K. Stelmaszczyk, L. Woste, R. Ackermann, G. Mejean, E. Salmon, J. Kasparian, J. Yu, and J.-P. Wolf, 2005, "Filament-induced remote surface ablation for long range laser-induced breakdown spectroscopy operation," *Spectrochimica Acta Part B* 60: 1025.

[92] C.G. Brown, R. Bernatha, M. Fishera, M.C. Richardson, M. Sigman, R.A. Walters, A. Miziolek, H. Bereket, and L.E. Johnson, 2006, "Remote femtosecond laser induced breakdown spectroscopy (LIBS) in a standoff detection regime," in *Enabling Technologies and Design of Nonlethal Weapons*, G.T. Shwaery, J.G. Blitch, and C. Land, ed., *Proc. of SPIE* 6219: 62190B.

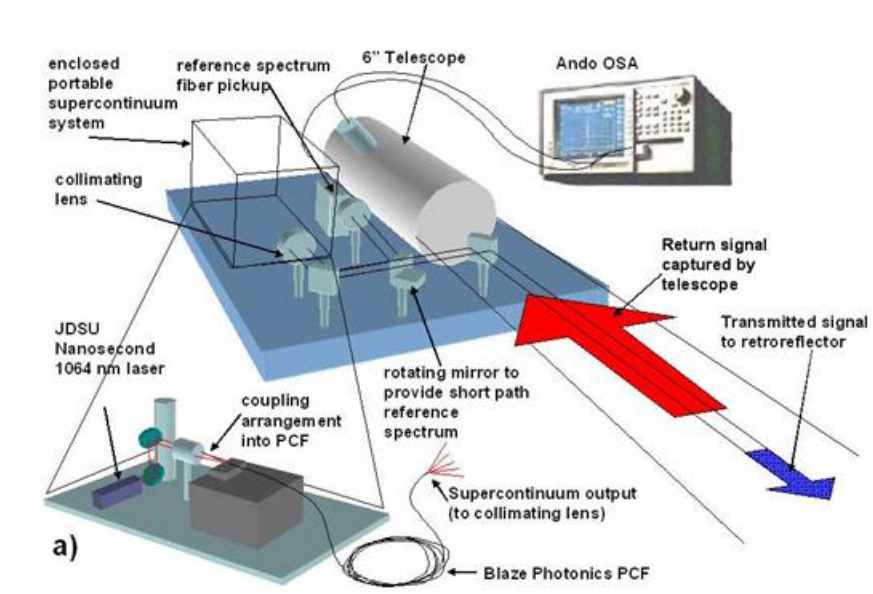

FIGURE 3-17 Experimental setup for path-averaged DIAL based on fiber SC. SOURCE: D. Brown, K. Zhiwen Liu, and C. Philbrick, "Long-path supercontinuum absorption spectroscopy for measurement of atmospheric constituents," 2008, *Optics Express* 16 (12): 8457.

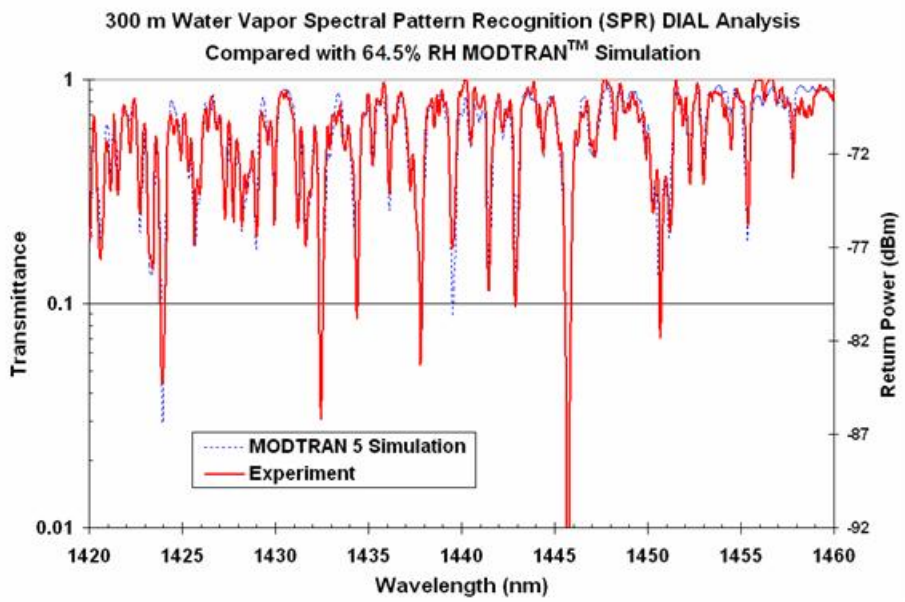

FIGURE 3-18 Data generated by the system shown in Figure 3-17 over a 300-m outdoor path with compared to standard model of atmospheric transmission for water vapor. SOURCE: D. Brown, K. Zhiwen Liu, and C. Philbrick, 2008, "Long-path supercontinuum absorption spectroscopy for measurement of atmospheric constituents," *Optics Express* 16 (12): 8457.

10 to 500, showing the advantage of SC-based DIAL over conventional DIAL.[93] Further measurements have shown the ability, over a 600-m path, with wavelengths in the 760-nm region, to calculate the oxygen concentration to an accuracy of 2×10^{-4}.[94]

Another approach to measurements with SC sources has employed a Fourier-transform spectrometer to analyze the return signals with a processing time of several seconds.[95]

Even faster data processing would be possible through the combination of a grating or other dispersive element along with a multielement detector, although the ultimate spectral resolution would still be less than possible with Fourier-transform techniques.

Use of SC sources for spectroscopy-based ranging systems (hyperspectral lidar) has been demonstrated for the discrimination of ground reflections from trees and inorganic material, while also generating range information, to date at short (10-m) ranges but with the expectation of eventual deployment to longer-range, airborne platforms.[96] Simulations on the use of SC sources for aerosol backscatter illustrate the possible advantages of being better able to characterize the aerosols by determining the wavelength variation of the scattered signal.[97]

More complex and higher-performance DIAL systems are based on the frequency combs generated by mode-locked lasers, with techniques similar to that used for the precision ranging described in the preceding section. Consider two frequency combs that are slightly offset in their repetition rate and mixed together on a detector. The resultant signal also contains a comb of frequencies, with each frequency representing a specific optical comb frequency. The optical combs have been essentially converted into combs that can be readily processed by high-speed electronics, and a Fourier-transform of the detected signal can extract the intensity information for each optical frequency. If an absorbing gas is placed in one (or both) of the beam paths, the Fourier transform will exhibit the absorption features of the gas in the wavelength region covered by the comb.

In one effort, the outputs of two 10-20 fs, 800-nm Ti:sapphire lasers were focused into GaSe crystals optimized to generate difference-frequency output in the 8-12 μm "fingerprint" region of the IR.[98] Figure 3-19a shows the detected signal with and without NH_3 in the beam path, with Figure 3-19b showing the resultant Fourier-transformed spectra, with 2 cm^{-1} of resolution, along with a comparison of data from a conventional, interferometer-based Fourier-transform spectrometer. The notable result with the comb-based source is that the spectral data resulted from a 70-μs signal, compared to the 60 s required for the conventional spectrometer, thus pointing the way to rapid data acquisition, as might be desired for a system that scans large regions for the presence of gases. Another system employed two fully stabilized combs locked to a common frequency standard, with a 1 kHz difference in pulse rate, and was able to fully resolve 155,000 different comb lines in the 1550-nm region to perform high-resolution spectroscopy of HCN gas.[99]

---

[93] D.M. Brown, K. Shi, Z. Liu, and C.R. Philbrick, 2008, "Long-path supercontinuum absorption spectroscopy for measurement of atmospheric constituents," *Opt. Express* 16: 8457.

[94] P.S. Edwards, D.M. Brown, A.M. Wyant, Z. Liu, and C.R. Philbrick, 2009, "Atmospheric absorption spectroscopy using supercontinuum lasers," in *Conference on Lasers and Electro-Optics*/International Quantum Electronics Conference, OSA Technical Digest (CD), Optical Society of America, CFJ3.

[95] J. Mandon, E. Sorokin, I.T. Sorokina, G. Guelachvili, and N. Picqué, 2008, "Supercontinua for high-resolution absorption multiplex infrared spectroscopy," *Opt. Lett.* 33: 285.

[96] Y. Chen, E. Räikkönen, S. Kaasalainen, J. Suomalainen, T. Hakala, J. Hyyppä, and R. Chen, 2010, "Two-channel hyperspectral LiDAR with a supercontinuum laser source," *Sensors*, 10: 7057.

[97] S. Kaasalainen, T. Lindroos, and J. Hyyppä, 2007, "Toward hyperspectral lidar: measurement of spectral backscatter intensity with a supercontinuum laser source," *IEEE Geoscience And Remote Sensing Letters* 4: 211.

[98] A. Schliesser, M. Brehm and F. Keilmann, 2005, "Frequency-comb infrared spectrometer for rapid, remote chemical sensing," *Opt. Express* 13: 9029.

[99] I. Coddington, W.C. Swann, and N.R. Newbury, 2008, "Coherent multiheterodyne spectroscopy using stabilized optical frequency combs," *Phys. Rev. Lett.* 100: 013902.

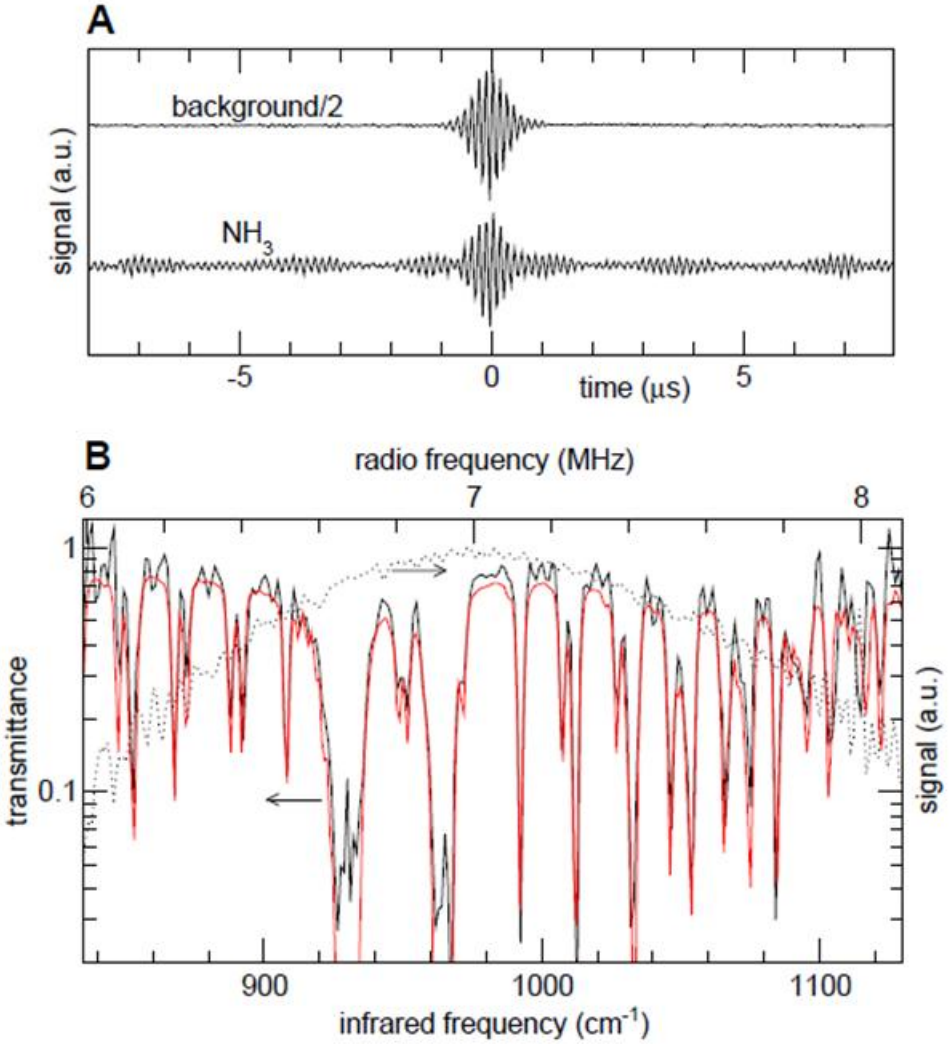

FIGURE 3-19 "(A) Single interferograms from dual-comb system with 70-μs acquisition time window each (16 μs displayed, t = 0 arbitrarily chosen at window center), without (background) and with a gas absorption cell (NH_3). (B) IR spectra calculated from (A) by Fourier transformation, vs. both RFand IR frequency scales; background spectrum with maximum normalized to 1 (dotted), NH_3 cell transmittance (full black). For comparison the red trace gives the transmittance of conventional Fourier transform infrared (FTIR) obtained at 2 cm^{-1} resolution and 60 s acquisition time, with 32 spectra averaged." SOURCE: A. Schliesser, M. Brehm and F. Keilmann, 2005, "Frequency-comb infrared spectrometer for rapid, remote chemical sensing," *Opt. Express* 13: 9029.

An alternative to dual-comb systems is the National Institute of Standards (NIST)-developed, high-resolution, crossed-spectral disperser to project the various frequency-comb modes from a single laser onto a two-dimensional digital camera. The NIST design employs a side-entrance etalon called a virtually imaged phased array (VIPA) disperser that provides ~1 GHz resolution in the visible spectral range (and a resolution of ~500 MHz at 1,550 nm).[100] When combined with a lower-dispersion grating in the orthogonal spatial direction, 5–10 THz of bandwidth can be captured in a single measurement taking a few milliseconds.[101] The technique has shown the ability to detect absorption levels in the 1×10^{-9} cm^{-1} range in the 1,500-nm region, with long-pass cells or cavity-enhanced systems. Shifting to mid-IR combs

[100]S. Diddams, 2010 "The evolving optical frequency comb [Invited]," *J. Opt. Soc. Am*. B 27: B51-B62.
[101] S.A. Diddams, 2010, "The evolving optical frequency comb [Invited]," *J. Opt. Soc. Am* B27: 851.

would allow an increase of one to three orders of magnitude in sensitivity, providing, theoretically, parts-per-trillion sensitivities for certain gases.[102]

One of the most dramatic applications of femtosecond lasers has been in the use of atmospheric filaments generated by the launching of high-power femtosecond lasers into the atmosphere. Lidar applications include aerosol measurements, DIAL and laser-induced fluorescence (LIF), discussed below.

A key concept behind filament formation is the self-focusing effect[103] for laser beams, where propagation in a medium with a positive value of nonlinear refractive index leads to a collapse of the beam diameter from the intensity-created positive lens in the medium. Analysis of the effect shows that the threshold for it to occur is a function of the peak power in the beam, not the intensity. In the following discussions, all of the results apply to the use of 800-nm Ti:sapphire lasers, which have predominated in filament studies.

In atmospheric-pressure air, the threshold power for self-focusing has been measured to be about 10 GW for a 42-fs laser pulse, gradually decreasing to about 5 GW for a chirped pulse with duration longer than 200 fs.[104] The change is due to the frequency-dependent nature of the atmospheric nonlinear refraction. Self-focusing alone would not result in filaments but would instead bring about optical breakdown, as is commonly seen inside solids as damage spots.

In air, filaments appear as a result of the dynamic balance between self-focusing and defocusing by the plasma produced from the air molecules.[105,106,107] In contrast to self-focusing, which is set by the peak power of the beam, the plasma defocusing does depend on intensity of the beam, and if it is too small, the filament will not form. On the other hand, the equilibrium places an upper limit on the laser intensity inside the filament core to about 5×10^{13} W/cm^2, so called "intensity clamping."[108] The single-filament beam diameter is about 100 μm, and filaments can typically extend tens to hundreds of meters, orders of magnitude longer than the Rayleigh length for a conventional beam of that diameter. Focusing to higher intensities results in the formation of multiple filaments in one beam. The peak intensity inside a single filament is high enough to dissociate/ionize other gas molecules, generate higher harmonics, induce other parametric processes as well as generate THz radiation,[109] explode dust particles and aerosols or induce partial breakdown on solid targets,[110] as noted in the prior section on femtosecond LIBS.

Filamentation can be formed at a far distance in an adverse atmospheric environment[111] and is only weakly disturbed by turbulence.[112] It has been demonstrated that the low-intensity background

---

[102] M.J. Thorpe, D. Balslev-Clausen, M.S. Kirchner, and J. Ye, 2008, "Cavity-enhanced optical frequency comb spectroscopy: Application to human breath analysis," *Opt. Express* 16: 2387.

[103] R.Y. Chiao, E. Garmire, and C.H. Townes, 1964, "Self-trapping of optical beams," *Phys. Rev. Lett.* 13: 479.

[104] W. Liu, and S.L. Chin, 2005, "Direct measurement of the critical power of femtosecond Ti:sapphire laser pulse in air," *Opt. Express* 13: 5750.

[105] A. Couairon and A. Mysyrowicz, 2007, "Femtosecond filamentation in transparent media," *Phys. Rep.* 441: 47.

[106] L. Berge, S. Skupin, R. Nuter, J. Kasparian, and J.-P. Wolf, 2007, "Ultrashort filaments of light in weakly ionized, optically transparent media," *Rep. Prog. Phys.* 70: 1633.

[107] V.P. Kandidov, S.A. Shlenov, and O.G. Kosareva, 2009, "Filamentation of high-power femtosecond laser radiation," *Quant. Electron.* 39: 205.

[108] J. Kasparian, R. Sauerbrey, and S.L. Chin, 2000 "The critical laser intensity of self-guided light filaments in air," *Appl. Phys.* B71: 877.

[109] C. D'Amico, A. Houard, M. Franco, B. Prade, A. Mysyrowicz, A. Couairon, and V.T. Tikhonchuk, 2007, "Conical forward THz emission from femtosecond-laser-beam filamentation in air," *Phys. Rev. Lett.* 98: 235002.

[110] K. Stelmaszczyk, P. Rohwetter, G. Méjean, J. Yu, E. Salmon, J. Kasparian, R. Ackermann, J.-P. Wolf, and L. Wöste, 2004, "Long-distance remote laser-induced breakdown spectroscopy using filamentation in air," *Appl. Phys. Lett.* 85: 3977.

[111] G. Méchain, G. Méjean, R. Ackermann, P. Rohwetter, Y.-B. André, J. Kasparian, B. Prade, K. Stelmaszczyk, J. Yu, E. Salmon, W. Winn, L.A. Schlie, A. Mysyrowicz, R. Sauerbrey, L. Wöste, and J.-P. Wolf, 2005, "Propagation of fs TW laser filaments in adverse atmospheric conditions," *Appl. Phys.* B80: 785.

[112] R. Salame, N. Lascoux, E. Salmon, R. Ackermann, J. Kasparian, and J.-P. Wolf, 2007, "Propagation of laser filaments through an extended turbulent region" *Appl. Phys. Lett.* 91: 171106.

energy (energy reservoir) surrounding filaments can play an important role in their formation.[113] When particles in the propagation path like water droplets, snow or dust, block the filament, the energy in the reservoir will refill the filament core (replenishment).[114] Such filamentation properties clarify why filaments can be formed and propagate under adverse atmospheric conditions such as rain, compared to the linear propagation of the beam. The background can contain up to 90 percent of the pulse energy, which is beneficial for maintaining the filament formation.[115] Calculations of the spatial evolution of filaments are complicated by the high level of nonlinearities and provide a major challenge to numerical modeling.

Filaments in the atmosphere, in common with high-intensity propagation of light in fibers, will generate SC emission, from the UV to the IR. The generation of the SC is assumed to be primarily the result of spectral broadening of the laser energy by self-phase modulation. Emission in the UV is enhanced via third-harmonic generation in the atmosphere, which mixes with the SC generated by self-phase-modulation of the fundamental.[116] Figure 3-20 shows the laboratory-measured spectra of the SC light for different levels of frequency chirp in the pulse as well as different pulsewidths.[117] The pulse chirp can be controlled to correct for atmospheric dispersion over a given path so the pulsewidth is minimized (and peak power maximized) at the desired location in the atmosphere. Subsequent measurements of the intensity of backscattered light from atmospheric filaments showed an enhancement in the amount of light beyond that expected by Rayleigh scattering, and this was proposed to be the result of longitudinal index variations in the filament, acting as a Bragg reflector to the generated SC.[118] Spectral measurements of the SC produced over a long vertical path in the atmosphere and reflected from a cloud at 4 km indicated a much higher level of energy in the 1,000-2,000-nm region than indicated by Figure 3-20, by about an order of magnitude, attributed to the much longer generation path than that in the laboratory.[119]

The ability to generate high intensities at long distances and create UV-IR SC light at some distance above ground suggested applications of filaments to various lidar applications, and led to the construction and deployment of the Teramobile lidar system,[120] built in 2000-2001 as part of a French-German effort. The Ti:sapphire laser (supplied by Thales in France) has the specifications listed in Table 3-2, and the receiver employs a 40-cm-diameter telescope, along with a variety of detectors and a 50-cm spectrograph for spectral analysis. A plan drawing and photograph of the system appear in Figure 3-21. Figure 3-22 shows a nighttime photograph of the SC light generated by the Teramobile system, and Figure 3-23 provides both aerosol and DIAL data generated from the system, the latter showing H_2O

[113] M. Mlejnek, E.M. Wright, and J.V. Moloney, 1999, "Moving-focus versus self-waveguiding model for long-distance propagation of femtosecond pulses in air," *IEEE J. Quant. Electron.* 35: 1771.

[114] F. Courvoisier, V. Boutou, J. Kasparian, E. Salmon, G. Méjean, J. Yu, and J.-P. Wolf, 2003, "Ultra-intense light filaments transmitted through clouds," *Appl. Phys. Lett.* 83: 213.

[115] W. Liu, F. Théberge, E. Arévalo, J.F. Gravel, A. Becker, and S.L. Chin, 2005, "Experiment and simulations on the energy reservoir effect in femtosecond light filaments," *Opt. Lett.* 30: 2602.

[116] L. Bergé, S. Skupin, G. Méjean, J. Kasparian, J. Yu, S. Frey, E. Salmon, and J.P. Wolf, 2005, "Supercontinuum emission and enhanced self-guiding of infrared femtosecond filaments sustained by third-harmonic generation in air," *Phys. Rev.* E71: 016602.

[117] J. Kasparian, R. Sauerbrey, D. Mondelain, S. Niedermeier, J. Yu, J.-P. Wolf, Y.-B. André, M. Franco, B. Prade, S. Tzortzakis, A. Mysyrowicz, A.M. Rodriguez, H. Wille, and L. Wöste, 2000, "Infrared extension of the supercontinuum generated by femtosecond terawatt laser pulses propagating in the atmosphere," *Opt. Lett.* 25: 1397.

[118] J. Yu, D. Mondelain, G. Ange, R. Volk, S. Niedermeier, J.-P. Wolf, J. Kasparian, and R. Sauerbrey, 2001, "Backward supercontinuum emission from a filament generated by ultrashort laser pulses in air," *Opt. Lett.* 26: 533.

[119] G. Mejean, J. Kasparian, E. Salmon, J. Yu, J.-P. Wolf, R. Bourayou, R. Sauerbrey, M. Rodriguez, L. Woste, H. Lehmann, B. Stecklum, U. Laux, J. Eisloffel, A. Scholz, and A.P. Hatzes, 2003, "Towards a supercontinuum-based infrared lidar," *Appl. Phys.* B77: 357.

[120] H. Wille, M. Rodriguez, J. Kasparian, D. Mondelain, J. Yu, A. Mysyrowicz, R. Sauerbrey, J.-P. Wolf, and L. Woste, 2002, "Teramobile: A mobile femtosecond-terawatt laser and detection system," *Eur. Phys. J. AP* 20: 183.

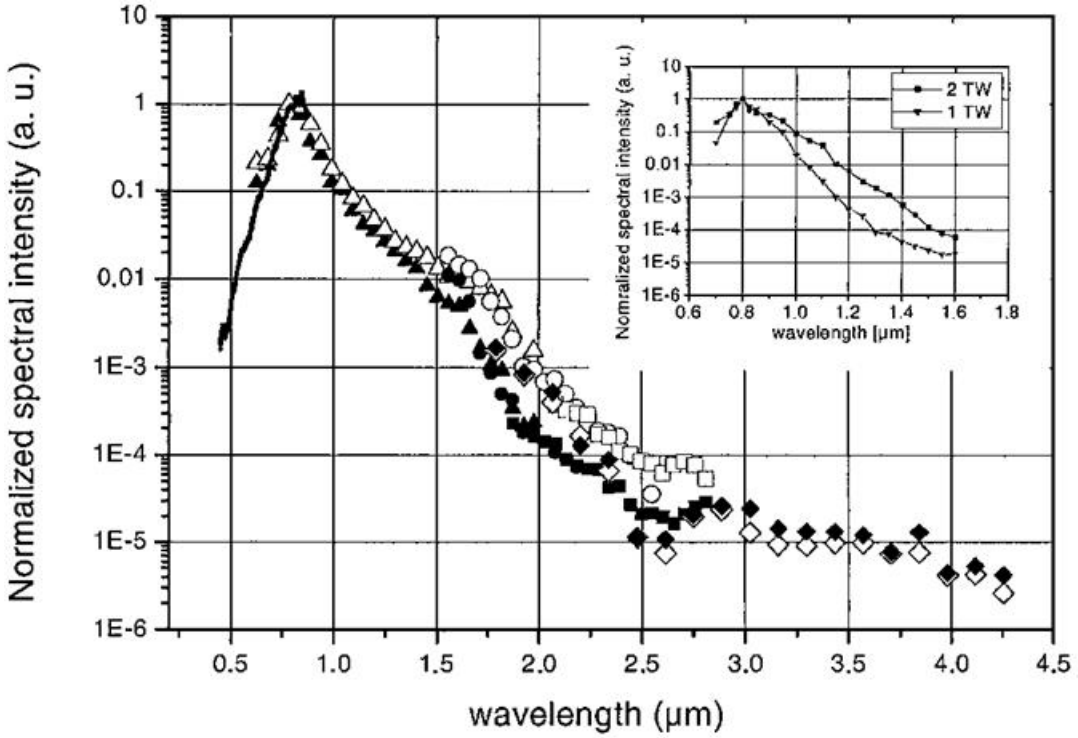

FIGURE 3-20 "Measured spectrum of the supercontinuum generated in the center of the beam by 2-TW laser pulses. The results are shown for two different chirp settings that correspond to an initial pulse duration of 35 fs without chirp after the compressor (filled symbols) and a 55-fs initial pulse duration with negative chirp after the compressor (open symbols). Inset, spectrum of the SC generated in the center of the beam by 100 fs pulses as a function of pulse power value (200 and 100 mJ for 2 and 1 TW, respectively). The two curves have the same normalization factor (24)." SOURCE: J. Kasparian, R. Sauerbrey, D. Mondelain, S. Niedermeier, J. Yu, J.-P. Wolf, Y.-B. André, M. Franco, B. Prade, S. Tzortzakis, A. Mysyrowicz, A. M. Rodriguez, H. Wille, and L. Wöste, 2000, "Infrared extension of the supercontinuum generated by femtosecond terawatt laser pulses propagating in the atmosphere," *Opt. Lett.* 25: 1397.

TABLE 3-2 Specifications for Ti:sapphire Laser in the Teramobile Lidar System

Center wavelength	793 nm
Bandwidth	16 nm
Pulse energy	350 mJ
Pulse duration	70 fs ($sech^2$)
Peak power	5 TW
Repetition rate	10 Hz
Output beam diameter	50 mm
Chirped pulse duration	70 fs to 2 ps, positive or negative chirp
Energy stability	2.5 percent RMS over 400 shots
Dimensions	3.5 × 2.2 m

SOURCE: H. Wille, M. Rodriguez, J. Kasparian, D. Mondelain, J. Yu, A. Mysyrowicz, R. Sauerbrey, J.-P. Wolf, and L. Woste, 2002, "Teramobile: A mobile femtosecond-terawatt laser and detection system," *Eur. Phys. J. AP* 20: 183. With kind permission of The European Physical Journal (PEG).

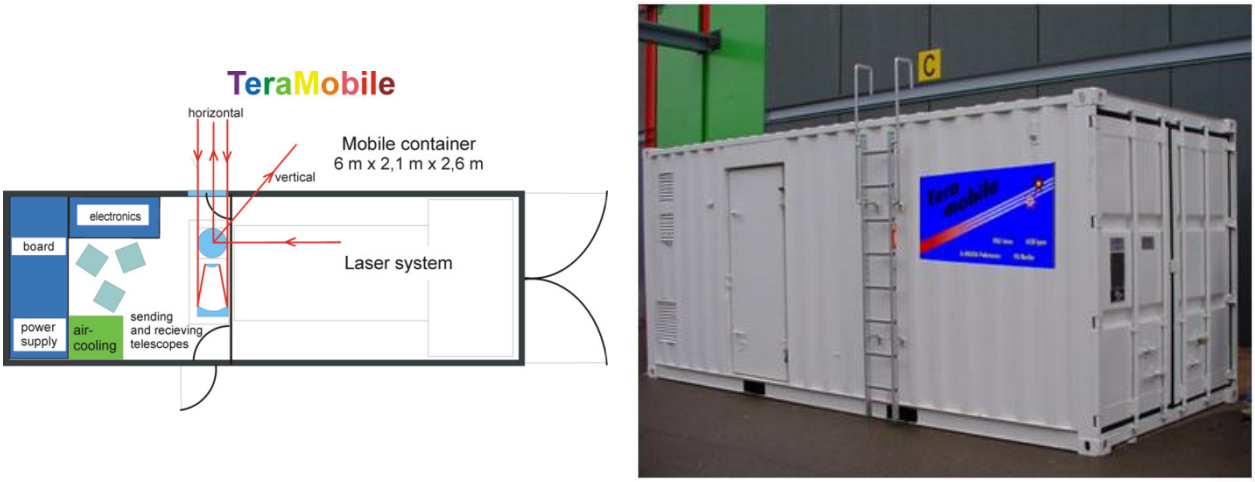

FIGURE 3-21 Plan for Teramobile lidar system (*left*) and photograph of system (*right*), built in a standard ISO 20 ft sea container. SOURCE: Copyright Teramobile. Used with permission.

FIGURE 3-22 Nighttime photograph of SC light generated by a vertically directed beam from the Teramobile system. SOURCE: Copyright Teramobile. Used with permission.

vapor and O_2 absorption lines.[121] In principle, spectral analysis of the near-IR data can provide (through the intensity of the water absorption lines) a probe of humidity and (through analysis of the spectral distribution of O_2 lines around 760 nm) a probe of atmospheric temperature. Under some conditions, SC light from as high as 20 km has been observed.

Other applications of the Teramobile system have been in aerosol characterization, through the simultaneous generation of backscatter from a wide variety of wavelengths. In particular, the broad SC spectrum allows probing clouds and determining not only size distributions of the aerosols, but also,

[121] J. Kasparian, M. Rodriguez, G. Méjean, J. Yu, E. Salmon, H. Wille, R. Bourayou, S. Frey, Y.-B. Andre, A. Mysyrowicz, R. Sauerbrey, J.-P. Wolf, and L. Wöste, 2003, "White-light filaments for atmospheric analysis," *Science* 301: 61.

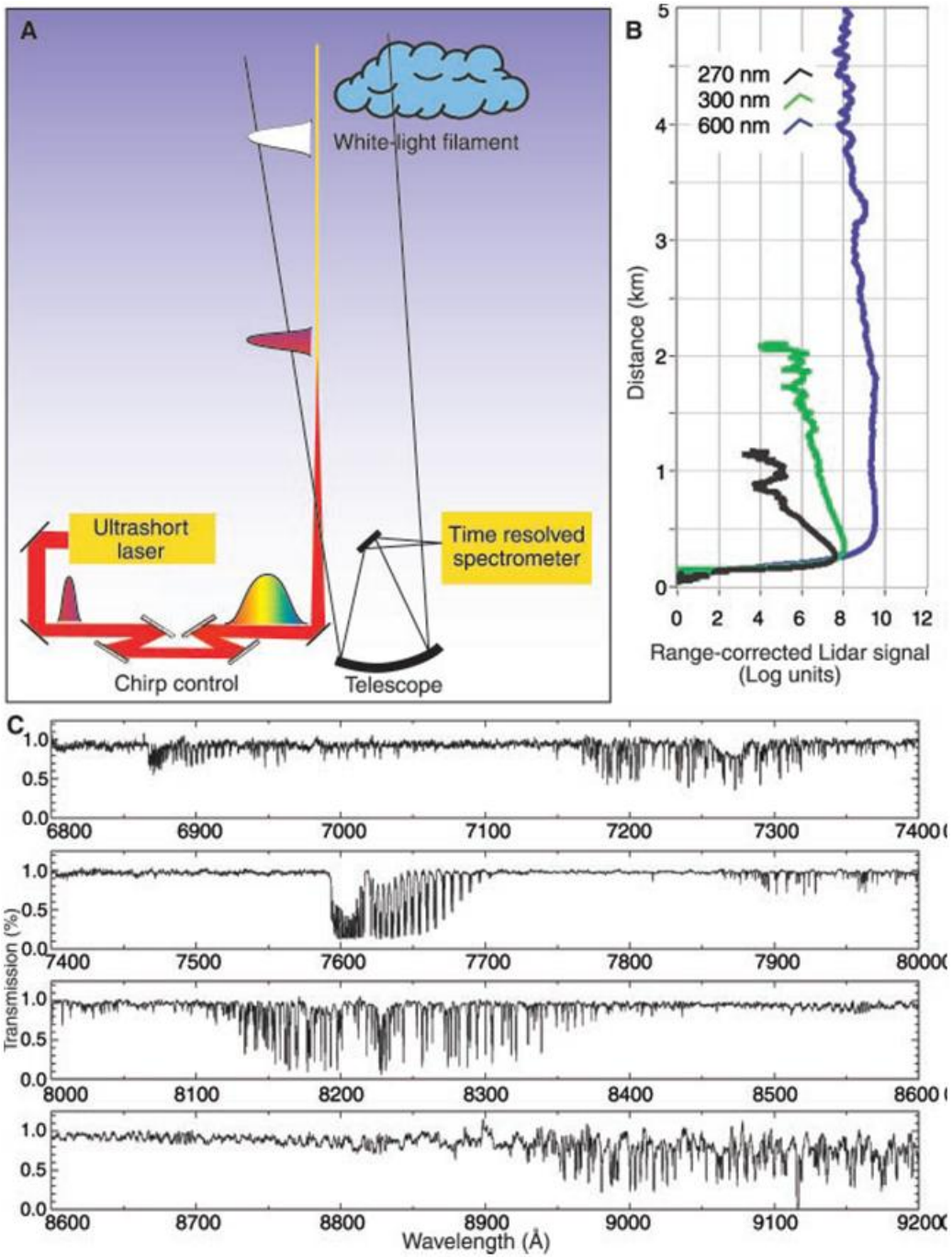

FIGURE 3-23 "(A) Schematic of the Teramobile lidar experimental setup. Before launch into the atmosphere, the pulse is given a chirp, which counteracts group velocity dispersion during its propagation in air. Hence, the pulse recombines temporally at a predetermined altitude, where a white light continuum is produced and then is backscattered and detected by the receiver. (B) Vertical SC aerosol/Rayleigh backscatter profile at three wavelengths: 270 nm (third harmonic), 300 nm, and 600 nm. (C) High-resolution atmospheric absorption spectrum from an altitude of 4.5 km, measured in a DIAL configuration." SOURCE: From J. Kasparian, M. Rodriguez, G. Méjean, J. Yu, E. Salmon, H. Wille, R. Bourayou, S. Frey, Y-B. Andre, A. Mysyrowicz, R. Sauerbrey, J.-P. Wolf, and L. Wöste, 2003, "White-light filaments for atmospheric analysis," *Science* 301: 61. Reprinted with permission from AAAS.

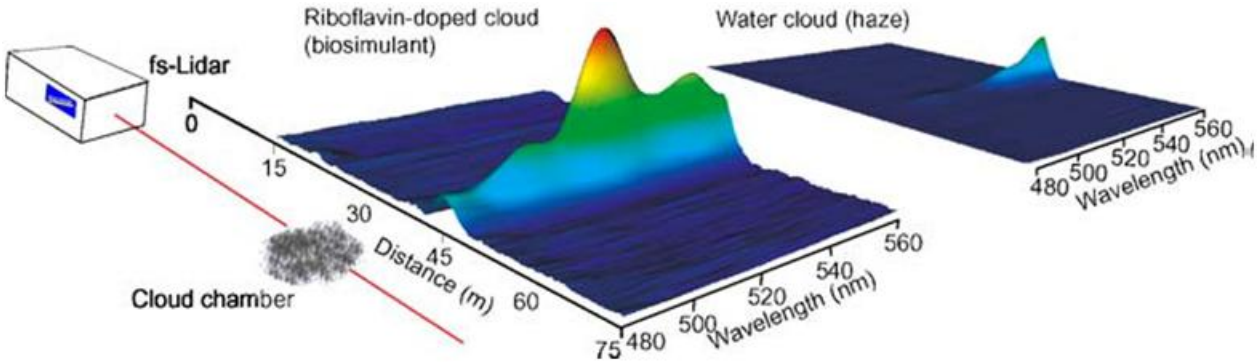

FIGURE 3-24 "Remote detection and identification of bioaerosols with Teramobile system. The femtosecond laser illuminates a plume of riboflavin (RBF)-containing microparticles 45 m away (*left*). The backward-emitted two-photon-excited fluorescence, recorded as a function of distance and wavelength, exhibits the specific RBF fluorescence signature for the bioaerosols (*middle*) but not for pure water droplets" (simulating haze, *right*).[122] SOURCE: G. Mejean, J. Kasparian, J. Yu, S. Frey, E. Salmon, and J.-P. Wolf, 2004, "Remote detection and identification of biological aerosols using a femtosecond terawatt lidar system," *Appl. Phys.* B78: 535.

through spectral analysis of the H_2O and O_2 absorption lines present in the returned spectra, cloud humidity and temperature.[123]

Finally, studies with the Teramobile system attempted to simulate the detection of bio-active aerosols, through LIF detection of clouds of 1-μm water droplets containing either riboflavin or just pure water. The system excited riboflavin fluorescence in the blue-green region by two-photon processes, by virtue of the high peak power in the pulse. The experimental setup and data appear in Figure 3-24.[124] A major advantage of the technique lies in the better atmospheric propagation of the near-IR light, compared to the short-wavelength light needed to directly excite the fluorescence. Presumably, tryptophan or nicotinamide adenine dinucleotide (NADH) fluorescence in bioactive aerosols could also be excited via higher order excitation processes.

Future work on femtosecond lidar systems may be able to employ newer, high-energy ultrafast sources using directly diode-pumped Yb-doped crystals, which would permit construction of less expensive, smaller, and more efficient sources.

In summary, the development of continuum sources driven by femtosecond-pulse lasers (and in some cases by nanosecond-pulse sources) has provided significant improvement in the measurement accuracy of path-averaged DIAL sensors. Continuum sources that are the result of coherent generation processes—and are thus precision frequency combs—have provided greatly enhanced gas detection sensitivities over other active sensors through the dual-comb technique, and have allowed broad spectral scans to be taken in under 100 microseconds. Filament-based white-light generation has enabled a new class of range-resolved DIAL atmospheric measurements. Intense near-IR femtosecond sources that excite bioactive molecules through multiphoton processes have been demonstrated. Finally, femtosecond laser technology, for short-range applications at least, can employ fiber-laser-based sources and can find use in UAV-based sensors.

[122] J. Kasparian and J.-P. Wolf, 2008, "Physics and applications of atmospheric nonlinear optics and filamentation," *Optics Express* 16:1.

[123] R. Bourayou, G. Mejean, J. Kasparian, M. Rodriguez, E. Salmon, J. Yu, H. Lehmann, B. Stecklum, U. Laux, J. Eisloffel, A. Scholz, A.P. Hatzes, R. Sauerbrey, L. Wöste, and J.-P. Wolf, 2005, "White-light filaments for multiparameter analysis of cloud microphysics," *J. Opt. Soc. Am.* B22: 369.

[124] G. Mejean, J. Kasparian, J. Yu, S. Frey, E. Salmon, and J.-P. Wolf, 2004, "Remote detection and identification of biological aerosols using a femtosecond terawatt lidar system," *Appl. Phys.* B78: 535.

Conclusion 3-11: A significant performance enhancement of both path-averaged and range-resolved differential absorption lidar can be facilitated with the use of femtosecond-pulse sources.

Recommendation 3-2: The general application of femtosecond sources should be encouraged at the development level and monitored worldwide.

Recommendation 3-3: Programs to deploy short-range sensors with fiber-laser-based femtosecond sources for use on unmanned aerial vehicles should be supported.

ADVANCED QUANTUM APPROACHES

In most of this report, the physical processes under consideration can be fully understood from a semiclassical physics perspective. That is, the optical field is treated classically (Maxwell's equations, the wave equation, etc.), and the interactions between the optical field and materials (targets, intervening media, detectors, etc.) are understood in terms of a classical electromagnetic field and either a continuous medium or quantized material model. To be sure, "photons" are commonly referred to in these discussions, but this is merely a convenient, if somewhat sloppy, expedient.[125] Even in the very low light limit, "photon-counting," shot noise limit, etc., the physics is that of a classical electromagnetic field interacting with matter.

In this section, however, potential new sensing modalities that can be realized by exploiting the truly quantum nature of the optical field are considered. There are quantum states of light that produce physical behavior that cannot be understood in terms of a classical stochastic electromagnetic field (e.g., antibunched light, squeezed states, and entangled photons) and these can be exploited.

When the term "quantum light" is used, the general reference is to optical fields exhibiting behavior—typically statistical behavior—not possible under the laws of classical physics. Many of the statistical properties of light that pose noise limitations for laser remote sensing are due to the stochastic nature of the electromagnetic field. For example, consider light (electromagnetic radiation of any frequency) from a thermal source (a lightbulb, or the Sun, say) falling onto a photodetector. Whether the detector is a linear mode device such as a p-i-n photodiode or a "photon counting" device such as a Geiger-mode avalanche photodiode, the resulting signal will fluctuate, and the statistics of these fluctuations are well understood. One aspect of the statistics of such a thermal source is "photon bunching"—that is, the photodetection events occur in "clumps." These clumps are easy to understand by thinking of this optical field as a classical stochastic process. That is to say (taking the photon counting detection case as a concrete example), if a photodetection event has just occurred, chances are the field is fluctuating "high," and another event is likely to follow immediately. If there has been a significant lapse since the last event, the field is probably fluctuating low, and more waiting time is likely. The photodetections "bunch." Other sources exhibit fluctuations as well—for example, a laser well above threshold emits light exhibiting Poissonian photocount statistics. This is the same statistical behavior one would observe in the radiation from a classical current. In Poisson photostatistics the arrival times of photons are highly random, exhibiting no temporal correlation at all.

However, not all sources exhibit fluctuations that can be explained as simply photon bunching or Poisson statistics. Light emitted from "resonance fluorescence" exhibits photon antibunching.[126] That is, the photodetection statistics are more regular than from a laser beam—like pearls on a string (see Figure 3-25). In a sense, then, antibunched light is more regular (or less noisy) than classical physics allows. Photon antibunching can be observed by sampling the light with a beamsplitter; it manifests as a minimum in the photodetection correlation function between the two output ports for zero delay time.

[125] W.E. Lamb, 1995, "Anti-photon," *Appl. Phys. B* 60: 77.

[126] H.J. Kimble, M. Dagenais, and L. Mandel, 1977, "Photon antibunching in resonance fluorescence," *Phys. Rev. Lett.* 39: 691.

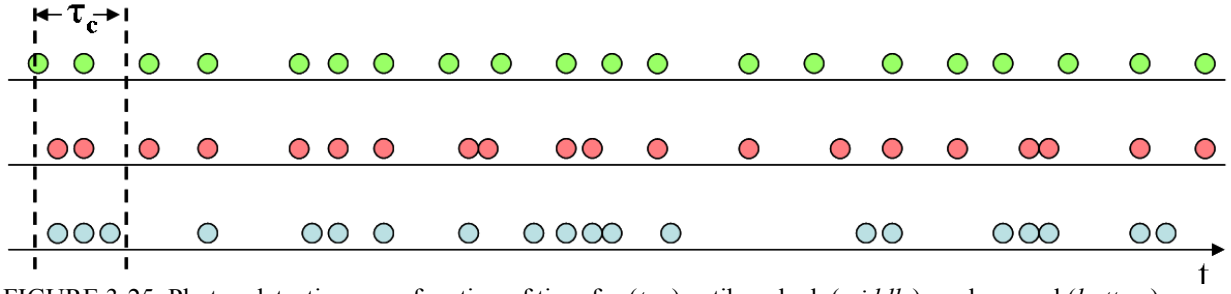

FIGURE 3-25 Photon detections as a function of time for (*top*) antibunched, (*middle*) random, and (*bottom*) bunched light. At high loss rates, the photon statistics approach that of Poissonian classical light and there is no advantage over using a conventional laser. SOURCE: By J.S. Lundeen at en.wikipedia [GPL (http://www.gnu.org/licenses/gpl.html)], from Wikimedia Commons. See http://upload.wikimedia.org/wikipedia/commons/8/86/Photon_bunching.png.

Antibunched light can also exhibit sub-Poissonian photostatistics—that is, the variance in the photon count in a given interval is less than the mean. This raises the possibility that such "ultra-regular" light might be used to improve the signal to noise ratio in a remote sensing system. Unfortunately, detailed, but straightforward, calculations show that this potential advantage does not survive the very high loss rates inherent in long-range active imaging.

Another kind of non-classical light that has drawn considerable interest is squeezed light.[127] Quantum mechanics states that there are fundamental limits to the fluctuations of the electromagnetic field. These fluctuations can be expressed in various ways, and for the purposes of this discussion, the focus is on fluctuations in amplitude and phase of the field, as this is relatively intuitive. The product of standard deviation of the number of photons and the standard deviation of the phase of the field is limited by a Heisenberg uncertainty principle of the form

$$\sigma_n \cdot \sigma_\varphi \geq \frac{1}{2}$$

For a coherent state, a good approximation to the output of a laser far above threshold, σ_n is proportional to the standard deviation of the electric field strength times the square root of the number of photons,[128] and σ_φ is proportional to the standard deviation of the electric field strength divided by the square root of the number of photons; this limit is the so-called "standard quantum limit" to measurements of amplitude and phase. The essence of squeezed light is to reduce the variance in one parameter at the expense of the other. Reducing the photon number variance is possible at the expense of increasing the phase uncertainty and vice versa (see Figure 3-26). Squeezed light was first demonstrated in the late 1980s and has been used to demonstrate phase measurement below the shot noise limit,[129] absorption spectroscopy below the vacuum state limit,[130] and a variety of other measurements beyond the standard quantum limit. Sources of squeezed light with squeezing 10 dB below the shot noise limit have been demonstrated and are finding application in interferometry applications such as gravitational wave detection.[131,132] Another application of squeezed light that attracted significant attention in the 1980s was

[127] M.C. Teich and B.E.A. Saleh, 1989, "Tutorial: squeezed states of light," *Quantum Opt.* l: 152.

[128] Recall that the number of photons is proportional to the square of the electric field.

[129] M. Xiao, L.-A. Wu, and H.J. Kimble, 1987, "Precision measurement beyond the shot-noise limit," *Phys. Rev. Lett.* 59: 278.

[130] E.S. Polzik, J. Carri, and H.J. Kimble, 1992, "Spectroscopy with squeezed light," *Phys. Rev. Lett.* 68: 3020.

[131] H. Vahlbruch, M. Mehmet, S. Chelkowski, B. Hage, A. Franzen, N. Lastzka, S. Goßler, K. Danzmann, and R. Schnabel, 2008, "Observation of squeezed light with 10-db quantum-noise reduction," *Phys. Rev. Lett.* 100: 033602.

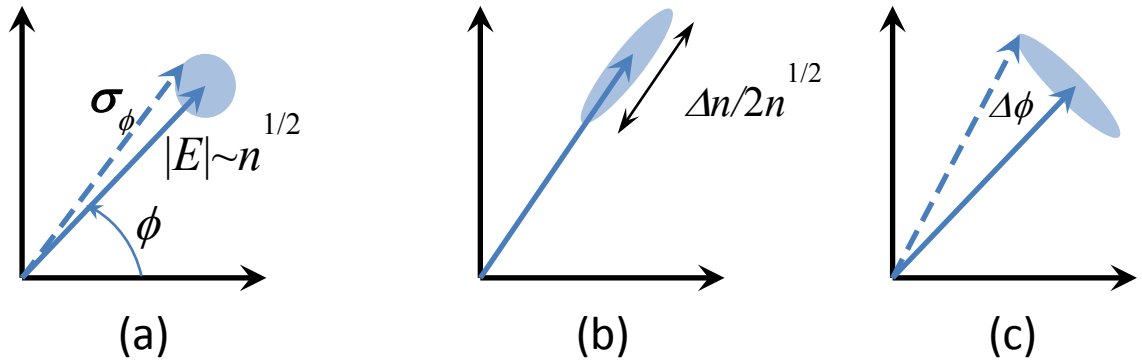

FIGURE 3-26 Squeezed states of light in polar representation. The length of the vector indicates the strength of the electric field (square root of the photon number), and the angle represents the phase of the field. (a) Typical representation of a coherent state where the shaded region represents the standard deviation of the photon number and phase. (b) "Phase-squeezed" field where the uncertainty in the phase is reduced at the expanse of larger uncertainty in the photon number. (c) "Number-squeezed" state where the uncertainty in the photon number is reduced at the expanse of larger uncertainty in the optical phase.

optical waveguide taps with infinitesimal insertion loss (enabling undetectable fiberoptic taps).[133,134] A natural question is whether these laboratory demonstrations are extensible to real-world remote-sensing applications. For example, might spectroscopy with squeezed light improve the sensitivity of remote DIAL systems, or might a phase squeezed local oscillator be used to improve the sensitivity of coherent lidar?

Unfortunately, when examined in detail, these schemes tend not to realize the initial promise for remote sensing. For coherent lidar, Rubin and Kaushik examined the problem in detail and concluded that the signal-to-noise ratio for heterodyne laser radar with a coherent target-return beam and a squeezed LO beam is lower than that obtained using a coherent LO, regardless of the method employed to combine the beams at the detector.[135,136]

One of the most intriguing aspects of quantum physics is quantum entanglement. Entanglement is more than just classical correlation; rather, it is a degree of correlation and predictability that exceeds that possible in classical physics. A detailed discussion of entanglement is beyond the scope of this report, and the reader is referred to the extensive literature.[137] In the past two decades, the resources of entanglement have been exploited in various ways in the burgeoning field of quantum information—quantum communication, quantum cryptography, quantum computation and quantum teleportation. Although these fascinating, and potentially groundbreaking, developments are beyond the scope of this study, quantum entanglement has also been proposed as a resource to enable new capabilities in remote sensing. These proposals have not engendered practical gains, but the field bears watching for the development of disruptive capabilities.

---

[132] H. Vahlbruch, A. Khalaidovski, N. Lastzka, C. Gräf, K. Danzmann, and R. Schnabel, 2010, "The GEO600 squeezed light source," *Class. Quantum Grav.* 27: 084027.

[133] R. Bruckmeier, H. Hansen, S. Schiller, and J. Mlynek, 1997, "Realization of a paradigm for quantum measurements: The squeezed light beam splitter," *Phys. Rev. Lett.* 79, 43.

[134] J.H. Shapiro, 1980, "Optical waveguide tap with infinitesimal insertion loss," *Opt. Lett.* 5: 351.

[135] M.A. Rubin and S. Kaushik, "Squeezing the local oscillator does not improve signal-to-noise ratio in heterodyne laser radar," 2007, *Optics Lett.* 32(11): 1369.

[136] M.A. Rubin and S. Kaushik, 2009, "Signal-to-noise ratio in squeezed-light laser radar," *App. Opt.* 48(23): 4597.

[137] A. Peres, 1993, *Quantum Theory: Concepts and Methods*, Kluwer Academic Publishers.

One application area for entangled photon states that has been touted[138] as offering a new capability is quantum superresolution. The idea is that in many respects, a maximally entangled state of N photons can behave as a single photon of wavelength λ/N. Experiments have demonstrated this behavior in several scenarios—namely, interferometry with fringe spacing $\lambda/2N$ and near-field optics. Unfortunately, despite the promise of λ/N quantum superresolution, nobody has yet demonstrated far field ($R \gg D^2/\lambda$) imaging performance exceeding conventional diffraction limits with entangled photon states. (Notionally, this would mean far field spatial resolution of $\lambda R/ND$, where D represents either the transmitter diameter in a flying spot lidar or the receive aperture in a focal plane lidar.)

Another application of entangled photon states is measurement below the shot noise limit, reaching the Heisenberg limit. The standard quantum limit for phase measurement of an optical field scales as the inverse of the square root of the number of photons in the field, N:

$$\delta\varphi \sim 1/\sqrt{N}.$$

Using entangled states of N photons it is theoretically possible to reach the Heisenberg limit to phase measurement which scales as the inverse of the number of photons:

$$\delta\varphi \sim 1/N.$$

Thus, for a field of 100 photons, one could improve on the performance of a measurement of optical phase by a factor of 10. Again, however, as with sub-Poissonian light, optical losses in a remote sensing system limit the effectiveness of this method. It has been shown[139] that once losses exceed the modest level of about 6.7 dB, the phase measurement actually degrades with increasing N.

One concept employing entangled photons that does not appear to be subject to the same deleterious effects of loss is quantum illumination.[140] In this technique, a series of single photon signal pulses is directed at a target. According to Llyod "each signal sent out is entangled with an ancilla, which is retained…Detection takes place via an entangling measurement on the returning signal together with the ancilla."[141] Quantum illumination with n bits of entanglement increases the effective signal-to-noise ratio of detection and imaging by a factor of $2n$, an exponential improvement over unentangled illumination. The entangled detection serves as a sort of filter, improving the SNR performance by identifying the received photons as the same ones that were transmitted. What is remarkable is that this performance enhancement is retained, even in a long-range remote-sensing application where noise and loss completely destroy the entanglement between signal and ancilla. One application area where such a capability could have a significant impact is missile defense ladar, where one is trying to determine the presence or absence of a target at very long range. Imaging applications with resolution enhancements have also been discussed in the literature, based on similar arguments as the quantum super-resolution discussed above, but it must be made clear that this again applies to near-field geometries. The key barrier to the realization of quantum illumination is the entangling measurement for multibit entanglement. Although possible by the laws of physics, implementations of this type of photodetection are not readily available.

---

[138] "Quantum Lidar - Remote Sensing at the Ultimate Limit," 2009 AFRL-RI-RS-TR-2009-180 Final Technical Report, July.

[139] M.A. Rubin and S. Kaushik, 2007, "Loss-induced limits to phase measurement precision with maximally entangled states," *Phys Rev. A* 75: 053805.

[140] S. Lloyd, 2008, "Enhanced sensitivity of photodetection via quantum illumination," *Science* 321(5895): 1463.

[141] S. Lloyd, 2008, "Enhanced Sensitivity of Photodetection via Quantum Illumination" *Science* 321 (5895): 1463-1465.

Another intriguing area that has attracted significant attention in the past few years is "ghost imaging."[142,143] In ghost imaging, one exploits the correlations between two light beams in an active imaging scenario. One beam illuminates and is scattered off a target, with the scattered light being collected by a nonresolving "bucket" detector. The second, correlated beam illuminates a spatially resolving detector. The image is formed by cross-correlations of the two photodetection signals. Thus, neither beam produces a target image by itself—the beam interacting with the target provides no spatial resolution, and the beam falling on the detector array has not interacted with the target.

The correlations between the two beams can be either classical or quantum in nature, and ghost imaging can use either direct or phase-sensitive coherent detection. The details in ultimate system performance (spatial resolution, field of view, image contrast, and SNR) do depend on the nature of the correlations and the detection technique. The potential benefits of ghost imaging in practical applications are principally in the expanded design trade space afforded by the technique, but it remains to be seen what role such imaging will have to play in practical systems.

Nonquantum Advanced Techniques

Although not strictly advanced quantum techniques, there are other nonconventional advanced concepts that should be addressed here as well.

Metamaterials (engineered materials whose properties are determined by their physical rather than molecular structure) have been used to achieve electromagnetic properties that, while allowed by physics, are generally not present in nature. These include control of the dielectric permittivity and magnetic susceptibility of materials to create negative refractive index (NRI) materials,[144] which have been shown to have unusual optical properties, enabling "perfect" near-field imaging resolution far beyond the diffraction limit of the wavelength,[145] as well as "cloaking devices" that render the metamaterial invisible. There are significant technical obstacles to overcome for these materials before applications become realizable, such as the deleterious effects of absorption, but the field should be monitored closely for breakthroughs in the technical barriers. More fundamental, however, is to understand the validity of the actual applications. There have been claims,[146,147] for example, that the perfect lensing capability of NRI materials enables high-resolution imaging at long range beyond the diffraction-limited (λ/D) resolution of the optics. This is based on a misconception[148] about the capabilities of NRI telescopes and should not be considered a viable approach. However, there may be other very beneficial applications of NRI and "perfect lenses" in general, and the field should be closely monitored and supported.

In summary, the academic community has conceived and in some cases demonstrated numerous intriguing concepts in quantum imaging that exploit the unique physics of the quantized electromagnetic field. Several of these concepts at first examination appear to offer profound advantages for active remote-sensing systems. Despite the potential promise of exploiting the quantum nature of light, however, most of these concepts can be shown not to provide a real advantage for remote-sensing systems.

---

[142] Y. Shih, 2009, "The physics of ghost imaging," arXiv:0805.1166v5 [quant-ph], Sep. 29, 2009.

[143] B.I. Erkmen and J.H. Shapiro, 2010, "Ghost imaging: from quantum to classical to computational," *Advances in Optics and Photonics* 2: 405.

[144] S.A. Ramakrishna and T.M. Grzegorczyk, 2008, *Physics and Applications of Negative Refractive Index Materials*, CRC Press.

[145] J.B. Pendry, 2000, "Negative refraction makes a perfect lens," *Phys. Rev. Lett.* 85: 3966.

[146] J. May and A. Jennetti, 2006, "Telescope resolution using negative refractive index materials," *Proceedings of SPIE* 5166: 220, *UV/Optical/IR Space Telescopes: Innovative Technologies and Concepts*.

[147] J. May and S.D. Stearns, 2011, "Imaging system using a negative index of refraction lens," US Patent 8017894.

[148] S. Stanton, B. Corrado, and T. Grycewicz, 2006, *Comments on Negative Refractive Index Materials and Claims of Super-Imaging for Remote Sensing*, Aerospace Report No. TOR-2006(3907)-4650.

Conclusion 3-12: Advanced quantum approaches, including nonclassical photon statistics, squeezed light, and entangled photons, while intriguing and potentially promising, are not currently of added utility for practical remote-sensing systems. Nevertheless, it is important to pursue and monitor this family of approaches, since new concepts with breakthrough capabilities may emerge.

GENERAL CONCLUSIONS—EMERGING SYSTEMS

The following general conclusions regarding emerging active EO systems are derived from discussions in this chapter taken as a whole.

Conclusion 3-13: Emerging active EO systems show strong advantages (signal to noise ratio gain, phase compensation and thinner, lighter apertures) at the cost of increased system complexities (computational processing costs, narrow linewidth lasers, etc.).

Conclusion 3-14: Emerging active EO technologies can complement current conventional ladar systems.

Conclusion 3-15: High-level, active EO emerging technologies will most likely be pursued through funding at university, government, or industry laboratories, with indicators given by publications and presentations.

Conclusion 3-16: Coherent active EO systems will continue to develop for applications that require access to the optical field (not just intensity).

Conclusion 3-17: Large potential markets may propel progress for emerging active EO technologies for commercial applications.

4

Active Electro-Optical Component Technologies

As has been described in Chapters 2 and 3, current and emerging active electro-optical (EO) sensing systems are implemented in many different modalities. All require components such as lasers, detectors, optics, and processing techniques to generate photons, bounce them off targets, and transform detected photons into usable information. Other components may be required as well, depending on the implementation. The specific requirements, complexity and sophistication of the components vary with the manner of implementation of the active EO sensing system and the usable information it is trying to extract. For example, no particular laser or detector technology meets all the requirements of the various active EO sensing approaches. An important factor in the last ten years' progress has been technological advances in each of the components of the active EO system: improved lasers, detectors, software, advances in robotics, and improved manufacturing technologies. This chapter discusses the variety of components currently used and some of the key technologies being developed for future systems.

LASER SOURCES FOR IMAGING

Active EO sensors employ coherent sources in the wavelength region from the long-wavelength infrared (around 10 μm) to the atmospheric transmission limit for UV light, around 200 nm. The sources can be based either on lasers or on nonlinear optical systems driven by lasers. Lasers are typically categorized by the type and format of the medium used to generate their output, which at the highest level are gases, liquids, and solids.

Solid materials are further categorized by their electrical characteristics. Solid-state lasers employ insulating solids (crystals, ceramics, or glasses) with elements added (dopants) that provide the energy levels needed for laser action. Energy to excite the levels is provided by other sources of light, either conventional sources such as arc lamps or other lasers, a process called optical pumping.

Solid-state lasers in turn are divided into two broad categories, bulk or fiber, with the latter having recently emerged as an important technology for generation of high average powers with high beam quality, as discussed below.

Even though they are also made from solid-state materials, semiconductor lasers are considered to be a separate laser category. While the lasers can be made to operate by optical pumping, if the semiconductor material can be fabricated in the form of an appropriate p-doped, n-doped (PN) junction, it is possible to pass electrical current through the junction and generate laser output directly. These "diode lasers" are by far the most widely used form of semiconductor lasers, and have led to major advances in source technology for active EO sensors.

The use of lasers for active EO sensors started with the use of the first laser to be operated (1960), the solid-state ruby laser.[1] With the development of techniques to generate nanosecond-duration pulses, ruby lasers provided the first example of laser rangefinders. Since 1960, almost every type of laser has been employed in demonstrations of active EO sensors. At this point, with a few exceptions, sensors now

[1] T.H. Maiman, 1960, "Stimulated optical radiation in ruby" *Nature* 187(4736): 493.

employ either diode or solid-state lasers, with the latter often combined with nonlinear optics. This looks to be the case for the foreseeable future, owing to a favorable combination of output format, operating wavelength, relatively high efficiency, ruggedness, compact size, and reliability.

With regard to nonlinear optics, most systems employ crystals with special properties that can convert a laser output to shorter wavelengths by harmonic conversion or to longer wavelengths, by parametric processes. The latter have the added advantage that the wavelengths generated can be tuned by a variety of techniques, an advantage for sensors that require specific wavelengths—for example, in the detection of specific gases. In some cases parametric and harmonic processes can be combined for added wavelength tuning.

The next sections go into more detail regarding the dominant active EO sensor source technologies. A reasonable amount of space is devoted to diode lasers, because they now feature prominently as both stand-alone devices for low-cost, short-range sensors and as optical pumps for solid-state lasers used for long-range sensors.

Diode Lasers

There are two major types of diode lasers, interband and cascade. The former is by far the most widespread in use, but the recently developed cascade lasers have emerged as an important source of mid- and long-wave infrared emission.

Diode Lasers: Interband, Edge-Emitting

Interband diode lasers employ electronic transitions between energy states in the conduction band and the valence band of the semiconductor crystal. First and foremost, the semiconductor material used for the diode laser must have a direct optical bandgap—that is, the electronic transitions giving rise to laser operation must occur without the assistance of mechanical vibrations (phonons) in the semiconductor. This eliminates the most common semiconductor material, silicon, as well as the related material germanium.

Laser operation in direct-bandgap semiconductors occurs by optical transitions between the lowest-lying electronic states in the conduction band to the highest-lying states in the valence band, with the requirement that there be a higher density of states occupied in the conduction band than in the valence band. This nonequilibrium condition, or "population inversion," occurs in diode lasers through the injection of sufficient electrical current into the lasing region.

Fabrication of a diode laser (Figure 4-1 for an edge-emitting device) requires that the semiconductor material be available in both n-type (electrons the majority current carrier) and p-type (holes the majority current carrier), with this achieved by the addition of certain impurities to the materials. While there are many semiconductors with direct bandgaps, many of these, for a variety of reasons, cannot be doped to make both n- and p-type materials to form diodes, notably all of the so-called II-VI binary semiconductors such as CdS, CdTe, ZnS, ZnO, and ZnSe. (The Roman numerals refer to the relative positions of the elements in the periodic table.) Infrared lasers have been operated based on IV-VI materials such as PbS and PbSe but require cryogenic cooling for efficient operation. To date, III-V binary, direct-bandgap semiconductors such as GaAs, GaSb, InAs, InSb and, most recently, GaN, as well as alloys of these crystals with other III-V elements, are by far the most widely used materials for use as diode lasers. Noncryogenic operation in the 350-2,000-nm region is possible with devices based on III-V materials.

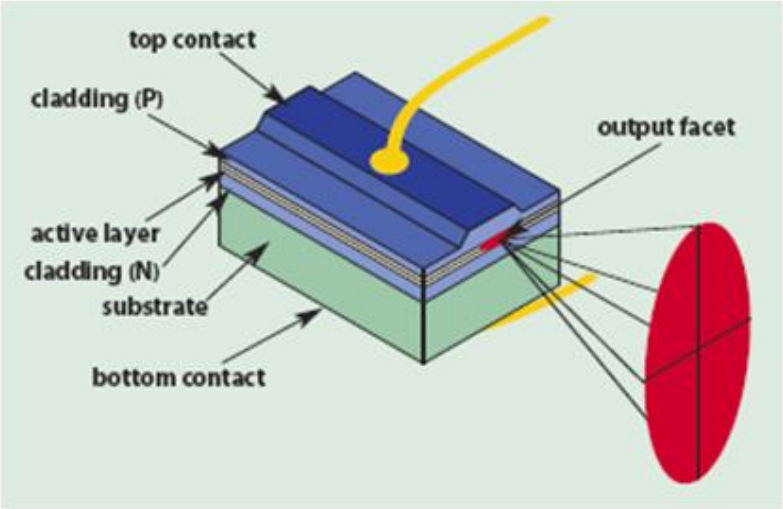

FIGURE 4-1 Reprinted with permission from The Photonics Handbook, online at http://www.PhotonicsHandbook.com. Copyright 2013. Laurin Publishing, Pittsfield, Mass.

Figure 4-1, though a very simplified diagram, shows the key features, of an edge-emitting interband diode laser. Electrical current passes through a wire bond to a stripe top contact, which confines the current in two dimensions. In the device shown, the current passes (is "injected") through p-type material into the "active layer," fundamentally a junction between p- and n-type material, then through n-type material to an electrical contact to complete the circuit. Laser gain occurs in a thin, 3-D region comprising the active layer, with the third dimension set primarily by the width of the stripe contact. The injected current and the structure of the device act to form an optical waveguide that confines the laser light to the same region as the gain, both vertically (perpendicular to the plane of the junction) and horizontally (parallel to the junction plane.) Most edge-emitting diode lasers operate with the laser cavity formed by the cleaved ends of the waveguide region. One end is coated with a deposited stack of dielectrics to form a highly reflecting mirror, while the other end is coated to provide only a small amount of reflectivity. Typical lengths between the two faces are in the sub- to several-millimeter range. In some cases one diode end has an antireflection coating to enable use of an external mirror and/or tuning element to obtain more control of the diode wavelength.

In reality, the details of the junction are much more complex than the simple PN junction implied by Figure 4-1. Major developments in diode-laser performance, from the first demonstrations of operation in the early 1960s, have been the result of developing the fabrication technology to make multilayer semiconductor structures (heterostructures) that better confine the lasing region and reduce the current needed to get laser operation, as well as increase the efficiency in converting electrical power to laser power. As the thickness of the layers has been reduced with improved processes to tens of nanometers, quantum effects that fundamentally change the nature of the semiconductor-material energy levels have been utilized to further improve performance. In addition to structuring in the plane of the junction, additional material structuring in the horizontal direction to confine the current and laser power has led to better laser properties.

A critical and fundamental characteristic of a diode is the nature of the region where laser action occurs. Referring to Figure 4-1, in the vertical direction the region of laser emission is limited to a dimension of about 0.5 μm, set by the fundamental nature of the junction. This is comparable to or smaller than the wavelength of the light emitted, which assures that the light ("fast axis") emitted in the vertical direction is diffraction limited, consisting of only a single transverse mode. In Figure 4-1 the beam dimension is larger in the vertical direction because the fast-axis light rapidly diverges from the

small emitting region. To effectively capture all of the power from the diode laser requires fast optics, at least for the fast-axis light, and this is accomplished with specialized aspheric optics, often fastened directly to the diode-laser package.

The limiting feature of the small dimension is that even relatively low absolute power levels lead to very high power densities at the surface of the edge-emitting diode laser, on the order of 10 MW/cm^2, and can lead to catastrophic optical destruction (COD) of the diode laser beyond certain power levels. COD results from the inevitable defects at the surfaces of semiconductors, which absorb the laser light, heat up, and, with enough power, melt the material at the surface. Means to improve the COD level are often considered proprietary or, at a minimum, patentable, by the diode manufacturers. For some semiconductor materials, particularly compounds designed to operate at wavelengths of 1,500 nm and longer, the diode power levels are limited by simple heating effects, well below the levels set by COD.

No matter the cause of the limited power, one can increase the power output of an edge-emitting diode laser by increasing the width of the lasing region along the horizontal direction, which reduces the intensity at the surface for a given absolute power level and also the generated heat density. Unfortunately, beyond a dimension of several microns, the light emitted in the direction along the plane of the junction becomes multimode. State-of-the-art diode lasers are able to generate diffraction-limited continuous wave (CW) powers on the order of 0.1 W/μm of horizontal width, so that diode lasers with strictly single-mode outputs are capable of only 0.1-0.2 W of output power. The low power is adequate for many applications, and for optical data storage devices (CDs, DVDs and Blue-Ray discs) the yearly volume of low-power diode lasers approaches 1 billion.

An upper limit to the power from a single emitting region is found when the laser action, rather than occurring through the cavity formed by the cleave ends, starts in a perpendicular direction along the width of the stripe. Typical junction widths for so-called broad-stripe lasers are in the 100-400μm range. State-of-the-art diode lasers operating in the 900-nm wavelength region can generate 10-15 W of CW power with a 100 μm stripe width.

In order to produce diode lasers with higher average power, the semiconductor laser industry, starting in the 1980s, took advantage of improvements in material quality and lithographic techniques to manufacture multiple diode lasers on one piece ("chip") of semiconductor material.

Figure 4-2 is a schematic of a multi-emitter diode "bar," and the most common overall horizontal dimension is 1 cm. The bar fabrication requires a means to suppress laser action along the length of the bar, and an additional manufacturing challenge is mounting the bar on a heat sink to remove the heat efficiently while also keeping the entire bar in one plane, so all of the emitters line up exactly. If they do not, the net divergence of all of the beams after collection by external optics increases. Key bar parameters are the number of emitters per bar and the stripe width of each emitter, leading to the "fill factor" for the bar, with typical stripe widths in the 100-200μm range. Unlike single-emitters, the power output of bars running in the CW mode is thermally limited, although recent improved thermal engineering of bar mounts and cooling may lead to COD starting to become a limit as well. The thermal limits to power can be overcome if the diodes are run in pulsed mode, sometimes referred to as quasi-CW (QCW). Typically, this mode is used for pumping solid-state lasers, with pulsewidths and on-time (duty cycle) percentages in the 0.1-1-msec and 1-10 percent ranges.

Present state-of-the art bars, designed for CW operation, have fill factors of 30-50 percent, and for the most efficient semiconductor materials, operating in the 900-nm region ("9xx devices"), commercially available power levels are as high as 160 W. Higher powers are possible with reduced device lifetimes, which drop rapidly as the temperature of the semiconductor junction increases. QCW bars run with fill factors of 75-90 percent and peak power levels of 250-300 W for 9xx devices.

Appendix C provides a number of tables that review the state of the art today in lasers. Table C-1 summarizes the key features of edge-emitting diode lasers, in both single-emitter and bar format. Given the significance of such devices for a number of commercial and military applications, there is continuing development of devices over the entire spectral range, and the performance (power, efficiency) and, to a lesser extent, the wavelength coverage can be expected to improve in the next 10-15 years.

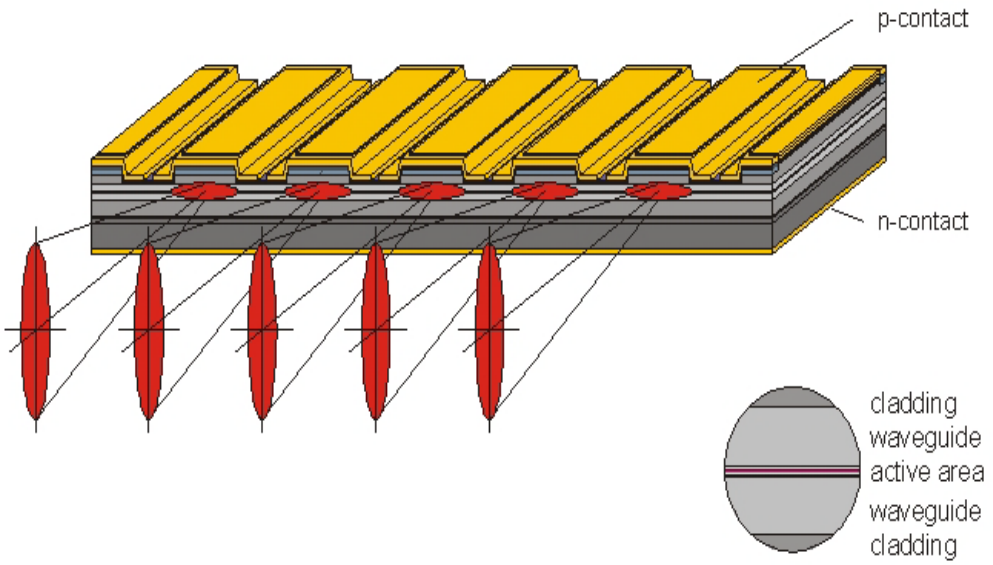

FIGURE 4-2 Diagram of "bar" lasers. SOURCE: © Jeniptik Laser GmbH.

The specific configurations of edge-emitting diode lasers as optical pumps for solid-state lasers are discussed in a later section.

Diode Lasers: Interband, Vertical Cavity

Another class of interband diode lasers, developed more recently than the edge-emitting devices just discussed, is based on improved semiconductor-processing technology. Several designs of these new devices, vertical-cavity surface-emitting diode lasers (VCSELs), appear in Figure 4-3. The key feature is that the device is constructed by building up a series of semiconductor layers and that the direction of optical gain in the junction is perpendicular to the plane of the junction. The latter, since the path length is very short, requires much higher reflectivity mirrors in the laser cavity to permit oscillation, with the output coupling mirror having less than 1 percent transmission. The bottom and top mirrors (distributed Bragg reflectors (DBRs in Figure 4-3) consist of alternating layers of high- and low-index semiconductor materials, and the active region, employing quantum-well-structure semiconductor heterojunctions, is sandwiched between the DBRs. Electrical contacts on the top and bottom allow electrical pumping with current passing through the structure, and a circular hole is etched into one of the contacts to allow the output beam to emerge. The entire length of the VCSEL is on the order of 10 μm for devices operating in the 900-nm region. To generate a diffraction-limited output in that region, the emitting-region diameter has to be kept below 4 μm, with power levels in the 5-mW range, but higher multimode powers are possible as a result of recent work, with Princeton Optronics claiming operation of a 5 W, 976 nm, CW laser with a 300 μm aperture.[2]

Once VCSELs became commercially available, they found widespread application as 850-nm-region transmitters for short-haul, high-bandwidth fiberoptic data links. The circular output beam of VCSELs and comparatively small natural divergence make their coupling to fibers an easy task. Because of the extremely short optical cavity, the devices are operate on a single frequency and can be turned on

[2] See http://www.princetonoptronics.com/technology/technology.php#VCSEL.

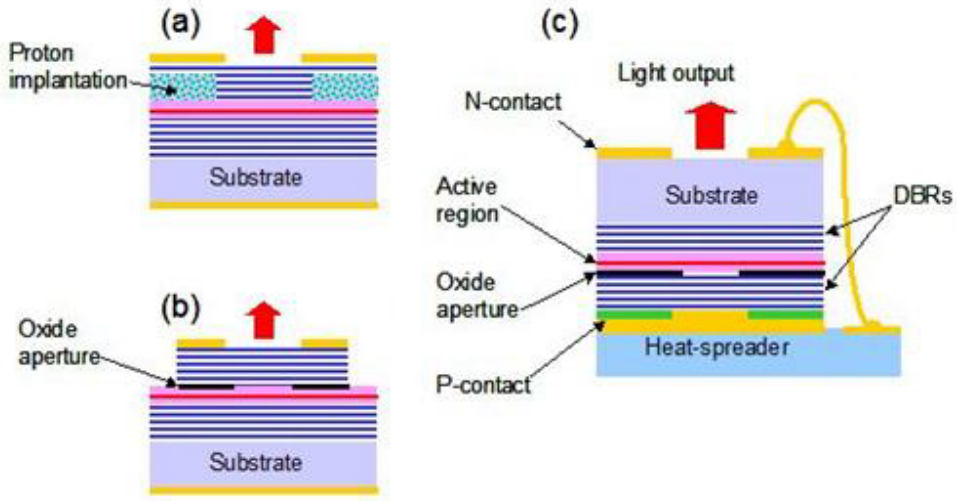

FIGURE 4-3 Several different VCSEL structures. SOURCE: Courtesy of Princeton Optronics, Inc.

and off in the 10-100 ps range, allowing direct current modulation for data rates in the tens of gigahertz. Since the fabrication process is similar to semiconductor integrated circuits (ICs), it is possible to construct many devices at one time, providing a major cost advantage over edge-emitting diode lasers. Finisar[3] claims to have shipped over 150 million devices, which also find application as the source in an optical mouse. Future large-scale applications are likely in optical interconnects for high-speed electronics systems and in fiber-to-the-home.

More recently several groups have developed arrays of VCSEL devices to produce much higher total powers, although the beam quality is not high since the devices are incoherently combined. The main challenge for high-power arrays is heat removal, where the desire to have a high areal density of devices to improve brightness conflicts with the need to remove power. Princeton Optronics has demonstrated 230 W of 976-nm CW power from a 5 × 5-mm array and 925 W of QCW power from the same area.[4]

VCSELs had been limited in wavelength coverage to 650-1,300 nm primarily because the GaAs-based technology used to manufacture the electrically pumped devices cannot extend beyond that region. Recently, however, devices operating in the 1,500-nm region have been reported, based on the InP semiconductor, and are expected to find wide use for higher-speed fiber applications where the low dispersion of silica fibers at 1,500 nm allows 10 GHz and higher long-distance links. Development of the longer-wavelength VCSELs is intense at this writing, given their large potential in the telecom industry.

Quantum Cascade Lasers

Unlike the interband diode lasers just discussed, quantum cascade lasers (QCLs) rely on a totally artificial gain medium made possible by quantum-well structures, a process sometimes referred to as band-structure engineering. The gain from a QCL relies on transitions entirely from electronic states within the conduction band of the material, as modified by the quantum well structure. Typical devices are fabricated on InP and employ GaInAs/AlInAs quantum wells, which are formed from nanometer-thick

[3] See http://www.finisar.com/sites/default/files/pdf/Finisar_infographic_timeline_web.pdf.

[4] J.-F. Seurin, C.L. Ghosh, V. Khalfin, A. Miglo, G. Xu, J.D. Wynn, P. Pradhan, and L.A. D'Asaro, 2008, "High-power vertical-cavity surface-emitting arrays," in High-Power Diode Laser Technology and Applications VI, M.S. Zediker, ed., *Proc. SPIE* 6876: 68760D.

layers, typically 10-15 in number. Devices based on other material combinations have been demonstrated but show inferior performance. Since the energy separation of the states is a function of the structure, it can be adjusted to generate a wide range of wavelengths, covering approximately 3.5-30 and 60-200 μm, with the latter falling in the terahertz region. The gap in long-wavelength coverage is due to the strong phonon absorption region of the semiconductor, which overcomes the gain in the quantum wells, while the short-wavelength limit is set by several factors, including fabrication difficulties with the required quantum-well alloys and losses due to scattering among different energy states in the conduction band. Different materials systems, now under development, allow operation at shorter wavelengths.

Laser operation comes through injection of a current of electrons through the quantum-well structure, resulting from a voltage applied across the structure. Both the efficiency and gain would be low if the device employed just one quantum well, since the energy of the emitted photon is small compared to the energy required to put an electron in the conduction band. This drawback is overcome by fabricating a structure consisting of multiple (typically 25-75) quantum wells in series, with another heterojunction structure, the "injector" through which the electron tunnels from one quantum well to another. Figure 4-4 is a simplified, partial schematic of a QCL structure with distance as the horizontal axis and energy as the vertical, showing the energy levels for the electron with voltage applied across the structure length that creates the gradient in energy. The QCL name comes from the "cascade" process of one electron interacting with multiple quantum wells as it crosses the structure. The one electron creates multiple photons, corresponding to the number of quantum wells in the device, thereby greatly improving the efficiency. As with edge-emitting interband lasers, the laser power is emitted perpendicular to the current flow through the layers, so that a given quantum well interacts with only a thin slice of the laser mode.

One of the remarkable features of QCLs is that they are able to operate in the mid- and long-wave IR regions without the need for cryogenic cooling, even in the CW mode. This is not possible with interband diode lasers at the same wavelengths, since the high thermal population of the conduction band turns such narrow-bandgap diodes into, essentially, short circuits at ambient temperatures.

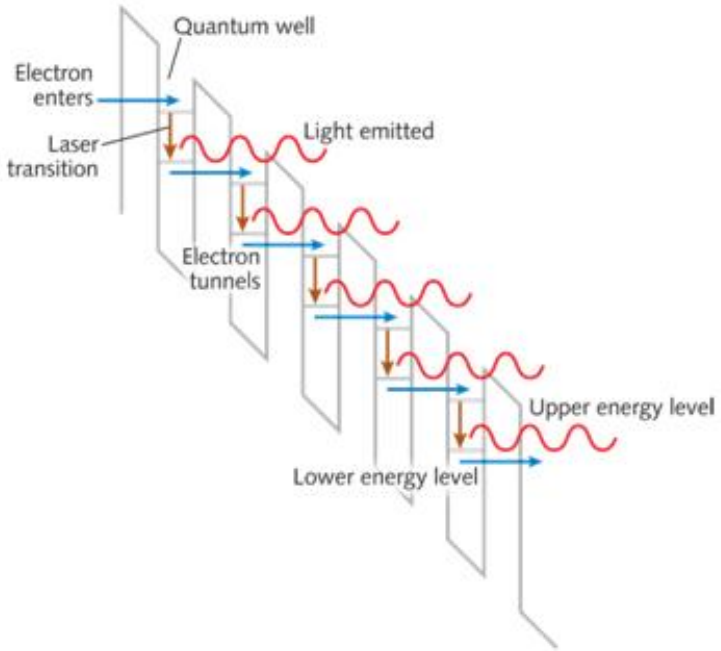

FIGURE 4-4 Simplified diagram of a quantum cascade laser. SOURCE: Courtesy of Laser Focus World.

Interband Cascade Lasers

A related semiconductor laser that has features of both interband and cascade diode lasers is the aptly named interband cascade laser (ICL). The operation of the device relies on multiple quantum-well structures, but now both electrons and holes are involved, and the quantum-well transitions are between an upper state in the conduction band and a lower state in the valence band. The material system commonly used involves either GaSb or InAs substrates, with alloys of InAs, GaSb, and AlSb employed for the quantum-well and injector structures. There are two advantages of the ICL over the QCL: (1) shorter-wavelength operation (at room temperature from 2.9 to 5.7 μm, with cryogenic operation from 2.7 to 10.4 μm), and (2) the ability to lase with much lower electrical powers (by a factor of 30 or so) than QCLs, since the transition lifetimes are much longer. As with conventional interband edge-emitting diode lasers, one would expect difficulty with long-wavelength operation at room temperature, and QCLs have an advantage there.

The current state of the art is summarized in Appendix C, Table C-2, for both QCLs and ICLs, but, as with conventional interband lasers, these numbers will change in the next 10-15 years.

Solid-State Lasers

The laser transitions for solid-state lasers occur (with a few exceptions) between energy levels of ionized atoms from the 3-D transition-metal group (Sc to Zn), from the ionized rare earth group, or from the lanthanide series of elements (La to Yb). The atoms are "hosted" in solids (crystals, glasses or ceramics) and are added as dopants to the mix of elements used to make the solids. The ions giving rise to laser operation are referred to as the active ions.

The first laser to operate was based on an artificial ruby crystal, which, in fact, consists of 3-D transition-metal Cr^{3+} ions hosted in Al_2O_3, a material known as corundum or sapphire. ("Sapphire" to the technical community is a transparent crystal, not to be confused with the naturally occurring, colored gemstone of the same name.) For transition metals, laser operation can be found from doubly to quadruply ionized atoms. The 3-D electronic states of transition metals have a reasonably large spatial extent, and thus interact strongly with the host crystal. Accordingly, the laser properties of transition-metal-doped solid-state lasers depend strongly on the host crystal.

In contrast, for the rare earths the laser transitions are primarily among different energy levels of 4f electronic states, and these, because they have a much smaller spatial extent than the 3-D states, are much more insensitive to the host crystal. With a few exceptions, a given rare earth in the same ionization state (almost exclusively triply ionized) will provide laser operation at a wavelength that is nearly independent of the host crystal.

The interaction between the host crystal and the active ion also has a profound effect on the span of frequencies/wavelengths where there is optical gain for the laser, referred to as the "linewidth." In the case of the rare earths, and for some levels of the transition metals, when the ion changes electronic state, the surrounding atoms (the lattice) stay in the same position. Figure 4-5, left, illustrates the case when there is no lattice change, with a drawing referred to as a configuration-coordinate diagram. The horizontal axis is some measure of the lattice position (say, the separation of all the surrounding atoms from the active ion) while the vertical axis represents the combined energy of the lattice and the active ion. In the so-called harmonic approximation, the system energy increases quadratically as the coordinate moves on either side from the lowest-energy, or equilibrium position. In the classical physics view, this movement will occur as the lattice vibrates from being at some finite temperature. In a more quantum-mechanical treatment, even at absolute zero, the uncertainty principal means that there is a distribution of displacements about equilibrium. The quantum-mechanical treatment also shows that the vibrational energy is quantized, with the vibrational quantum states called "phonons." From Figure 4-5, left, assuming that there is no displacement in the equilibrium position between the ground state and the excited state, and that the quadratic change energy with displacement is the same, then the energy

difference for light being absorbed (upward transition) and that being emitted (downward transition) is the same and independent of the displacement. One key and generally valid assumption is that the electronic transition occurs much faster than the lattice vibrates (Franck-Condon approximation) and so the lattice remains "frozen" during the transition. In the diagram, this is why upward and downward transitions are shown as vertical arrows.

In this simple picture, the spectral width of both absorption and emission would be zero, and would be at the exact same energy. In reality, other effects, especially that of phonons in disrupting the phase of the light making the transitions, lead to a finite linewidth that is a tiny fraction of the transition energy, on the order of 10^{-4} to 10^{-5} of the transition frequency/wavelength. This is several orders-of-magnitude smaller than for semiconductor diode lasers.

Figure 4-5 right, shows a configuration coordinate diagram for the case when the electronic transition causes the lattice to shift its equilibrium position. One important dynamic is that when the active ion is put into the excited state, the lattice reaches its new equilibrium position on a timescale of picoseconds. As noted, when an electronic transition takes place, it is much faster than this, but for the ions of interest, the probability of this transition actually occurring is very low on a picosecond timescale, so the lattice gets to its new position ("relaxes") before any light is emitted by a transition back to the ground state.

Two important things result from this. First, the energies, and hence the frequencies of light absorbed in electronic transitions from the ground state to the excited state, are generally higher than for light emitted by transitions from the excited to the ground states. Second, it is evident from Figure 4-5 and the lengths of the arrows, that there can be a range of energies for either absorption or emission, due to the displacement of the two parabolas. The linewidths for absorption and emission can be very broad, on the order of 10-20 percent of the central energy.

As noted in the introductory portion of this section, unlike semiconductor diode lasers, which convert electrical power directly to laser power, solid-state lasers need a source of light to operate, which typically is driven by electrical power. The optical pumping process requires that the solid medium absorb light, which then must result in a population inversion between the upper and lower laser levels.

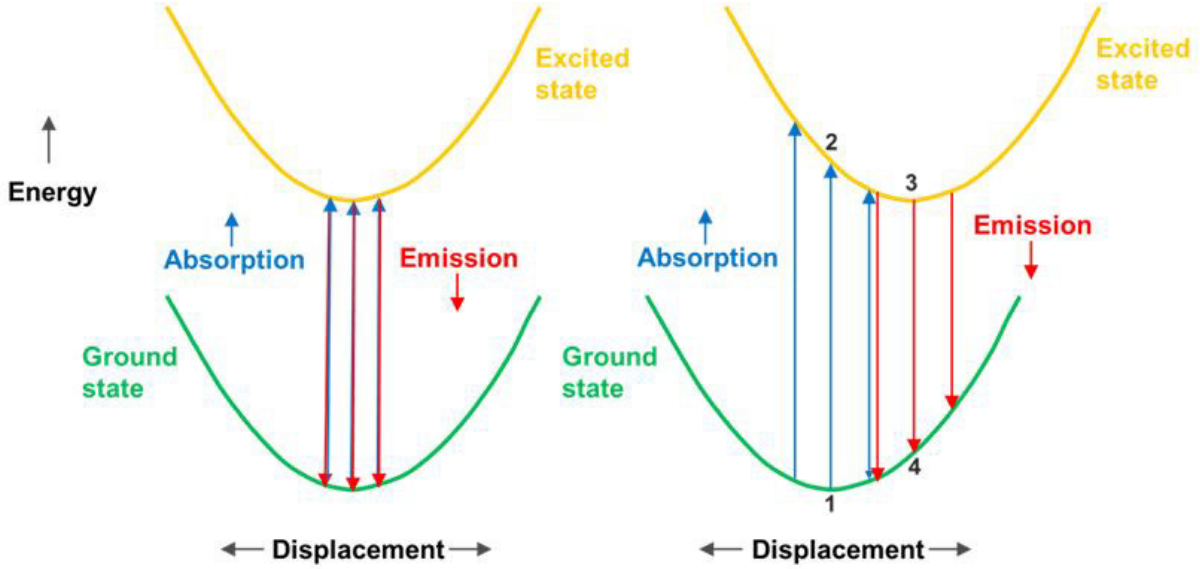

FIGURE 4-5 *Left:* Configuration coordinate diagram, no change in crystal. *Right:* Same but with a shift in crystal atomic positions.

A simple electronic system with just two energy levels and no lattice displacement (Figure 4-5 left) cannot be used to make a laser, since the material becomes transparent when a population inversion occurs. At least one higher level is needed to absorb the light, but that level must be able to transfer its excitation to the upper laser level. If the laser transition then occurs to the lowest energy level of the active ion, laser action is possible only when a substantial fraction of the ions are pumped into the upper laser level. This type of laser operation is referred to as "3-level." If the laser operation is not to the lowest energy level but to a higher level, and that level is high enough in energy that in thermal equilibrium the population of the level is small enough to not affect laser operation, the laser operation is "4-level." For systems where the lower level is above the ground state but has a non-negligible population, the designation is sometimes called "3½-level."

In Figure 4-5 right, it is evident that, even with only two electronic levels, it is possible to have 4-level operation. As marked in the diagram, pump light creates a transition from the ground state in equilibrium position (1) to the first excited state in nonequilibrium (2), which then relaxes to the equilibrium position (3) and makes a laser transition to the nonequilibrium ground state (4). This type of laser operation is often called phonon-assisted or vibronic.

The ruby laser is an example of a 3-level laser, with a narrow-line (nonvibronic) laser transition (the so-called "R-line") terminating on the ground state but with the pumping absorption lines caused by broadband vibronic transitions. These, with very high probability, deliver their excitation to the upper laser level by a so-called nonradiative process that makes up the difference in energy between the pump levels and the upper laser level by the generation of multiple phonons, often called "multiphonon relaxation." The measured absorption coefficient for ruby shows three broad peaks at 250, 400, and 550 nm due to vibronic transitions, all of which are effective in exciting the upper level of the R-line transition at 694.3 nm.

Shortly after the ruby laser was demonstrated, laser researchers were able to operate solid-state lasers based on rare earth elements. The challenge for rare earths was in finding the right combination of energy levels among all the possible ions from that series. The most effective dopant was quickly established as neodymium (Nd^{3+}); a partial, simplified energy-level diagram for the ion appears in Figure 4-6. The most common laser transitions for the ion occur in the 1050-1,080-nm region, from level E3 to level E2. Optical pumping is possible from the ground state (E1) to a multiplicity of levels (E4), all of which are efficient in exciting level E3 by multiphonon relaxation, with energy cascading down one level to another and finally ending up in the upper laser level. A key reason that the Nd^{3+} ion works well as a laser is that there is a large gap in energy from E3 to the next-lower level, so much that there is very little probability of multiphonon relaxation, and energy that winds up there can be extracted effectively by laser action. Another key to the success of Nd^{3+} lasers is that level E2 is sufficiently above the lowest energy level so that it has only a negligible population at normal temperatures. On the other hand, energy in that level can rapidly cascade down to the ground state, and not build up to the point that laser action stops. Thus, the Nd^{3+} ion provides true 4-level operation.

The right-hand side of Figure 4-6 indicates that the simplified energy levels shown as thick bars are in fact composed of multiple levels, a result of the splitting of the energy states of the rare earth ions by interactions with the surrounding atoms of the host crystal. The cluster of multiple levels indicated by the thick bars are often referred to as "manifolds."

Given all these conditions for effective laser operation, it may not be surprising that, among the total of 13 triply ionized rare earth ions that could support laser operation, only Nd^{3+} has all of the characteristics just described. The other common rare earth laser ions, which include Er^{3+}, Tm^{3+}, Ho^{3+}, and Yb^{3+}, will be discussed below, but they all have one or two unfavorable characteristics that must be overcome for effective laser operation. As discussed below, a combination of laser-based optical pumping and the use of fiber formats has opened up a wide variety of uses for these ions.

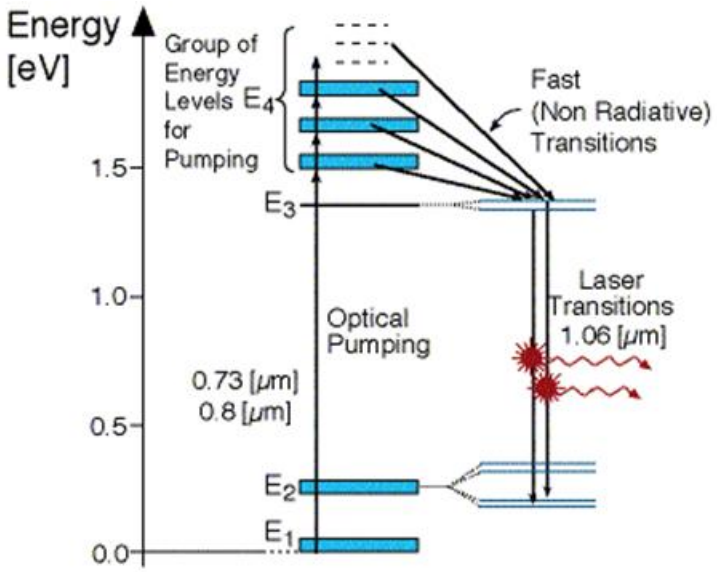

FIGURE 4-6 Simplified energy-level diagram for an Nd^{3+} ion.
SOURCE: Courtesy of Rani Arieli, Weizmann Institute of Science.

An alternative Cr^{3+}-doped crystal, $BeAl_2O_4$, known as alexandrite from the gem of the same name, emerged from work in the late 1970s and presented a case where the level giving rise to the vibronic transition was, like ruby, higher in energy than the R-line level but close enough so that it was partially excited by optical pumping. Vibronic laser action is possible at room temperature around 760 nm, but the optical gain is low and the tuning range is reduced due to excited state absorption (ESA). By running the crystal well above room temperature, it is possible to improve the laser characteristics and provide tuning from about 720-860 nm. The low gain has limited applications in areas requiring generation of high-energy pulses.

The demonstration in 1982 of the Ti^{3+}-doped sapphire (Ti:sapphire) laser provided a system with a well-known host crystal and a new laser ion. The single d electron of the Ti^{3+} ion provided a system similar to that diagrammed in Figure 4-5 right, with somewhat more complexity due to the details of the vibronic levels. ESA is not a factor since there are no higher-lying 3-D electronic levels and, in sapphire, other possible transitions are too high in energy. Absorption and emission intensities as a function of wavelength appear in Figure 4-7, along with the spectral distribution of the optical gain, which is red-shifted owing to the details of the Einstein relation of the measured emission spectra to the gain spectrum. The half-width of the gain spectrum is about 100 THz in frequency space, the widest of any known laser medium, and the system has been tuned from 660 to 1,100 nm.

A unique class of solid-state media has recently emerged, based on divalent transition metals doped in semiconductors, the most well-developed of which is the dopant Cr^{2+} in ZnSe. As with Ti:sapphire, the vibronic transitions lead to very broad linewidths, but the transitions are centered in the mid-IR region, around 2,500 nm for the case of Cr:ZnSe, which has been tuned from about 2,000-3,500 nm.

Further discussion of solid-state lasers below considers the two main formats for these systems, bulk and fiber.

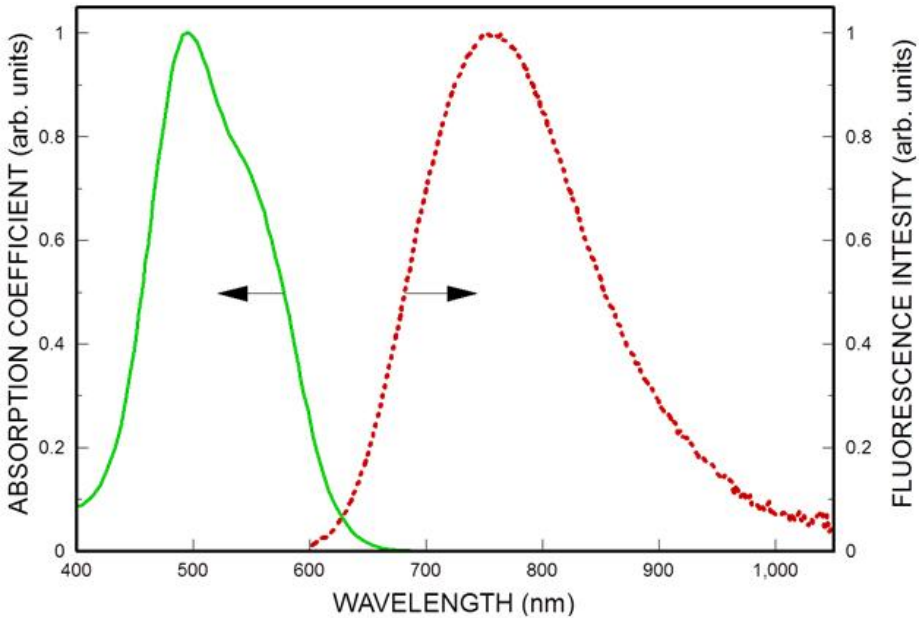

FIGURE 4-7 Absorption and emission spectra for Ti:sapphire. SOURCE: Courtesy of Peter Moulton, Q-Peak, Inc., http://www.qpeak.com/.

Bulk Format

Solid-state lasers in bulk format are based on active ions doped in crystals, glasses, or ceramics. The material can be fabricated in a variety of shapes and must allow optical pump power to be transmitted into the material, as well as allow the generated laser power to spatially overlap with the region of material that is pumped. Figure 4-8 shows a schematic of the first ruby laser, where the ruby crystal was fabricated in the form of a rod, 1 cm in diameter by 1.5 cm long and the laser cavity employed mirrors deposited directly on the polished ends of the rod. A helical xenon-gas-discharge, quartz-envelope lamp, originally developed for flash photography, surrounded the rod and was in turn surrounded by a reflecting cylinder ("pump cavity") to help concentrate the light from the lamp onto the barrel surface of the rod. Ruby has strong visible-UV-region absorption bands, which overlap well with the emission spectrum of the lamp.

The basic arrangement in Figure 4-8 is still employed today for many lamp-pumped solid-state laser systems. The helical lamp is now usually replaced with a linear-discharge lamp, and pump cavities may be formed as a cylindrical ellipse, whose inside is coated with a highly reflecting metal like silver and with the rod and lamp each centered in one of the foci. Alternative cavities may employ a diffuse reflector to help even out the distribution of pump power in the rod. Typical rod diameters are 3-12 mm and lengths are 50-150 mm, though high-energy systems may use much larger sizes.

The simple example of the lamp-pumped laser of Figure 4-8 is used to make three key points about bulk-format solid-state lasers:

- Compared to semiconductor lasers, the spatial region over which laser action occurs is generally many orders of magnitude larger, for both the laser beam area and the volume involved in lasing. Bulk solid-state lasers can generate much higher powers from a single laser in CW operation and even more peak power simply because of the larger areas and volumes involved. For efficient laser operation, it is important that (1) a large fraction of incident pump light is absorbed in the material and (2) the spatial overlap is high between the large pumped volume and the laser beam either generated or

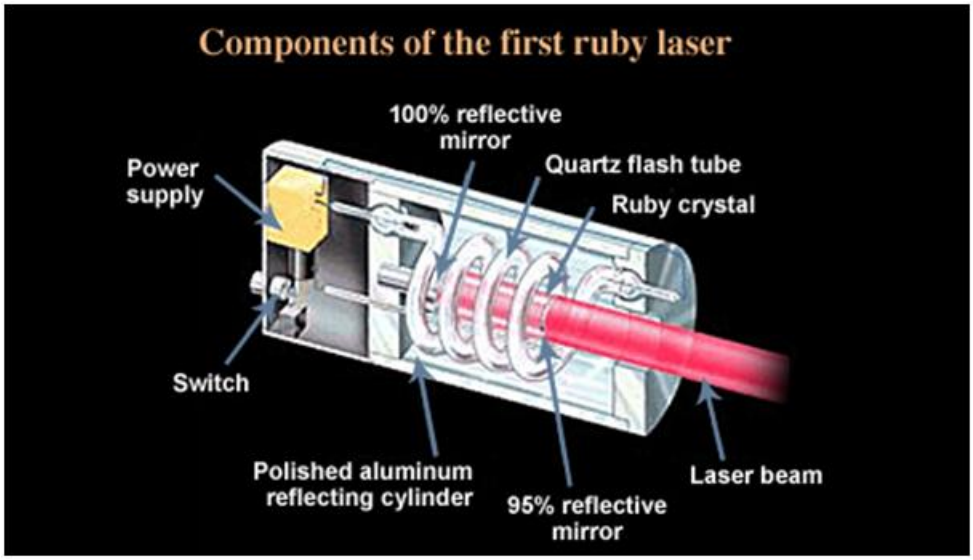

FIGURE 4-8 Schematic diagram of first ruby laser. SOURCE: Lawrence Livermore National Security, LLC and the Department of Energy.

amplified in the material. For the simple design in Figure 4-8, the latter is accomplished by making an optical cavity out of the rod itself.

- For optical pumping, at least some fraction of the absorbed pump power must be of a shorter wavelength than that of the laser, in order to create the inverted energy level population in the material. Generally, all of the energy difference between pump and laser photons is taken up via nonradiative processes that generate phonons, and the measurable effect is a rise in the laser material temperature. Unless the heat is removed, laser action will eventually stop from a variety of high-temperature effects, the most extreme being melting of the material. For the laser in Figure 4-8, heat from the rod could be removed by flowing air around the rod or, more commonly, by placing the rod inside a transparent cylinder and flowing a transparent liquid, usually water, around the outer rod surface.
- Once the laser material is actively cooled, since the material has a finite thermal conductivity, the flow of heat leads to a variation in temperature in the material. For the case of the rod in Figure 4-8, the center of the rod will be hotter than the cooled outside surface. All materials expand (or in a few cases contract) with increased temperature, and the temperature difference inside the rod leads to different levels of expansion, which in turn create stress in the material. Ultimately, at sufficiently high levels of power the material may fracture from this stress. Below the fracture limit, the optical properties of the material are changed. The stress changes the refractive index and polarization properties of the material through the stress-optic effect. More important, since the refractive index of materials changes with temperature, the gradient in temperature leads to a nonuniform refractive index. The net effect of the changes is to change the properties of the laser beam passing through the material. The changes are pump-power dependent.

Many of the engineering challenges and performance limitations of bulk solid-state lasers center on the issues described above.

Host Materials

The effect that the choice of host material has on these issues is considered first. The desirable host material properties are these:

1. Availability in a size suited for the pumping, lasing configuration; low optical loss at the pump and laser wavelengths; ability to support active-ion doping levels high enough to absorb the pump light in the desired volume.
2. High mechanical strength to avoid fracture at the desired power level, high thermal conductivity, and low thermal expansion coefficient.
3. High thermal conductivity, small change in refractive index with temperature, and small change in refractive index and polarization properties with stress.

Particularly for the rare earth dopants, several hundred different host crystals have demonstrated laser operation, but only a handful have found widespread use because they best satisfy all of the requirements listed above:

- *$Y_3Al_5O_{12}$ (YAG)*. A member of the very large family of garnet-structured crystals, YAG emerged from research in the 1960s as having the best combination of properties for hosting rare earth materials, including high thermal conductivity and good mechanical strength. It remains in wide use at present for a variety of industrial and medical applications and is also the basis for all laser-designator systems, as its primary operational wavelength when doped with Nd^{3+} ions, 1,064 nm, matches the current inventory of laser-guided munitions. A serious drawback is in the growth of single crystals, which require furnaces to operate above 1,950° C to produce a liquid from which crystals are pulled. (Reportedly, at one point the production of YAG in New Jersey represented the largest single use of electrical power in that state.) Recent advances in ceramic YAG, which is made up of randomly oriented, micrometer-sized single crystals, have allowed fabrication of much larger materials. The successful use of ceramic materials for lasers is a recent development that required significant advances in technology to essentially eliminate the scattering of light associated with the boundaries between the microcrystals making up the ceramic. The demonstration of the most powerful CW solid-state laser, 100 kW from Nd:YAG, as part of the Joint High Power Solid State Laser (JHPSSL) program at the U.S. High Energy Laser Joint Technology Office involved the use of ceramic laser material.[5]
- *YVO_4 (vanadate)*. One important parameter of the electronic transitions involved in pumping and laser action is the "cross section," given in units of area. When it is multiplied by the volume concentration of laser ions, the resultant value provides a measure of the pump absorption or laser gain per unit length. Electronic transitions for solid-state lasers ions generally take place in the same electronic shell (3-d or 4-f) and are thus "parity forbidden," which in practical terms means they would have very small, or in some cases zero, cross sections if they could somehow be suspended in free space. Once the ion is placed in a crystal, the surrounding ions create a lower symmetry than free space, mix in other electronic states to the electronic wave functions for the ion, and greatly enhance the cross section. This process is sometimes called activation of the transitions. For most of the host crystals for rare earths, the laser cross sections fall in a fairly limited range of values. The crystal YVO_4 and some related compounds have unusually low-symmetry surroundings for rare earths and induce cross sections about an order of magnitude higher than other hosts. This leads to efficient absorption of pump light in short lengths of material and very high optical gains for rare earths, allowing construction of diode-laser-pumped miniaturized devices, such the common "green" laser pointer, as well as moderate-power, efficient, high-pulse rate lasers for applications in laser materials processing and high-pulse rate laser imaging systems. By a fortunate coincidence, the prime Nd:YVO_4 wavelength of 1,064 nm overlaps that of Nd:YAG, making it possible to build specialized hybrid lasers for some applications.
- *$LiYF_4$ (YLF)*. The fluoride-based material YLF has several unique properties, including much weaker thermal distortion effects than oxide materials and natural birefringence (see below), along with favorable laser properties for some rare earth dopants. The main drawback is a reduced mechanical strength compared to most oxides.

[5] See http://www.acq.osd.mil/dsb/reports/ADA476320.pdf.

- *Al_2O_3 (sapphire)*. Sapphire has thermal and mechanical properties superior to those of all other host crystals, and when undoped it finds applications ranging from high-end watch crystals to tank windows. These properties were key to operation of the ruby laser, as many other host crystals would have fractured when pumped to the levels (several hundred Joules into the flash lamp) required by the 3-level Cr^{3+} transition. Unfortunately, sapphire does not accommodate rare earths at any reasonable doping level, and so after ruby it found little use in solid-state lasers until the demonstration of the Ti:sapphire broadly tunable laser. The sapphire host, besides having excellent thermo-mechanical properties, also situates the Ti^{3+} ion in a low-symmetry crystal environment, thereby inducing large absorption and emission cross sections. Due to this unique combination of properties, sapphire remains the only effective host for Ti^{3+}-based, broadly tunable solid-state lasers.
- *ZnSe*. The material ZnSe and similar II-VI compounds have in the past found widespread use for IR-transmitting windows, with acceptable levels of mechanical strength and hardness. Chemical vapor deposition (CVD) techniques have been developed to grow large slabs of polycrystalline (i.e., ceramic) material with very low optical losses. As a host material for Cr^{2+} and other divalent transition metals, it provides a low-symmetry (tetragonal) environment, which leads to lower-energy transitions with large cross sections. The low phonon energy of the material minimizes nonradiative decay of the longer-wavelength transitions characteristic of divalent ions in the tetragonal environment.
- *Glasses*. Glass materials represent a disordered, or amorphous, arrangement of atoms. For lasers, glasses are generally based on silicate or phosphate structures, or "formers." On a microscopic scale, active ions in glasses have slightly differing surroundings for each ion. The fine-detailed structure in the manifold levels for rare earth ions in glasses is thus smeared out (inhomogeneously broadened) when one measures their absorption and emission properties. This allows glass-based materials to have much broader absorption bands for pump light and broader continuous tuning ranges than crystal and ceramic hosts. On the other hand, the effective cross sections are lowered as well, reducing the gain and length for laser action. A big advantage for glass hosts is that they can be cast from liquid form into arbitrary and very large shapes to allow generation of very high energy pulses. The world's most energetic laser, the Nd:glass-based, 4-megajoule-level, National Ignition Facility in Livermore, California, uses about 3,000 large plates of laser glass, each 0.46 × 0.81 m in area,[6] occupying a space equivalent to three football fields. The significant disadvantage of glasses for bulk lasers is their poor thermal conductivity, a direct result of their amorphous nature. They are not capable of high-average-power operation and find use primarily in single-shot or low-pulse-rate applications ranging from handheld rangefinders (1 pulse/s) to the NIF (1 pulse/day). (A 4-MJ pulse in 24 hours leads to an average power of around 50 W, less than that of a 1-cm-long diode laser.)

Beam Quality and Thermal Effects

One of the significant challenges in bulk solid-state laser design and engineering centers on the effects of thermal distortion of the laser material's optical properties ("thermo-optic distortion") on the beam quality of the laser output generated by the material. The choice of pumping devices (lamps or diode lasers) plays an important role and is discussed in detail in the next section.

For bulk solid-state lasers, modern system designs for oscillators almost always employ "external" resonators formed by dielectric-coated mirrors, sometimes including internal lenses and sometimes using a coated end of the laser material as one of the resonator mirrors. The properties of the laser output beam are set by the resonator configuration, as modified by any thermo-optic effects in the laser material.

For materials shaped like simple cylindrical rods and cooled through the cylindrical surface, the temperature distribution can be approximated by a parabolic profile (transverse to the beam), peaking in the center of the rod. The corresponding optical effect is to make a rod behave, to a first level of

[6] See https://lasers.llnl.gov/about/nif/how_nif_works/amplifiers.php.

approximation, as a lens. The lens is positive for the majority of materials, which exhibit a positive change in the refractive index with temperature. Additional optical distortions come about from the stress-optic effect, which converts the stress in the material from the thermal gradient into a change in refractive index. In rod configurations the stress-optic effect can also be approximated as creating a simple lens. For many materials, the two effects are additive, but in a few, particularly fluoride-based materials, the change of refractive index with temperature is negative, while the stress term is positive, leading to a reduced level of "lensing." Smaller levels of optical distortion result from changes in the shape of the laser material due to nonuniform thermal expansion of the material as a result of the temperature gradient.

Another stress optic-effect is a distortion of the polarization of light passing through the material due to "stress-induced birefringence," which in rods leads to a polarization state that changes over the cross section of beam. For crystals that are naturally birefringent, such as vanadate and sapphire, the natural birefringence typically dominates stress birefringence, and one can avoid any change in, say, a linearly polarized beam by aligning the polarization with one of the crystal axes. Isotropic materials such as YAG and all glasses are, however, susceptible to stress-induced birefringence, and this is an issue since linearly polarized light is required to drive many of the nonlinear processes discussed below.

In theory, it should be possible with any bulk solid-state laser to effectively generate a large fraction of power from a single, diffraction-limited, transverse mode of the resonator, since the diameter of the single mode can be made arbitrarily large by the choice of cavity design. In practice, for the laser wavelengths and resonator lengths used for bulk solid-state lasers, operating with "stable" cavities that produce more than several millimeters of beam diameters is impractical, since either the required angular alignment of the mirrors becomes unreasonably exacting or the thermo-optic effects in the laser material make the beam properties highly dependent on the exact pump power of the system. For high-gain lasers, such as Nd:YAG-based devices running in a pulse-pumped mode, it is possible to employ so-called "unstable" resonators that can provide single-mode 5-10-mm diameters, but these have to date been employed only in laboratory-based lasers.

To get around the limits of cavities, one can design a relatively low-power laser oscillator optimized for single-transverse-mode operation, and then amplify the oscillator in one or several amplifier stages, where the laser mode is expanded by telescopes to fill the size of the pumped region in the amplifier. This is the standard approach for most high-power systems and is termed a master oscillator power amplifier (MOPA) system.

Ultimately, thermo-optic effects in the laser material will limit the beam quality of the system by adding phase distortions across the area of the laser beam. While simple external lenses can correct the first-order effects in rod-geometry designs, as the thermal gradients increase with pump power, the higher-order levels of distortion become significant. In recent high-power laser systems design, so-called adaptive optics, first developed to correct for the linear phase aberration effects of atmospheric turbulence on laser beams, can be applied to correcting the more complex distortions in materials running at high powers, typically through the use of feedback loops that seek to maximize beam quality.

There are several material geometries beyond simple rods that attempt to reduce thermo-optic effects. Notably, if the heat flow is along the direction of the optical beam, the thermal gradient is as well. When the material is uniformly pumped, all sections of the beam experience the same refractive-index change and thus the beam phase-front is undistorted. Typical designs employ thin plates, often in the shape of a disk, where the large surface can be cooled by high-velocity liquids or by contact with a transparent, high-thermal conductivity crystal. Power limits are from stress, as always, and this is minimized by the use of thin crystals. However, if the crystal becomes too thin, it will not absorb, in a single pass of light through the material, a sufficient fraction of the incident pump power.

A recent design, the so-called thin-disk geometry, has been made possible through the development of diode pump lasers (discussed below.) Here a thin disk of material (on the order of 0.1-mm thick) is cooled by contact of one of the large faces to a metal heat sink, with the contacted face employing a dielectric coating that is designed to be highly reflective at both pump and laser wavelengths. The design employs optics that arrange for the pump light to make many passes back and forth through the laser material, effectively increasing the optical path for the pump light and facilitating efficient

absorption. (This arrangement would not be possible with lamp pumping.) Ultimate limits to beam quality are set by the inevitable deviation of heat flow from the desired direction along the thin axis of the material and the resultant stresses. For thin-disk lasers the mechanical distortion of the material cannot be ignored as a thermo-optic effect, and there is a limit to how large an area of the crystal can be used while still maintaining a high beam quality. One disadvantage of the thin-disk design is the relatively low gain of the system because of the short length of active material, and typical designs are with oscillators only, with low-loss laser cavities.

Another geometry used for high-power, high-beam quality systems is the zig-zag slab design. In this case the material is fabricated as a thin slab, but with a rectangular geometry. The laser beam is arranged to reflect back and forth in the thin direction (hence the descriptor zig-zag), off polished faces of the slab. This has the advantage that all parts of the optical beam experience the same thermo-optic distortion in the zig-zag direction. Distortion in the other direction is minimized by designing the pump beam to be as uniform as possible. Lamp pumping is typically done through both large faces of the slab, with cooling of the same faces by a flowing liquid. Diode pumping can be done through one face, with the other face cooled by contact with a metal heat sink. Another configuration, with the diode pump light directed along the long length of the slab, is discussed in the next section. This is the configuration used in the most powerful bulk solid-state laser (as of this writing).

Many common solid-state lasers, such as those used for materials processing, laser target designation, range-finding, and imaging at relatively short distances do not require diffraction-limited output, and resonator designs for such systems emphasize some combination of efficiency and high stability against vibration or system temperature change.

The actual relation between power output and beam quality is set not only by the laser material and resonator design but also by the pumping source.

Pumping Systems

Incoherent, lamp-based pump sources were the standard for bulk solid-state lasers until the 1980s, when development of high-power diode lasers brought about a major revolution in high-performance systems.

Lamps are still widely used for systems where cost or simplicity is important. For lamps operating in the pulsed mode ("flashlamps"), the typical design uses approximately atmospheric-pressure xenon gas, typically contained in a fused silica tube with tungsten-based electrodes on both ends and driven by a high-current electrical discharge. Lamp diameters are from 3 to 19 mm, and lengths can range from 2 cm to 1 m, depending on the system and energy required. Well-designed xenon lamps are 50-60 percent efficient in converting electricity to light, which they produce in a spectrum that resembles a high-temperature (6000-10000 K) blackbody with discrete emission lines in the 820-1,000-nm region. As the lamp discharge current increases, the line spectrum becomes a smaller fraction of the total spectrum. The energy limit for a flashlamp is set by its "explosion energy," which is proportional to the inner area of the containing tube, and if the lamps are run at less than 30 percent of their explosion energy typical shot lifetimes before lamp failure are 10^7 to 10^8, with NASA claiming to have shown a 10^9-shot lifetime in one system.[7]

For CW lasers, one employs arc lamps, which differ in some details from flashlamps in the electrode and tube designs, as well as use of a multi-atmosphere fill pressure. For the most widely used laser application for pumping Nd:YAG lasers, arc lamps employ krypton rather than xenon gas, since the narrow-line krypton spectrum is a better match to the absorption spectrum of that material. Arc lamp lifetimes range from 200 to 2,000 h, depending on the operating cycle of the laser, with the longest lifetime seen for lamps running 24/7.

[7] L.J. Richter, F. Schuda, and J.J. Degnan, 1990, "Billion-shot flash lamp for spaceborne lasers," *Proc. SPIE* 1223: 142, Solid State Lasers, doi:10.1117/12.18406.

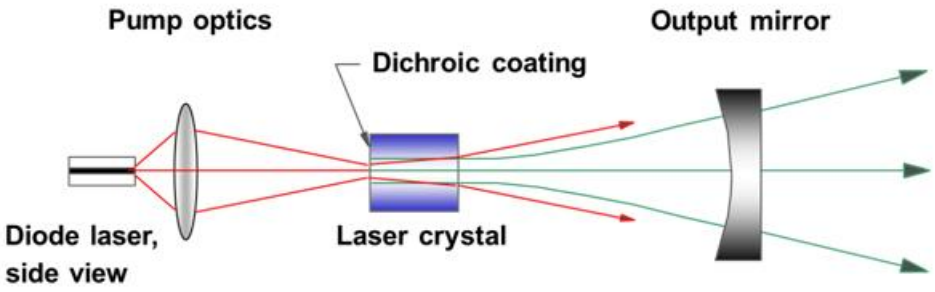

FIGURE 4-9 Schematic of diode end-pumped solid-state laser.

Lamp-pumped Nd:YAG lasers convert 2-3 percent of the lamp energy into laser output, with much of the lamp output not absorbed by the laser material. For flash lamps, as the pulse energy increases more of the pump power is delivered into higher levels, increasing the heat dissipated in the crystal. Despite these difficulties, lamps remain in widespread use, especially in pulsed lasers, for reasons to be discussed below.

The use of diode lasers as optical pumps for solid-state lasers was demonstrated in the 1960s, but the limited power and performance of early diode lasers kept them from widespread adoption. With improvements in semiconductor materials and processing technology in the 1980s, commercial high power single-emitter diodes with outputs of 0.2 W became available, with operating wavelengths around 808 nm, suited for pumping Nd:YAG lasers. The first important design scheme for diodes was the so-called longitudinal (or end-) pumping configuration, shown schematically in Figure 4-9. The single-emitter diode laser output is collected and focused into an approximately round beam at the surface of the laser crystal through appropriate optical design. The pump beam diameter, and the diameter of the fundamental laser-cavity mode formed by one coated mirror on the crystal and one curved external mirror, are designed to be about the same, which results in the solid-state laser output being single-transverse mode. Typical laser crystal lengths are from 1 to 10 mm.

Several major advantages of diode over lamp pumping are evident with the end-pumping design of Figure 4-9:

- Essentially all of the diode output, rather than the small fraction typical with lamps, can be absorbed by the laser material. Thus, even though lamps and diode lasers are roughly comparable in overall conversion of electricity to light, the diode output to laser output (optical-to-optical) efficiency is more than ten times greater than that of lamps.
- With the proper choice of wavelength, the amount of heating in the material from laser action for a given power is smaller with diode pumping and reaches a minimum when the pump wavelength directly excites the upper laser level in a normally 4-level laser. This leads, in turn, to less thermo-optical distortion of the laser beam by the laser material.
- As noted above, the correct choice of pump-beam and laser-cavity beam area produces a diffraction-limited output, an important characteristic. Diode lasers are limited in their ability to produce large diffraction-limited powers, but they can make high multimode powers. The diode-pumped solid-state laser acts as a “brightness converter” for high-power diode lasers, and even with the inevitable loss in absolute power from the diode pumps, the brightness of the solid-state laser is between one and two orders of magnitude brighter.
- Heat generated in the laser material can be extracted by bringing the outer surface of the laser material into contact with a high-thermal conductivity material, eliminating the need for flowing a coolant over the material. Combined with the generally more stable pump power from a diode laser, this can lead to very stable operation of the solid-state laser, important for, say, generation of a stable laser wavelength/frequency.

- Diodes can be powered by low-voltage power supplies or batteries, in contrast to the high-voltages needed to drive lamps.
- With improvements in diode fabrication, laser operating times before pump replacement are typically three orders of magnitude longer than with lamps.

Disadvantages of the diode pumps compared to lamps include the following:

- The much higher cost of diodes.
- The need to maintain the diode operating temperature within a fairly narrow range, since the wavelength of diodes does shift with temperature. For typical diode lasers in the 800-1,000-nm range, the tuning rate is about 0.3 nm/°C. For Nd:YAG pumping at 808 nm, where the absorption linewidth is about 2 nm, one must hold the diode temperature well within a 10° C range for a simple end-pumped design. More complex designs can work over much wider ranges by using several diodes, each of which works for a certain temperature range, optical designs that provide multiple passes of the pump light, and/or diodes that have built-in wavelength-stabilization optics.

Initially, with the first single-emitter diodes, the power available for pumping was well below 1 W. Today, the available power from a single device with a 100-μm-widestrip is on the order of 10 W. For higher powers there are a number of techniques that make it possible to combine the outputs of multiple single-emitters into one pump beam. Figure 4-10 shows a simple approach that converts the multiple outputs of a diode bar laser into a single fiber, whose output can then be used for end-pumping. Because the simple design does not take advantage of the fact that diode outputs are single-mode in one transverse direction, it produces a low-brightness pump beam. More sophisticated bar-based pump sources use complex imaging systems to "stack" the diffraction-limited beam components of multiple emitter outputs in a way that greatly improves brightness, leading to fiber-coupled devices that can provide CW power levels of 200-300 W from a fiber with a 200 μm diameter and a 0.22 numerical aperture fiber.

At some power level, the simple end-pumped solid-state laser design of Figure 4-9 becomes impractical. Even with improvements in diode brightness, the pump beam in the crystal diverges rapidly enough such that, for good overlap of pump and laser beams, the length over which pump power is absorbed must be kept to 1 cm or less. Power limits are due either to crystal fracture from the large thermal load in a small volume or a large amount of thermal distortion.

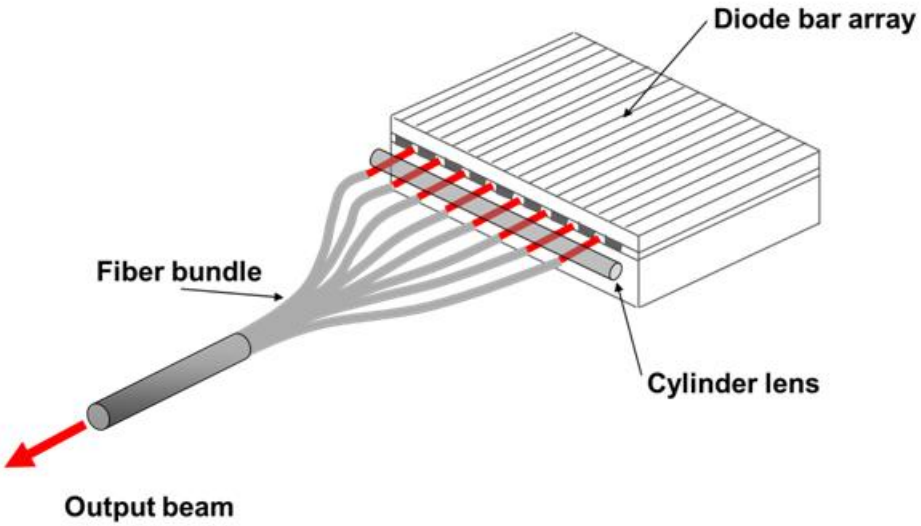

FIGURE 4-10 Optical schematic showing coupling of bar output to a single fiber.

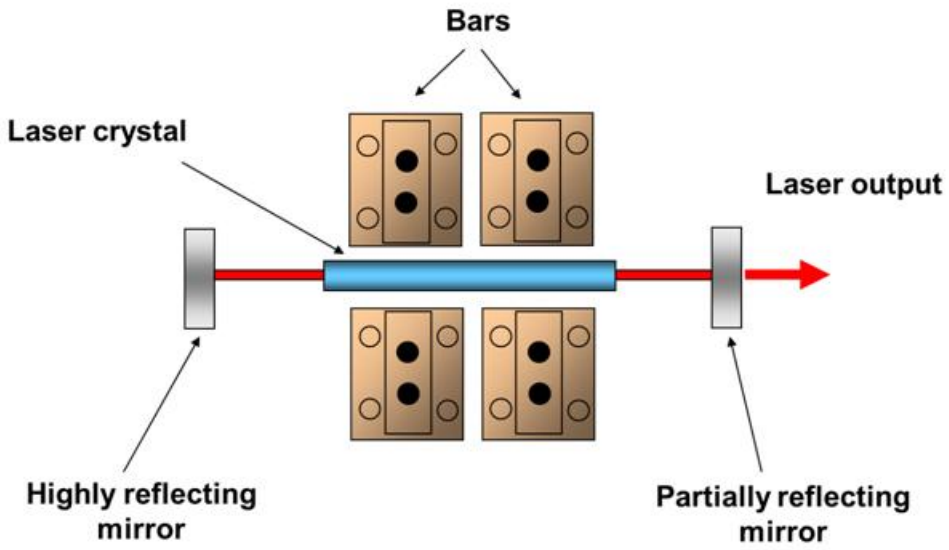

FIGURE 4-11 Schematic of diode side-pumped laser.

So-called side- or transverse-pumped designs with diodes can be employed to generate higher powers (and energies in the pulsed mode, see below), as shown schematically in Figure 4-11. Similar to lamp-pumped lasers, the pump light direction is roughly perpendicular to the laser direction, requiring the laser material to be fabricated appropriately. Cylindrical rods, thin, zig-zag slabs and other geometries can be employed, with cooling by either flowing liquid or bringing a portion of the material into contact with a metal heat sink. Obtaining good overlap with the lowest-order mode of the laser cavity is much more of a challenge than with end-pumped designs, and optical-optical efficiencies in general are lower. On the other hand, it is possible to couple much higher powers from the diode pump lasers, as many pumps can be used with one laser medium.

For side-pumping designs, diode bars are more effective in transporting light to the laser materials when they are closely stacked together, and extracting heat becomes a thermal engineering challenge. For pulsed systems with low average powers but high peak powers many more bars per square area can be put together than for CW systems. There are designs for low-average-power systems, as well as water-cooled bars designed for CW lasers such as that shown in Figure 4-12.

An example of the current state of the art of diode-pumped bulk lasers is the "gain module" design employed by Northrop Grumman in the 100-kW, CW laser built as part of the U.S. High-Energy-Laser Joint Technology Office, JHPSSL program. The module design has been commercialized by Northrop Grumman as part of the Vesta series of lasers.[8]

A Northrop-Grumman array employs two 36-bar vertical stacks with spacing between bars of 1.8 mm. Each bar has a cylinder lens attached to collimate the fast axis of the bar and, after that, a narrow-wavelength reflecting element (a volume Bragg grating) designed to form an external resonator and control the diode wavelength. For the device design used for the JHPSSL, the operating wavelength is set to center at 885 nm and efficiently pump the upper laser level of Nd:YAG, minimizing the heat generated

[8] See http://www.irconnect.com/noc/press/pages/news_releases.html?d=107455.

FIGURE 4-12 High-brightness, high-power stacked bars. SOURCE: Courtesy of DILAS.

in the solid-state laser material. The output of the two stacks is combined with an interleaving mirror to increase the intensity of the pump beam. Micro-channel water coolers, which provide high heat extraction per unit area, remove heat from the arrays and allow each 36-bar stack to generate a total CW power of 2.5 kW.[9]

Figure 4-13 (left panel) shows a schematic of the pumping and laser-material configuration for an earlier version of the gain module, based on the material Yb:YAG.[10] The solid-state laser material is configured as a thin zig-zag slab, with a pumping source on each end of the slab, and pump light entering through one of the slab faces, coated to act as a reflector for the pump power. As with longitudinal pumping, the pump and laser light propagate in the same general direction, with both reflecting off the large slab faces. A majority of the slab surface area is cooled by conduction to water-cooled metal heatsinks. The figure shows lens ducts used to collect the light from the pump arrays, while the newer, Nd:YAG-based JHPSSL uses simple lenses to properly format the pump light, as is evident in Figure 4-13 (right panel), a photograph of the Vesta gain module. Four gain modules for the basis for the lowest-power Vesta laser system, and generate 15-kW of CW power in a near-diffraction-limited beam. The JHPSSL system coherently combined the outputs of seven 15-kW modules to generate the record 100 kW of power.

Alternative Rare Earths

Diode pumping has also brought back into common use several rare earth dopants, all of which had been demonstrated in the early years of solid-state lasers, but subsequently abandoned in favor of Nd^{3+}-doped lasers for lamp-pumped lasers. These include:

[9] See http://www.northropgrumman.com/BusinessVentures/CEO/Documents/TechNote17_wavelength_stabiliz.pdf.

[10] G.D. Goodno, S.Palese, J. Harkenrider, and H. Injeyan, 2001, "Yb:YAG power oscillator with high brightness and linear polarization," *Opt. Lett.* 26: 1672.

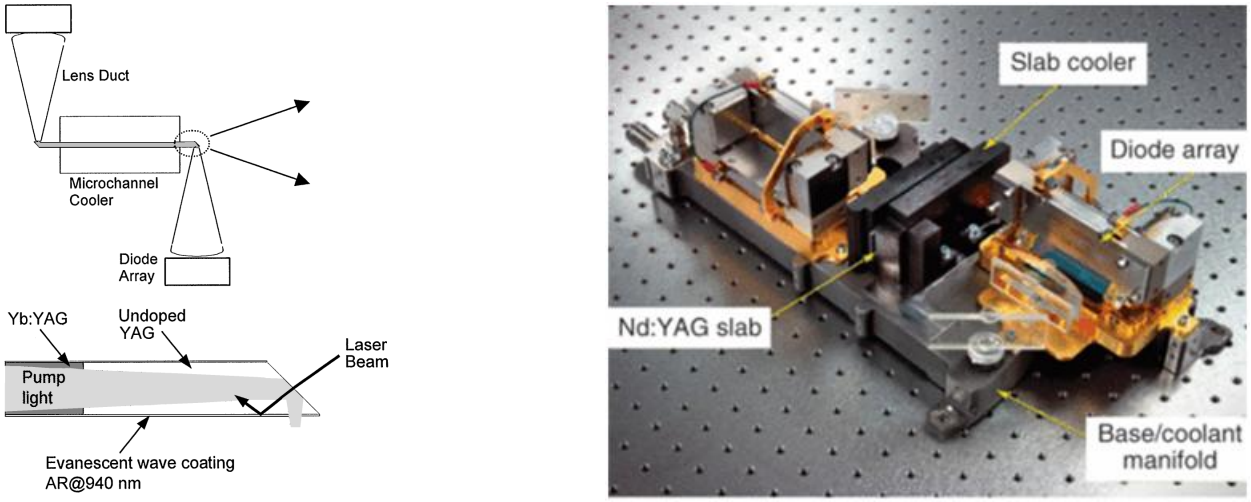

FIGURE 4-13 *Left*: Diagram of end-pumped slab design used for Vesta gain module. *Right:* Photograph of the gain module. SOURCE: G.D. Goodno, S. Palese, J. Harkenrider, and H. Injeyan, 2001, "Yb:YAG power oscillator with high brightness and linear polarization," *Opt. Lett.* 26: 1672 (*left*) and Courtesy of Northrop Grumman Corporation (*right*).

- *Yb^{3+}*. The Yb^{3+} has only a single 4-f electron and as a result has a very simple 4-fenergy-level structure, a ground manifold, and one excited-state manifold. There is enough energy separation among the levels within the ground manifold to facilitate 3½-level laser operation. For lamp pumping, the lack of excited levels leads to low efficiency in absorbing broadband pump power, but with diodes tuned to the ground-state-to-excited-state absorption band, efficient laser operation is possible. In Yb:YAG, the typical laser wavelength is 1,030 nm, while a common pump band is centered at 941 nm. The difference in pump and laser wavelength is much smaller than for typical Nd:YAG pump/laser wavelengths of 808/1,064 nm, leading to a thermal load in Yb:YAG that is about 38 percent that of Nd:YAG and 53 percent for the less common 885/1,064 combination. There are some drawbacks to Yb:YAG compared to Nd:YAG, including a much higher threshold for operation at room temperature due to the partial thermal occupation of the lower laser level, as well as a smaller gain cross section. If Yb:YAG crystals are cooled to cryogenic temperatures, laser operation becomes truly 4-level, the gain becomes comparable to Nd:YAG since the spectral linewidth decreases, and the thermal conductivity rises by an order of magnitude. Moreover, recent demonstrations of cryogenically cooled systems suggest they could find use as very high power, pulsed sources for a variety of applications. Other host crystals for Yb exhibit large gain linewidths, and these have been exploited to build diode-pumped, 100-femtosecond-pulse, high-energy sources. The Yb ionic size is very close to the Y-ion size in YAG and other Y-based materials and can be doped to levels as high as a 100 percent replacement of the Y ions. This is used to advantage in Yb:YAG thin-disk lasers, where typical doping levels can be on the order of 10 percent to raise the amount of absorbed single-pass power.
- *Er^{3+}*. The ground-to-first-excited-state transition of the Er^{3+} ion provides laser operation in the 1,530 to 1650-nm wavelength region, desirable for active sensors since it falls in the eye-safe wavelength range. These laser transitions are 3½-level, leading to the need for high populations of the upper level to reach threshold. In addition, the energy-level structure of the ion promotes upconversion processes, where two nearby ions in the upper laser level interact, with one ion dropping to the ground state and the other ion moving to a higher energy state. This process competes with optical pumping and makes it difficult to obtain efficient operation. The best way to minimize this process is to reduce the Er doping level, but that also reduces the absorption strength and makes pumping less efficient. This

problem is worsened by the relatively small cross sections for the Er transitions, for both absorption and emission. One solution to the problem is to add Yb^{3+} ions into the crystal. Energy put into the Yb energy levels can transfer over to Er ions, since the two rare earths have excited states that are coincident in energy, a process called sensitization. Materials can be doped with high levels of Yb ions to provide good pump absorption and low-enough levels of Er to minimize upconversion. (For lamp pumping, Cr^{3+} ions can also be added, and they in turn transfer energy to Yb ions.) Unfortunately, to date, the only host materials where the sensitization process is effective are phosphate-based glasses. With diode pumping, the average-power capabilities and efficiencies of glass-based Er lasers have been increased but are still limited.[11] Yb and Er glasses have found use in low-pulse-rate range finders, particularly handheld, short-range devices. In the fiber-laser section, hybrid systems that use fiber lasers to pump bulk lasers are discussed, and Er:YAG is one bulk material that has been employed in hybrid designs.

- *Tm^{3+} and Ho^{3+}*. As with Er-doped lasers in the 1,550-nm region, the common laser transitions for Tm- and Ho-doped materials in the 1,900-2,100-nm wavelength region, also eye-safe, terminate on the ground manifold and are thus 3½-level systems. Of the two ions, the Ho laser-transition cross sections are generally higher (by a factor of 5 in YAG, for example), but Ho has limited and weak absorption transitions for either lamp pumping or diode pumping. A common scheme for pumping employs Tm sensitization of the Ho laser, with Cr^{3+} sensitization added for lamp pumps. Pumping of Tm in the region around 795 nm has the special property that, with a high enough Tm doping level, there is a cross relaxation between pairs of nearby ion that generates two ions in the upper level per absorbed pump photon. Recent developments in diode-laser materials show promise for the emergence of efficient, high-power pump lasers that can drive the upper laser levels of Tm or Ho. The ions are of interest for application in a variety of eye-safe active sensors, including coherent-detection wind-sensing systems. Ho-doped bulk materials also feature in hybrid bulk-fiber solid-state lasers, discussed below.
- *Pr^{3+}*. All of the ions discussed so far generate wavelengths in the infrared. Pr-doped materials have laser transitions at green and red wavelengths, as well as some near-IR transitions. The ion has narrow and weak pumping transitions at blue and shorter wavelengths and suffers from cross-relaxation processes that reduce laser efficiency at high doping levels and thus can barely operate with lamp pumping. The emergence of GaN-based blue diode lasers has made it possible to construct diode-pumped systems, and as GaN technology improves, Pr-doped lasers may emerge as a new source of direct laser operation in the visible wavelength range, with possible applications in underwater sensing.

Energy Storage and Q-Switching

Pumping power, either electrical for diodes or optical for solid-state lasers, leads to excitation of the upper laser level. If there is no laser operation, and the pump power is turned off, energy remains stored in the laser material, which decays away by spontaneous emission of photons and, in some materials, also by generation of phonons through nonradiative relaxation. In most cases, the decay is exponential, characterized by a single number, the upper-state lifetime or "storage" time. For a given pump power, the stored energy is directly proportional to the decay time, provided that only a small fraction of the ions in the material have been excited. With laser action, the stimulated emission process that gives rise to optical gain acts to extract the stored energy. The rate of extraction (or inversely, the stimulated lifetime) depends on the power in the laser beam but is of the same order as that of the spontaneous emission.

For semiconductor lasers, storage times are in the nanosecond range at best and lower in quantum-cascade devices. Transitions in solid-state lasers have storage times between 0.1 and 15 ms, with the notable exception of the 3.2-μs lifetime of Ti:sapphire. Thus, per watt of pump power, stored energies are in the nanojoule range for diode lasers and the millijoule range for solid-state lasers.

[11] T. Yanagisawa, K. Asaka, K. Hamazu, and Y. Hirano, 2001, "11-mJ, 15-Hz single-frequency diode-pumped Q switched Er, Yb:phosphate glass laser," *Opt. Lett.* 26: 1262.

The ability to store large amounts of energy in the laser medium is one of the key advantages of bulk solid-state lasers and is unique to them. (As noted below, fiber-format lasers, although capable of high CW power levels, have major challenges in storing energy.) The process of Q-switching allows bulk materials used as laser oscillators to deliver the stored energy in very short pulses, thereby generating high peak powers useful for many applications, including, notably, active sensing.

Consider if the pump power to a laser is turned on but a shutter is put in the optical cavity to prevent laser action. The laser material will then, after several decay times, reach a steady amount of stored energy. Also assume that the pump power is well over the threshold for laser operation. If the shutter is rapidly opened, the power output of the laser will start to build up from the thermal noise background, as the laser power circulates back and forth between the mirrors. For typical solid-state laser optical cavity lengths and levels of laser gain, this buildup takes place on a timescale of tens of nanoseconds. Eventually the power in the cavity is sufficient to start to extract the stored energy by stimulated emission and the population of the upper laser level falls. Once the stored energy returns to the threshold level, power in the cavity stops increasing and the output of the laser falls, on a timescale commensurate with the time it takes for light to decay in the laser cavity without any optical gain from the laser medium, a quantity called the "cavity lifetime."

With the advent of diode-laser pumping, some new types of Q-switched lasers could be developed. So-called microchip lasers, made from millimeter-thick, longitudinally pumped active media such as Nd:YAG bonded directly to solid media that have saturable absorption at the laser wavelength, can generate pulses that last much less than 1 ns; this is useful for active sensing requiring high range precision.

This transient condition results in the generation of a pulse containing a large fraction of the energy initially stored in the laser material, with a width that is related to the round-trip time of light in the laser cavity. A key element in the process is the shutter, which, to avoid any loss of energy in the pulse, must change from a high-loss to a low-loss state in less time than the time required by the pulse to build up from the noise background. The shutter, for reasons dating back to the microwave origins of lasers, is called a "Q-switch" and the generation of pulses is corresponding called Q-switching.

For flash lamp-pumped or pulsed-diode-pumped lasers the optimum time at which the Q-switch goes to a low-loss state, or "opens," is at the point in the pumping process when the stored energy is greatest. To avoid losing energy to decay processes in the upper level, optimal pulse durations should be shorter than the storage time. For flash lamps, if the lamp energy is kept constant and the pulse is made too short, the risk of lamp explosion increases; in general, however, the same lamp running at a 10 J can produce a pulse lasting from 10 μs to 10 ms, thus operating over a range of powers from 1 kW to 1 MW. For diodes, which cannot be Q-switched and generally are limited to the nearly the same power independent of pulsewidth, the pulse energy decreases with pulsewidth, so that the optimum choice of pulsewidth is slightly shorter than the decay time.

CW-pumped, repetitively Q-switched bulk solid-state lasers running at 1-100 kHz and primarily based on Nd:YAG have found wide use in micromachining a variety of materials used by the electronics industry, where the short pulses facilitate precise drilling, forming, and marking. With diode-pumping, the performance of the devices has improved in terms of beam quality and power. End-pumped designs such as that shown in Figure 4-9, short optical cavities, and high-gain laser materials such as Nd:vanadate combine to enable the generation of pulsewidths at the 1-ns level at tens of kilohertz, opening up applications for 3-D imaging sensors with high data collection rates.

One related mode of laser operation is "gain-switching," whereby the pump power is in the form of a pulse whose duration is preferably shorter than or at least comparable to the laser buildup time. In this case the energy stored in the laser material can reach a high level before laser operation begins, and the result is generation of a short pulse similar to that of a Q-switched laser. Diode lasers can be gain-switched.

Clearly, with diodes, materials having long decay times lead to higher stored energies for the same pump power. The laser engineering trade-off is that there is an inverse relation between the gain cross section and the storage time, for desirable laser systems where the storage time is not reduced by

nonradiative processes. In addition, for the same lifetime, the gain cross section is inversely proportional to the spectral linewidth of the gain. Thus, materials system with long storage times may suffer from small cross sections, as do systems with large linewidths. One potential problem with low cross sections is that the energy per square centimeter, or fluence, in the laser pulse passing through the material has to be sufficiently high to efficiently extract the stored energy. A measure of the required level is the "saturation fluence," which is simply the energy of a laser photon divided by the cross section. If the fluence is too high the laser system optics, including the laser material itself, may suffer optical damage and cease to function. The damage level for optics in terms of fluence is a complex and often not-well-characterized function of the optical material, the surface (bare or dielectric coated), pulsewidth, beam area, pulse rate, and laser wavelength, with dielectric coatings often exhibiting the lowest threshold. Threshold fluences for damage can fall in the 1-100 J/cm^2 range for 10-ns pulses, with coated surfaces typically failing in the region 5-20 J/cm^2.

In the other direction, if a material has too high a cross section it may not be able to store the desired energy. At high levels of stored energy the optical gain may be so high that, for oscillators, the loss on the Q-switch (hold-off) may be insufficient. Even for infinite hold-off, the laser material may start to oscillate from optical cavities formed by unwanted reflections in the material. Ultimately, stored energy may be reduced by amplified spontaneous emission (ASE), where the light emitted by the normal decay process stimulates transitions from other excited ions, thus speeding up the loss of stored energy. With oscillator-amplifier chains, the gain through the entire system may get so large that the stored energy is limited by ASE.

Thus, there are a series of trade-offs regarding the choice of materials and system designs for pulsed systems working with Q-switched pulses. Even when the source of the pulses is not a Q-switched solid-state laser but some other device—say, a pulsed semiconductor laser or, as is discussed more below, picosecond and femtosecond pulses—similar issues arise.

Table C-3 lists the characteristics of several common bulk laser materials that impact design of short-pulse systems. There is a relation between cross section and lifetime that depends on the center wavelength of the laser, with the product of the cross section and lifetime following the 4th power of wavelength, thus favoring longer wavelengths to get long storage times for the same cross section.

In terms of lasers for use around 1000-nm wavelengths, the Nd-doped crystals (except for vanadate) provide a reasonable balance of good storage time and low saturation fluences but tend to be limited in storage by ASE issues. As noted above, the vanadate material, although not well suited for high-energy storage, provides a unique capability: it generates short pulses at high rates because of its large gain cross section. The smaller gain of Nd:glass allows significantly higher levels of storage before ASE limits set in, but glass hosts are not suited for high-average-power systems. (There is a continuing interest in finding new Nd host crystals that have glasslike storage capability but high thermal conductivity, but to date no material has proven useful.) Yb:YAG at cryogenic temperatures has a Nd-like saturation fluence combined with very low thermo-optic distortion, but the cooling requirements are a challenge for many applications. At room temperature the much higher fluence levels of optics with Yb:YAG is an issue for short-pulse systems.

For the longer eye-safe wavelengths, high saturation fluences are one of the reasons that Er-doped materials are not well-suited for high-energy, short-pulse lasers, while Ho-doped materials have proven capable of moderately high pulse energies with nanosecond-duration devices. The 15-ms lifetime of Ho:YLF promoted substantial energy storage with CW-laser pumping.

Ti:sapphire stands out from the other materials because of its very large linewidth, and it has an acceptably low saturation fluence. However, its storage time requires pump pulses to be microseconds or less for efficient operation. The most common pump for high-energy Ti:sapphire lasers is a Q-switched, frequency doubled, Nd-doped laser, and the Ti:sapphire laser typically generates a gain-switched pulse as a result of the short pump pulse. Cr:ZnSe, referred to as the "Ti:sapphire of the mid-infrared," also has a large linewidth and a short storage time and is emerging also as potential source of high-energy pulses around 2000 nm wavelength when pumped by Q-switched lasers. While the bandwidth for Cr:ZnSe in

terms of wavelength is larger than for Ti:sapphire, in terms of frequency it is less, and the latter actually determines the shortest pulsewidth for mode-locked operation, as discussed below.

Many moderate-energy solid-state lasers (for example, those based on Nd:YAG) still use flash lamps rather than diodes. If the goal is to generate a 100-mJ-energy Q-switched pulse, one would typically run the pump for 0.15 ms (compared to the storage time of 0.24 ms). Accordingly, with diode pumping, if there is 25 percent conversion of pump power, an array of diodes producing 400 mJ for 0.15 ms would be needed. This corresponds to about 2.7 kW of power. At an estimated current pricing of $3/W for a pulsed diode array, this leads to a cost of $8,000 just for the pump devices. The single lamp required to produce the same 100 mJ of laser energy costs about $100. Even with this cost difference, systems for man-portable applications may use diodes, since the higher efficiency and simpler cooling leads to a significant reduction in weight. For airborne systems, the longer lifetime of diode pumps may justify their use, given the high costs of removing a laser from the system to do a pump replacement.

Mode-Locking and Chirped Pulse Amplification

Even when a solid-state laser generates a single spatial, or transverse mode, it can generate a large number of longitudinal modes. The resonant frequencies of the longitudinal modes of an optical cavity correspond to a large integer times the speed of light divided by the round-trip optical path length of the cavity. For a 50-cm-long cavity, the frequencies are spaced by about 0.3 GHz. For Nd:YAG, the linewidth of the laser gain is about 180 GHz, so, in principal, that laser could oscillate on hundreds of longitudinal modes. In reality the modes compete for power in the laser medium, and in principle, if the laser is allowed to reach a true steady state, only one mode would remain operational. More typically, the modes occupy slightly different spatial regions in the material, and Nd:YAG lasers may have multiple modes that span a 10-GHz range. Remarkably, this is true even for materials like Ti:sapphire that have a linewidth 500 times larger, because of the competition among modes for gain.

The modes typically operate with random phase relationships that fluctuate with slight changes in operating conditions. If one places a time-dependent loss element in the laser cavity at exactly the frequency of the mode spacing, the lowest-loss condition for the laser leads to the modes acquiring, after some time, a fixed phase relationship among the different frequencies. If this is translated from the frequency to the time domain, this phasing leads to the generation of a steady train of pulses at a rate inverse to the mode spacing. Such phasing leads to pulses that time themselves to pass through the loss element at its minimum loss point. This general process is called “mode-locking.” The more modes that are locked together, the shorter the pulse, and how quickly the loss element changes from a high to a low state influences the number of modes. The major influence on the pulse duration is from the gain linewidth of the laser, and that puts an upper limit on the number of modes that can be locked, or, equivalently, a lower limit on the pulse duration. The product of the minimum pulse duration in seconds and the gain linewidth in Hz is a constant, depending on the exact temporal shape of the pulse, and falls in the range 0.3-0.4. For Nd:YAG, the minimum pulsewidth is about 2 picoseconds ps while for Ti:sapphire the value is 3.5 femtoseconds fs, and for Cr:ZnSe about twice that. Pulses that have the minimum value or duration for a given spectral linewidth are called “time-bandwidth limited.”

Typical mode-locking schemes could employ an electro-optic or acousto-optic loss modulator, but they have sinusoidal variations in loss and generally are not able to generate pulses that fully occupy the available linewidth of the laser. Passive loss elements based on semiconductors (or recently, graphene) that show a rapid reduction in loss with intensity can produce much shorter pulses. A breakthrough in mode-locking of the large-bandwidth Ti:sapphire laser occurred with the accidental discovery that the laser medium itself could provide a fast-switching loss element based on the nonlinear refractive index of sapphire. This led to generation of 100-fs pulses, and subsequent refinement of this

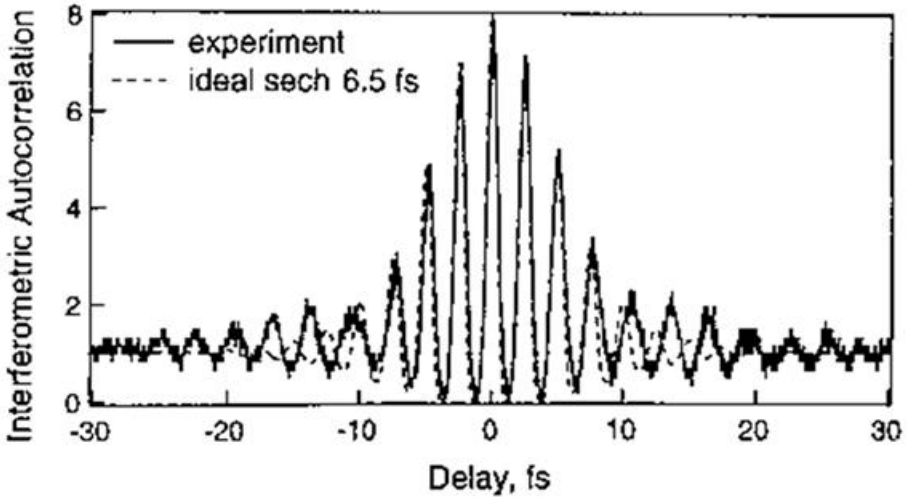

FIGURE 4-14 Interferometric autocorrelation signal from mode-locked Ti:sapphire laser showing data and fit to 6.5 ns pulse. SOURCE: I.D. Jung, F.X. Kärtner, N. Matuschek, D.H. Sutter, F. Morier-Genoud, G. Zhang, U. Keller, V. Scheuer, M. Tilsch, and T. Tschudi, 1997, "Self-starting 6.5-fs pulses from a Ti:sapphire laser," *Opt. Lett.* 22: 1009.

technique, combined with advances in the optical resonator mirrors, has led to Ti:sapphire lasers that can generate pulses approaching the 4 fs limit. The pulses contain only a few optical cycles, as is evident from the data in Figure 4-14, for a pulse with an estimated pulsewidth of 6.5 fs.[12]

Typical mode-locked oscillators operate at a pulse rate of around 100 MHz, with an average power of 1-10 W, leading to per-pulse energies of 10-100 nJ. For a Ti:sapphire laser producing 10-fs pulses, each pulse has a peak power of 1-10 MW, comparable to the peak power of high-energy Q-switched solid-state lasers. If one sends this pulse into a Ti:sapphire amplifier system, amplification to the relatively modest energy of 1 mJ leads to a peak power of 100 GW. At typical beam sizes, even if the fluence is comparable to the saturation fluence level, the high electric fields in the pulse will lead to substantial nonlinear effects in the material, as well as possible dielectric breakdown.

The chirped-pulse amplification (CPA) technique has been developed to circumvent the challenges of amplifying fs-duration pulses. When a short pulse passes through a dispersive optical element with an optical path that increases with decreasing wavelength ("normal dispersion"), the pulses get stretched in time, as the long-wavelength component of the pulse sees less time delay than the shorter wavelength component. Such elements include silica-based optical fibers at wavelengths shorter than 1,500 nm, specialized chirped volume Bragg gratings, and pairs of conventional reflection gratings with appropriate optics. The stretched pulses, increased to durations in the nanosecond range, can then be amplified by broadband gain media like Ti:sapphire that amplify all of the wavelength components. At that point, after amplification, the output beam area can be increased and the stretched pulse can be compressed, typically, by pairs of large-aperture gratings with suitable optics or, with more recent technology, chirped volume Bragg gratings. Both elements undo the stretching by arranging an optical path that provides a delay that increases rather than decreases with wavelength. The technique is now widely used for scientific investigations of the interactions of matter with extremely high optical fields and has emerged (see Chapter 3) as the basis for advanced active sensing techniques that rely on the formation of active filaments.

[12] I.D. Jung, F. X. Kärtner, N. Matuschek, D.H. Sutter, F. Morier-Genoud, G. Zhang, U. Keller, V. Scheuer, M. Tilsch, and T. Tschudi, 1997, "Self-starting 6.5-fs pulses from a Ti:sapphire laser," *Opt. Lett.* 22: 1009.

Although Ti:sapphire has been the standard source for generation and amplification of fs-duration pulses, there are many applications that benefit from pulses with central wavelengths longer than 800 nm or simpler laser systems. With bulk crystals, systems employing Yb- or Ho-doped media can provide pulses with durations in the 50- to 200-fs range, and CPA techniques can be employed for amplification. As discussed below, fiber lasers can generate pulses in this duration range, but with modest energies, and a variety of techniques combining bulk solid-state lasers and nonlinear optics can be used to provide broadband amplification of stretched pulses at high energies.

Fiber Format

Optical fibers employ highly transparent optical materials, structured to form a waveguide for light. The simplest fiber designs employ a circular cross section with a central circular region (core) that has higher refractive index than the surrounding material (cladding). Through the process of total internal reflection at the interface between the high-and low-index regions, light incident at a sufficiently low angle to the interface is reflected with no loss of power, and the only loss of power for the light contained in the core is due to the inherent absorption and scattering losses in the core material. Typical core diameters for fibers supporting only the lowest-order (diffraction-limited) mode are on the order of 5-10 μm. Through the development of glass-processing techniques, fibers based on silica (SiO_2) glass capable of transmitting light with minimal loss over kilometer-length distances became available in the 1970s and have developed to the point that fiber-optic-based media are the basis for data transport in the telecommunications industry.

Although it was recognized in the 1960s that the glass in fibers could be doped with rare earths and provide optical gain for communications systems, there was no practical pump source for the fibers and the early materials were lossy. Improvements in fiber-related glass technology in the 1980s and development of reliable diode pump lasers made it possible to deploy Er-doped fibers as amplifiers for telecommunications applications, where they eliminated the need to employ electronic repeaters after a certain distance in a fiber link.

Er doping was chosen since the 1,550-nm wavelength for the first-excited-state to ground-state transition provides good overlap with the lowest-dispersion wavelength region for silica-based fibers, thereby maximizing the data rate that can be transmitted. (The same dispersion that provides useful stretching of pulses for CPA systems is a major limitation when the goal is to keep the pulsewidths short in high-data-rate transmission systems.) The problem of upconversion that limits performance of Er-doped bulk lasers is eliminated in fibers by keeping the Er doping level low and using the very long absorption paths possible for pump light that is coupled into the core of the fiber. The other issue with Er-doped systems, the partial occupation of the lower laser level, is not a problem for fibers because large intensities of pump light can be established at moderate pump-power levels in the small cores, so that population inversions well over 50 percent are easily obtained.

For pumping the core of a fiber laser, with single-mode fibers to efficiently couple in pump light the pump must clearly be single-mode as well, and with diode lasers the powers required for telecommunications fiber amplifiers are well within the limits for single-mode devices. If one wants to extract power from the core-pumped fiber there is a limit for diode-pumped systems. A major breakthrough in fiber-laser technology, first demonstrated in the late 1980s, was the recognition that pump light could be coupled into a much larger cladding region surrounding the core if the cladding itself had an outer cladding of a low-index material, typically a specially formulated polymer (plastic) material or in some cases silica glass with fluorine added. Figure 4-15 is a schematic of a so-called double-clad fiber laser, where pump light is launched into the cladding region, and, as it propagates down the figure, passes through the core region, where it gets absorbed and pumps the active ions. The characteristic

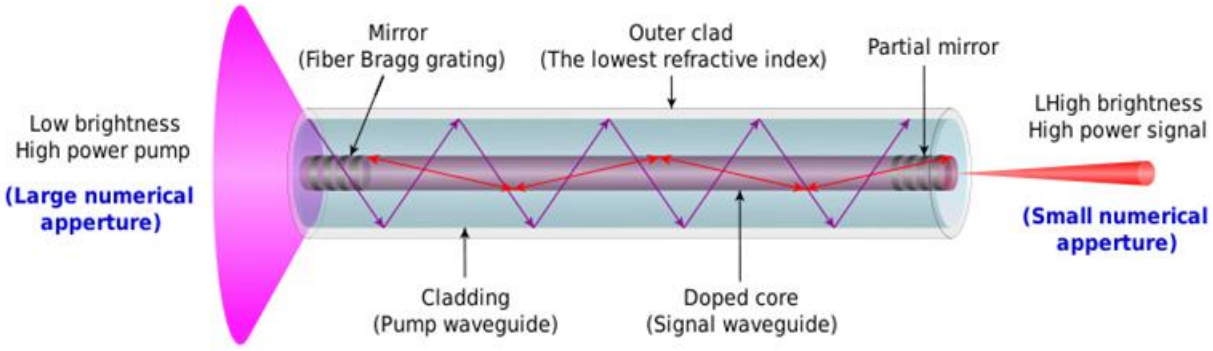

FIGURE 4-15 Schematic of pumped double-clad fiber laser. SOURCE: By Fl.png: Danielsoh8 derivative work: Das steinerne Herz (Fl.png) [CC0], via Wikimedia Commons. Available at https://en.wikipedia.org/wiki/File:Fl.svg.

length over which a certain fraction of the pump is absorbed is greater than the length for straight core pumping, in the best case by the ratio of the cladding to the core areas. With a purely cylindrical outer surface for the pump cladding, and a core located in the center, some of the pump light will never pass through the core, so actual double-clad fibers deviate from cylindrical surfaces to make sure that all the pump light does pass through the core.

With the development of double-clad fibers, fiber lasers were able to provide the same sort of "brightness enhancement" of pump light that bulk solid-state lasers routinely accomplish, albeit with pumping limited to diode lasers. As high-power diodes improved in performance, along with means to couple the power of many diodes into a multimode fiber, the power available from fiber lasers increased, crossing the 1 kW level for Yb-doped fibers in 2003.[13] Indeed, single fibers now are commercially available with multimode powers of 50 kW and single-mode powers exceeding 10 kW.

Since fibers are made of glass, it may seem puzzling that they can run at such high powers, given the limits of bulk glass lasers. The keys to high power generation in fibers are that (1) heat produced in the core from laser operation has a comparatively short (sub-millimeter) path to travel to the outer surface of the fiber, where it can be removed by active cooling;(2) the heat generation is typically spread out over a meters-long distance, so the thermal loading unit length is modest; and(3) the thermal gradient in the core itself is generally small, and the largest gradients are in the cladding region where laser operation occurs. Also, the fibers consist mostly of fused silica. The thermal lensing that does occur from the gradients is generally a small perturbation on the waveguiding of the fiber, which dominates in determining the spatial-mode properties of the laser beam.

There are ultimate thermal limits to fiber power, and, for operation in the 1000-nm range, Yb-doped fibers are favored over Nd-doped fibers because of the lower heat generated in laser operation. As with Er-doped fibers, the partial occupation of the lower Yb laser is not an issue in fiber-laser operation.

The small core size and long operating lengths that allow high CW power from fiber-format lasers also provide the biggest challenge for certain modes of operation. The long lengths and high intensities in the core promote nonlinear effects that are generally not seen with bulk lasers. For single-frequency generation, stimulated Brillouin scattering (SBS) of light, which involves an interaction with low-frequency ("acoustic") phonons in the glass, acts to amplify light in the fibers in a direction reverse to the laser propagation and can lead to destruction of optics at the beginning of a fiber-laser amplifier system, as well as limits to the power output. For typical core sizes and operating lengths in fiber lasers

[13] Y. Jeong, J. Sahu, D. Payne, and J. Nilsson, 2004, "Ytterbium-doped large-core fiber laser with 1.36 kW continuous-wave output power," *Opt. Express* 12: 6088.

this process may start with only tens of watts of power. SBS has a very limited gain linewidth, on the order of tens of magahertz, and can be circumvented by operating the fiber laser with a large spread of operating frequencies or adding frequency modulation onto a single-frequency input. Power levels on the order of 1 kW from Yb:fiber lasers have been produced via the frequency-modulation technique.[14]

Beyond SBS, for broadband fiber lasers, stimulated Raman scattering (SRS) provides an ultimate limit to power and for traditional small-core (5-10 μm diameter) fibers can set in at the 100-W level for typical lengths. High-power fiber lasers are operated with fibers having core sizes of 20-30-μm, where SRS limits are on the order of in the 10 kW.

For the pulsed mode of operation, SRS is clearly a limit, but fiber lasers are also much more subject to ASE problems owing to the long lengths of the gain region and the small volume of active material, and even though rare earths still have long storage times, conventional fibers are challenged to generate amplified pulse energies beyond 0.1-1 mJ because of ASE.

The glass host material for fibers produces large gain linewidths for the rare earth dopants and supports generation and amplification of pulses with durations on the order of 100 fs. At high pulse energies, beyond limits from SRS, the fact that the refractive index of the fiber material has weak dependence of on the electric field leads, over typical fiber lengths, to a distortion of the pulse spectrum and, eventually, the shape, through a process called self-phase modulation (SPM). As with bulk lasers, CPA techniques can and have been used with fibers to increase the energy outputs with 100-fs pulses, but the limits are considerably below those for bulk lasers.

All of these limits to operational modes can be mitigated by increasing the core size of the fiber. For single-mode operation, to use a larger core requires that the difference in refractive index between core and cladding be reduced, which also lowers the range of angles the fiber can guide (the numerical aperture, NA). At some point, the index difference becomes so small that fabrication of glass with the required uniformity and control of the refractive index becomes impractical; the amount that the fiber can be bent also becomes limited. One approach to increasing the single-mode capabilities of fibers is to fabricate a large-core fiber with a large NA that would typically support multiple transverse modes and then purposely wrap the fiber around a spool (typically in the 10-cm range), which acts to create high losses for all but the lowest-order, diffraction-limited mode. Fiber core diameters of 20-30μm diameter for operation around 1,000 nm can be used to generate single-mode output even when the fiber should be propagating multiple modes.

Conventional fibers are fabricated from solid glass, but developments in fiber technology have allowed development of so-called photonic crystal (PC) fibers that have a microstructure, on a scale of the wavelength of light, consisting of regular patterns of air holes in the glass cladding around the core. This is done by making the glass used to pull the fiber from capillaries rather than solid glass. The microstructure has special properties compared to the solid material, one of which allows fine control over the mode properties of the fiber and, ultimately, the use of larger core sizes for single-mode fibers than possible with solid glass. (PC fibers can also be made with very small cores when enhancement of nonlinear effects is desired.) The larger core sizes do not allow one to escape the problem of excessive bend losses with large cores, and the largest-core PC fibers are typically made with very thick glass claddings that prevent bending and that produce active material in the form of a long (typically 50-100 cm) millimeter-diameter rod that must be handled and mounted with care.

The use of PC fiber configurations has allowed single-mode generation with effective diameters of the optical mode of 50-100 μm and a large increase in the peak power available from fiber lasers. Recent results with Yb-doped fibers include generation at 1,040 nm of near-single-mode, 27 mJ energy, sub-60-ns-duration pulses at a 5 kHz rate, for 130 W of average power. The final amplifier stage had a

[14] C. Robin, I. Dajani, C. Zernigue, A. Flores, B. Pulford, A. Lanari, and S. Naderi, 2013, "Pseudo-random binary sequency phase modulation in high power Yb-doped fiber amplifiers," *Proc. SPIE* 8601, Fiber Lasers X: Technology, Systems, and Applications, 86010Z, doi:10.1117/12.2004486.

1.3-m-long rod stage.[15] For short-pulse operation, the same sort of large-mode PC fiber was employed with a CPA configuration to produce 500 fs, 2.2 mJ pulses at a 5 kHz rate.[16]

While these results show that reasonable peak powers can be generated by fibers, the use of rigid, rod-type structures raises the question whether the devices are closer to being bulk lasers rather than fiber lasers. Moreover, in terms of engineering, the use of gain media with meter-long lengths presents engineering challenges for ruggedizing the technology. The appeal of conventional fiber lasers is that the gain material can be easily bent to allow compact packaging, and also optical components in the laser can be fused together like telecommunications systems to eliminate alignment problems, an arrangement not possible with "rod" fibers. Finally, limits to average powers from rod-type gain media have been found related to thermal effects such as lensing, making them closer to bulk lasers in operational issues.

To date, the majority of work on high-power fiber lasers has involved Yb-doped silica glass, which exhibits optical-optical efficiencies of 70-90 percent. Er-doped silica works well for low-power telecommunications applications, but the Er-ion transitions that one could use to pump the laser are relatively weak, and too high a doping level leads to upconversion processes that reduce efficiency and gain. For typical double-clad designs at appropriate doping levels, the absorption lengths for the pump light are an order of magnitude or greater than for typical Yb-doped fibers, leading to long fibers that create more severe power limits from nonlinear processes. The most powerful Er-doped fiber lasers employ the same Yb sensitization process used with bulk lasers to reduce the length of fiber needed to absorb pump light. To date, this has allowed generation of around 300 W of output in the 1550-nm wavelength region, with the limit set by laser action from the Yb ion sensitizers starting up at high pump powers and stealing excitation.[17]

Tm-doped silica fiber lasers in the 2,000-nm wavelength region have advanced recently to CW power levels exceeding 1 kW.[18] The development allowing this was the recognition that, as with bulk-crystal lasers, if the Tm ion doping level is made high enough, pumping the ion in the 790-nm wavelength range generates two excited Tm states for one pump photon, leading to optical-optical efficiencies exceeding 60 percent, and the ability to use the high-power diode lasers available at 790 nm. To achieve the necessary high Tm doping levels, one needs to add a high level of Al-ion impurities to make more sites available for the laser ions. The nonlinear processes that limit powers for Yb-doped fibers are reduced at the longer wavelength, and, in addition, for the same fiber NA, one can operate a single-mode fiber with twice the core diameter of a Yb:fiber laser. The net effect is that power limits from SBS and SRS are considerably higher. To date, the highest CW, true single-frequency output from any fiber laser is the 600-W level obtained from a Tm:fiber laser.[19] The technology for Tm-doped silica fibers is not as advanced as that for Yb-doped devices, particularly regarding PC fibers, and the prospect for Tm-doped lasers to provide high peak powers and high-energy ultrafast pulses looks excellent.

Silica glass has thermomechanical properties that far exceed those of other glasses capable of being pulled into low-loss optical fibers. However, there is interest in so-called "soft" glasses for certain applications that need fibers doped to high concentrations of active ions, or with reduced multi-phonon

[15] F. Stutzki, F. Jansen, A. Liem, C. Jauregui, J. Limpert, and A. Tünnermann, 2012, "26 mJ, 130 W Q-switched fiber-laser system with near-diffraction-limited beam quality," *Opt. Lett.* 37: 1073.

[16] T. Eidam, J. Rothhardt, F. Stutzki, F. Jansen, S. Hädrich, H. Carstens, C. Jauregui, J. Limpert, and A. Tünnermann, 2011, "Fiber chirped-pulse amplification system emitting 3.8 GW peak power," *Opt. Express* 19: 255.

[17] Y. Jeong, S. Yoo, C. A. Codemard, J. Nilsson, J.K. Sahu, D.N. Payne, R. Horley, P.W. Turner, L.M.B. Hickey, A. Harker, M. Lovelady, and A. Piper, 2007, "Erbium:ytterbium codoped large-core fiber laser with 297 W continuous-wave output power," *IEEE J. Sel. Top. Quantum Electron.* 13: 573.

[18] T. Ehrenreich, R. Leveille, I. Majid, K. Tankala, G. Rines, and P.F. Moulton, 2010, "1-kW, all-glass Tm:fiber laser," SPIE Photonics West 2010: LASE Fiber Lasers VII: Technology, Systems, and Applications, Conference 7580 Session 16: Late-Breaking News.

[19] G.D. Goodno, L.D. Book, and J.E. Rothenberg, 2009, "Low phase-noise, single-frequency, single-mode 608 W thulium fiber amplifier," *Opt. Lett.* 34: 1204.

relaxation or for operation at longer wavelengths than possible from silica. Powers from the soft-glass fibers have tended not to exceed 100 W or so.

One of the most sought-after goals is the development of single-crystal fibers made from materials like YAG, but there are major technical challenges in fabricating practical, single-mode wave-guiding structures and, beyond that, double-clad structures for high-power pumping. The advantages of YAG and other crystals as a fiber material, compared to silica, include their much higher thermal conductivity, which would permit the use of shorter fiber before thermal limits set in. Crystal fibers would have narrower absorption and emission linewidths and, according to a recent study, a much higher threshold for SBS, at least in YAG.[20] The latter, combined with the shorter lengths, would allow generation of much higher single-frequency powers than possible with silica glass material.

The materials challenges with fiber-based lasers are major, and advances in high-power fiber lasers have been able to leverage the substantial interest in development of low-loss silica fibers for telecommunications. The main challenge is in making any fiber material low-loss enough so that light is barely attenuated by scatter and other imperfections over a path of 10 m, much more difficult than making laser crystals with a typical path of 1-10 cm.

Advances in diode-laser technology will contribute to further advances in fiber lasers. The cladding diameter for double-clad designs must be large enough to guide the desired pump power. There is interest in employing diode lasers for direct cutting and welding of materials, and that has motivated development of schemes such as (1) wavelength beam-combining of multiple diode arrays[21] or (2) techniques to combine high-brightness, single-emitter diodes,[22] to increase the brightness of diode-based sources. These schemes will allow double-clad fiber lasers to employ smaller-diameter claddings and reduce the length of fiber required to absorb the pump power, which will act to increase any nonlinear limits to power. Ultimately thermal limits will set in, but for many fiber-laser systems those limits have not been reached.

A state of the art, 1070 nm, 10 kW, CW, single-mode, Yb:silica fiber laser is offered as a commercial product by IPG Photonics.[23] The key to operation at such a high power is the use of fiber-laser pumps for the output amplifier, in this case 57 270 W, 1,018 nm single-mode Yb:fiber lasers, which are combined in a fiberoptic element and coupled into the pump cladding of the amplifier fiber. The latter operates with a core diameter of 30 μm and a length of 15 m. The small difference in pump and laser wavelengths reduces the heat load in the fiber, while the use of fiber-laser pumps provides a high-brightness source that minimizes the required cladding diameter for the amplifier fiber. The spectral linewidth of the output is very broad (10 nm) and prevents SBS; the laser operates close to the limit set by SRS. The JHPSSL 100 kW bulk crystal laser produced single-frequency power through combination of multiple, single-frequency, 15 kW modules, and thus fiber-format lasers have a considerable challenge at this writing in matching the narrow-linewidth output of bulk materials.

Research and development of high-power fiber lasers is being conducted throughout the world for applications in material processing, sensing, and directed-energy systems. The latter includes work to coherently combine the outputs of many fiber lasers to go beyond the power limits of single fibers, either through appropriate phasing of narrow-linewidth sources or spectral beam combination via dispersive elements such as gratings. Phasing schemes provide the opportunity to do rapid beam steering but pose significant engineering challenges to establishing reliable operation and countering nonlinear effects in

[20] B.T. Do and A.V. Smith, 2009, "Bulk optical damage thresholds for doped and undoped, crystalline and ceramic yttrium aluminum garnet," *Appl. Opt.* 48: 3509.

[21] B. Chann, R.K. Huang, L.J. Missaggia, C.T. Harris, Z.L. Liau, A.K. Goyal, J.P. Donnelly, T.Y. Fan, A. Sanchez-Rubio, and G.W. Turner, 2005, "Near-diffraction-limited diode laser arrays by wavelength beam combining," *Opt. Lett.* 30: 2104.

[22] See http://www.directphotonics.com/images/technology/whitepaper_advancedpumping.pdf.

[23] V. Gapontsev, V. Fomin, and A. Yusim, 2009, "Recent progress in scaling high-power fiber lasers at IPG photonics," 22nd Annual Solid State and Diode Laser Technology Review, Newton, MA, July.

the fibers. At this point, none of the beam combining techniques has moved out of the basic developmental level to applications, and would present an expensive approach for active sensors. However, the technology may advance to practicality for use in the future. In general, the next 10-15 years are likely to see important advances in a number of areas related to high-power fiber lasers.

Hybrid Laser Designs

Two hybrid laser designs combine fiber lasers as pump sources with bulk materials as the laser medium providing the desired output. Both are of interest to active-sensing applications since they provide sources capable of high peak powers with high beam quality.

As mentioned above, Er-doped materials provide operation in the eye-safe wavelength region around 1,600 nm but are challenged by the fundamental properties of the Er ion. Er:YAG can provide high energy storage but with a high saturation is problematic for high-energy pulse generation. The best operation of Er:YAG to date employs crystals with low doping levels to reduce upconversion and longitudinal pumping with a CW, Yb, Er:silica fiber laser operating around 1,530 nm. With the high-brightness pump, laser designs can operate with small cavity mode sizes consistent with diffraction-limited output. The fiber-laser pump source allows high-intensity pumping over a sufficiently long crystal to absorb all of the pump despite the low doping. Another benefit is distribution of pump power over a long distance in the crystal, reducing stress and allowing higher-average-power generation. Finally, the small difference in pump and laser wavelengths minimizes the heat load in general.

More power and higher pulse energies have been generated by Tm:fiber pumped Ho-doped crystals, including Ho:YAG (pumped at 1908 and lasing at 2090 nm) and Ho:YLF (pumped at 1940 and lasing at 2050 nm). As Table C-3 shows, the high cross sections and resultant lower saturation fluences allow higher gain and pulse energies compared to Er-doped materials. The fiber pump provides the same advantages in terms of beam quality and reduced heating as for Er-doped lasers. Besides being useful for active sensors, the Q-switched, Ho-doped lasers provide good pump sources for Cr:ZnSe and similar mid-IR tunable lasers and have facilitated recent CPA-based amplification of a mode-locked Cr:ZnSe laser to the 1 GW level (0.3 mJ in a 300 fs pulse), with promise for high energies and shorter pulses. Table C-4 summarizes the recently reported properties of Er-and Ho-doped hybrid lasers.

NONLINEAR-OPTICS-BASED SOURCES

The development of lasers enabled the high optical intensities required to observe nonlinear optical effects in materials. Generally, light passing into a material interacts with the electrons and, at longer wavelengths, the vibrational states of the atoms in the material. For transparent materials, the main interaction is to cause the electrons to move (be polarized) at the same frequency as the incident light, in linear proportion to the electric field of the light. The interaction between the light and the polarization causes the speed of light in a material to slow down by an amount related to the (linear) refractive index of the material. The intensities of light generated by lasers allowed observation of nonlinear effects in the electronic polarization, where the polarization has elements proportional to the square of the electric field, the cube, and so on.

One can use symmetry arguments to show that squared (second-order) terms in the electronic polarization in crystals can be possible only in materials that lack inversion symmetry. This implies that if one moves from an atomic location in one direction in the crystal, and compares the surrounding environment with what one would experience by moving in exactly the opposite direction, there is at least one direction where the environment would be different. There are many materials that lack inversion symmetry, including all of the III-V and II-VI compound semiconductors, as well as crystalline quartz (SiO_2), but not silica glass, which, like all truly amorphous materials, exhibits second-order nonlinearities

only over very short lengths of material. Two important nonlinear processes are discussed next, both of which depend on second-order nonlinear effects in crystals.

Harmonic Generation and Related Processes

The immediate effect of the presence of second-order terms in the polarization is to generate an optical frequency that is twice the incident frequency. The first observation of this was in a quartz crystal with a ruby laser beam incident,[24] where a faint amount of light was detected at the second harmonic of the 694.3 nm ruby laser, at 347.15 nm.

Analysis of the second-order nonlinear process shows that the power in the second harmonic can grow along the path of light in the crystal if the input, "fundamental" wavelength light and the second-harmonic light can stay in phase. If they do not, the power initially grows along the path in the crystal, then starts to decrease as the two beams get out of phase, falling to zero in a distance called the "coherence length," then continues to grow and fall in an oscillatory manner along the path of the light. In all materials, dephasing occurs in general because of the change of the linear refractive index (dispersion) with wavelength, which makes the fundamental and second-harmonic light get out of phase.

In birefringent crystals the refractive index for light in the crystal depends on the direction of the light in the crystal as well as its polarization. If one arranges the direction of light and the polarizations of the fundamental and the second-harmonic light to have the same refractive index in the crystal, the second-harmonic power will stay in phase with the fundamental and grow in power to constitute a substantial fraction of the fundamental power. This technique, called "birefringent phase-matching," was developed shortly after the first demonstration of second-harmonic generation in crystal quartz.

In general, light that propagates in a birefringent crystal exhibits a unique "walk-off" effect, in which the flow of power in the light beam deviates from the propagation direction. In phase-matching schemes, either the fundamental or the harmonic beam may exhibit walk-off, which serves to limit the length of crystal over which the harmonic-generation process occurs. In rare crystals, phase-matching can take place whereby the two beams propagate along one of the so-called principal axes of the crystal, in which case there is no walk-off. This is called non-critical phase-matching (NCPM), while the more general case is critical phase-matching (CPM). Often, NCPM is achieved by adjusting the crystal temperature to tune the refractive indices to the correct value. One advantage of NCPM is that the sensitivity of phase-matching to beam angle is much less than for CPM, making crystal alignment less of an issue, but more importantly, NCPM can provide efficient operation with non-diffraction-limited beams that have a relatively high spread of angles.

Not all materials that lack inversion symmetry have the necessary birefringent properties to achieve phase-matching. Even fewer can achieve NCPM. In addition to the necessary symmetry and birefringent properties, crystals used for efficient harmonic generation must be highly transparent at the fundamental and harmonic wavelengths in order to avoid heating, which not only creates optical distortion but can disturb phase-matching by changing the material's refractive properties. While the ability to temperature-tune phase-matching can be an advantage, materials that have a strong temperature dependence of phase-matching may not be able to generate high average powers, since they will de-tune even if the crystal absorbs a small amount of power. Given that high peak powers are often used, the material must also be resistant to optical damage. Unfortunately, materials that exhibit high nonlinearity are also more susceptible to optical damage from unwanted processes, created by higher-order nonlinear effects that generate free electrons in the material. Some materials are also susceptible to photorefractive effects, created by the interaction of high-intensity light with impurities in the material, leading to

[24] P.A. Franken, A.E. Hill, C.W. Peters and G. Weinreich, 1961, "Generation of optical harmonics," *Phys. Rev. Lett.* 7: 118.

creation of trapped electrons, an internal electric field, and a change in refractive index from this field that destroys the phase-matching condition.

The second-order term that allows generation of a harmonic can be viewed as the combination of two wavelengths (of the fundamental light) to produce a third, at the second harmonic. A more general interaction is possible where the two wavelengths are different, and with appropriate phase-matching, it is possible for a crystal to generate a third wavelength that is the sum ("sum-frequency generation") or the difference of the two wavelengths ("difference-frequency generation"). The most common interaction of this type is in so-called "third-harmonic" generation, where one crystal generates some fraction of light in the second harmonic, and another crystal combines the second harmonic with the fundamental light to produce the third harmonic. (Although it might be possible to utilize the third-order nonlinearity of a material to do this in one crystal, this nonlinearity is much weaker in materials, and a sequential process involving the second-order nonlinearity is much more efficient.) "Fourth-harmonic" generation involves the use of two second-harmonic crystal in series, the second arranged to double the harmonic generated by the first.

Given the many requirements for effective harmonic generation, it should not be surprising that there are a limited number of crystals that have been used with success in harmonic-, sum-, and difference-frequency generation applications. Table 4-1 lists the most widely used materials, along with some key properties and comments. The materials are listed in order of their first reported development. The quantity used to characterize the degree of second-order nonlinearity is a tensor called the *d* coefficient, which relates the nonlinear polarization in the material to the product of the electric fields of the incident beam or beams. In general, this quantity depends on the direction and polarization of the beams in the crystal. Analysis of the conversion efficiency of materials to the second harmonic shows that it depends on the square of the *d* coefficient for the particular interaction divided by the cube of the material's refractive index *n*, and this quantity is often used as a simple figure of merit for a nonlinear material, though the efficiency of the process depends on a number of other properties set by phase-matching in the material.

The list of materials is limited to those used with lasers in the 800-2,000 nm range. There are semiconductor-based materials that can be used for harmonic generation of CO_2 and similar gas lasers, but they are more widely employed for parametric generation, covered in the next section.

Given all of the constraints listed for nonlinear materials to work well, it may not be surprising that new nonlinear-crystal development features major advances on a 10-year timescale at best and seems to have slowed down in the last two decades. The materials listed above emerged from 1960 to the late 1980s and no major new materials have emerged since then. Advances in improving the size and quality of the materials listed do continue.

Even though LBO does not have the largest figure of merit of the materials shown, it is not subject to photorefractive and other optical damage issues and also exhibits an unusually low level of absorption at both the fundamental and harmonic wavelengths. As a result, it has been used for the generation of harmonic powers of several hundreds of watts at visible wavelengths and tens of watts at the Nd-laser third harmonic, limited more by the properties of the drive laser than the material itself. Conversion efficiencies are 50-65 percent for second-harmonic systems and 30-50 percent for third-harmonic devices. LBO is also widely employed in commercial CW, diode-pumped Nd-doped lasers, with intracavity doubling, to generate CW diffraction-limited powers at green wavelengths as high as 20 W.

TABLE 4-1 Common Birefringent Crystals Used for Harmonic and Related Generation.

Crystal (Acronym)	Figure of merit d^2/n^3	Transparency (nm)	Comments
KH_2PO_4 Potassium dihydrogen phosphate (KDP)	0.045	200-1,500	Several variants available, can be grown in very large sizes. Major use in fusion driver systems such as the National Ignition Facility, for third-harmonic generation.
$LiNbO_3$ Lithium niobate (LN)	30	330-5,500	Large nonlinearity. Major limitation due to photorefraction: not suited for generation of visible light except when doped with magnesium
$K(TiO)PO_4$ Potassium titanyl phosphate (KTP)	2.4	350-4,500	Major application in high-power second-harmonic generation of Nd-doped light, where high beam quality is not required, or in low-power devices such as laser pointers. Some long-term degradation due to formation of light-induced defect centers.
β-BaB_2O_4 Beta barium borate (BBO)	1.3	196-2,600	Robust material with capability for CPM over a wide range of wavelengths. Requires high beam quality for efficient operation. Used now primarily for fourth-harmonic generation of nanosecond-duration lasers or general harmonic generation with fs-ps pulsed lasers.
LiB_3O_5 Lithium borate (LBO)	0.18	160-2,600	Very robust material, can do NCPM for Nd-doped lasers with temperature phase-matching. Widely used in commercial CW and pulsed laser products for green- and UV-light generation.

One significant advance in nonlinear devices has allowed them to exploit the properties of existing crystals. Early in the initial work in nonlinear optics, in the 1960s, theoretical studies showed that another way to phase-match nonlinear generation was to make an artificial crystal by arranging alternating layers of material so that the sign of the nonlinear coefficient for the process of interest alternated in a periodic manner. As noted, without normal phase-matching, power along the beam path alternates in a distance given by the coherence length. If a crystal is structured properly, instead of power converting back to the fundamental due to dephasing, the change in sign of the nonlinearity results in power continuing to convert to the harmonic. While the buildup of harmonic power per length of material is not as rapid as with normal phase-matching, this quasi phase-matching (QPM) will allow efficient harmonic generation in nonbirefringent materials, and has features in terms of angular acceptance that are common with NCPM.

The first attempts at QPM employed materials that were polished into thin slices and bonded together with periodically alternating orientations. However, losses at the interfaces were too high for effective generation. Another approach, possible with certain materials that exhibit the ferroelectric effect, employs high-voltage electric fields to orient the crystal structure. Through a combination of lithography patterning of electrodes, setting up electrodes that alternate at a period of a few to tens of microns, and applying the appropriate voltage, one can permanently "pole" a ferroelectric crystal with an alternating

pattern of nonlinear-coefficient signs.[25] Such crystals are called periodically poled (PP). Crystals involved have included some of those listed in Table 4-1(LN and KTP) as well as variants such as $LiTaO_3$ and Mg:LN, which offer better resistance to photorefractive degradation. While these ferroelectrics can also employ birefringent phase-matching, QPM provides phase-matching over a much greater range of wavelengths, limited by the transparency of the material, and also allows use of crystal orientations that maximize the nonlinear coefficient. One drawback of ferroelectric QPM is that the thickness of the material is limited by the length of material over which poling is possible, typically on the order of 1 mm, although recent advances have shown that in LN doped with MgO thicknesses more than 5 mm are possible.[26]

Another approach to fabricating QPM materials employs semiconductors, most prominently GaAs,[27] where lithography and semiconductor processing technology is employed to form a patterned, thin substrate of material. One can then employ thick-film growth techniques on this substrate to fabricate QPM crystals with (to date) 1-2 mm thicknesses.

As with the development of nonlinear crystals, progress in new QPM materials has been slow and requires dedicated, long-term efforts. Practical ferroelectric and semiconductor devices emerged from 10-year-long, collaborative academic-industry programs. While QPM nonlinear materials can be used for harmonic generation, they have seen more widespread use in parametric generation and will be discussed in more detail in the next section.

Optical Parametric Generation

When there is no phase-matching in a crystal with a second-order nonlinearity, the harmonic wavelength converts back into the fundamental as it propagates in the crystal. In fact, this process, where the harmonic generates longer-wavelength light, is a special case of "optical parametric generation," which is a descendant of a process first observed and employed at radio and microwave frequencies. In general, the same second-order nonlinearity that can produce shorter wavelength light also allows a process in which a beam of light at one wavelength converts into two beams, where, for energy-conservation reasons, the sum of the frequencies of the two beams has to sum to the frequency of the original beam. The optical parametric process thus allows generation of longer wavelength light. Of the two beams created, the one with the shorter wavelength is called the signal while the longer-wavelength beam is the idler, and the original beam is called the pump, terms that derive from the microwave inheritance.

Solution of the equations for the interaction of the pump, signal, and idler beams show that with phase-matching of the three waves, application of the pump light in the crystal leads to optical gain for both the signal and idler waves along the path of the pump beam. The signal and idler build up together, so if one uses the parametric process to amplify an input at the signal wavelength, creating an optical parametric amplifier (OPA), power builds up in the idler wave, and vice versa. (Every signal photon generated is accompanied by an idler photon.) If one employs an optical cavity around the crystal for either the signal, the idler, or both, the device is called an optical parametric oscillator (OPO). OPOs where the cavity is resonant for either the signal or idler are called singly resonant, while doubly resonant devices attempt to resonate both frequencies. The latter devices are rarely employed because instabilities arise in the output owing to interactions between signal and idler frequencies. When the pump intensity is

[25] M.M. Fejer, G.A. Magel, D.H. Jundt, and R.L. Byer, 1992, "Quasi-phase-matched second harmonic generation: Tuning and tolerances," *IEEE J. Quantum Electron.* 28: 2631.

[26] H. Ishizuki and T. Taira, 2005, "High-energy quasi-phase-matched optical parametric oscillation in a periodically poled MgO:LiNbO3 device with a 5 mm × 5 mm aperture," *Opt. Lett.* 30: 2918.

[27] L.A. Eyres, P.J. Tourreau, T.J. Pinguet, C.B. Erbert, J.S. Harris, M.M. Fejer, L. Becourarn, B. Gerard, E. Lallier, 2001, "All-epitaxial fabrication of thick, orientation-patterned GaAs films for nonlinear frequency conversion," *Appl. Phys. Lett.* 79: 904.

high enough and the interaction length long enough, the gain in the OPA may be high enough to allow significant signal or idler power to build up from noise, and the device is referred to as an optical parametric generator (OPG). Parametric generation where the signal and idler beams have the same wavelength is described as degenerate, and gain is at a maximum for this case.

In contrast to lasers in general, optical parametric devices have no energy storage, as gain is present only when the pump power is present. Their ability to generate peak power comes about purely from the peak power of the pump. Energy conservation requires, at best, that the sum of the energy in the signal and idler outputs must equal that of the pump, with the energy apportioned according to the wavelengths/frequencies of the signal and idler. In the amplification process, as the signal and idler build up in power, the power in the pump beam is depleted. In theory, it is possible for all of the energy in the pump to convert to signal and idler power, but in practice, due to non-uniform power distribution in the pump beam parametric devices will achieve about 50 percent of full conversion, though some devices have pushed to 90 percent. Barring any optical absorption in the crystal, the parametric process is completely lossless, and the sum of the power in all three beams remains constant. As a result, no heat is, in theory, deposited in a parametric device.

Given that the gain only occurs during the pump, the time needed for beams to pass back and forth in the optical resonator must be accounted for in constructing an OPO. If the resonator is too long, the gain may be gone before enough passes back and forth take place to build up power from the noise. For many Q-switched solid-state pump lasers, the gain is on for 10-20 ns, so the cavity typically must be 10 cm or less in length. OPO efficiency is reduced as the pulse takes longer to build up since no power is being extracted from the pump light.

For a given crystal and pump-beam orientation, the phase-matching process leads to specific signal and idler wavelengths. As with harmonic generation, CPM and NCPM processes exist, with the latter much more desirable. The ability to tune the generated signal and idler wavelengths by adjusting the angle or temperature of the crystal or the pump wavelength (with tunable pump lasers) is one of the most important features of optical parametric devices.

The first optical parametric devices were demonstrated in the 1960s and looked to be a major breakthrough in coherent sources. However, the limits of materials like LN owing to photorefraction, along with optical damage from the high powers required for operation, prevented widespread use and gave OPOs a reputation for being unreliable.

The development of more robust materials, BBO and LBO, made OPOs more practical, and commercial devices for scientific applications, based on tripled, Q-switched Nd:YAG pump lasers driving BBO-based OPOs, provide nanosecond-duration, tunable light over the 410-2,500 nm range.

The material KTP was initially developed for frequency doubling of Nd:YAG lasers, but examination of its phase-matching characteristics showed that it could be used as a NCPM OPO, also pumped by Nd-doped lasers, to generate a signal wavelength around 1550 nm. The device thus acts to convert a very unsafe (in terms of eye damage) laser for applications like active sensing into a much eye-safer pulsed source. Originally, KTP crystals could only be grown in very small (millimeter) sizes through hydrothermal processes, but development of flux-growth technology allowed scale-up to centimeter-sized materials. At that point, construction of high-energy OPOs was possible, and to date, the highest energy OPO, 450 mJ of signal at 1570 nm, pumped by 1.1 J from a Q-switched Nd:YAG laser,[28] is based on a centimeter-aperture KTP crystal. One drawback with KTP is that the idler wavelength is absorbed in the material, limiting average powers to <10 W due to thermal effects. Motivated by this issue, crystal growers developed flux-growth of an isomorph of KTP, $K(TiO)AsO_4$ (KTA), which has similar phase-matching properties but is highly transparent at the idler wavelength, allowing signal

[28] G.A. Rines, D.M. Rines, and P.F. Moulton, 1994, "Efficient, high-energy, ktp optical parametric oscillators pumped with 1 micron Nd-lasers," in *Advanced Solid State Lasers*, T. Fan and B. Chai, eds., Vol. 20 of OSA Proceedings Series (Optical Society of America), paper PO9.

average-power scaling to 33 W (330 mJ at 100 Hz),[29] which held the record for a time as the highest-average-power OPO. At present, KTP OPOs have replaced Er:glass lasers for eye-safe rangefinder applications, owing to their higher performance and higher pulserates.

As noted, semiconductor nonlinear crystals provide transparency much further into the IR than the oxide materials listed above and are suited for optical parametric generation at long wavelengths, particularly beyond 4 μm. With a few exceptions (CdSe being one) the III-V and II-VI materials lacking inversion symmetry also lack birefringence, and one has to look at more complex compounds, such as the ternary (three-element) chalcopyrite structure (I-III-VI_2) materials, to obtain birefringent phase-matching crystals. The semiconductor materials generally have much higher nonlinear coefficients, but at the expense of smaller bandgaps than oxide materials. Some materials are completely opaque to the common Nd-doped pump lasers, while others suffer from multiphoton absorption with those pumps. In general, the chalcopyrites, because of their structure, have complex thermal expansion properties and are a challenge to grow.

Because of the strong interest in the development of tunable IR sources for counter-sensor applications, there has been considerable investment in the development of chalcopyrite materials, and at present the most widely used material is $ZnGeP_2$ (ZGP), which combines very high nonlinearity (figure of merit = 200) with good thermal conductivity. It is not transparent to Nd-doped lasers but can be pumped by 2,000 nm lasers (as well as OPOs) and can be phase-matched to generate a wide variety of wavelengths from 2.5 to 10 μm, although the material has some absorption at wavelengths beyond 8 μm. Recent reported ZGP-based OPOs pumped by a Q-switched, 35-kHz pulse rate, Ho:YAG hybrid laser have generated average powers of 27 W in the 4 μm wavelength region.[30]

For longer wavelength OPO operation, covering the LWIR 8-12 μm atmospheric window, there are other semiconductor materials with transparency extending beyond the range of ZGP, including CdSe and GaSe, and often the systems to generate these wavelengths employ "tandem OPO" schemes whereby a solid-state laser pumps one OPO and the signal or idler wavelength is then used to pump the OPO covering the LWIR range.

The development of QPM materials has probably had the greatest impact on recent development of OPOs, OPAs, and OPGs. For periodically poled LN (PPLN), the ability to obtain NCPM-like phase matching over a wide range of wavelengths by using different poling periods, along with the relatively high nonlinearity, has made it possible to operate Nd-laser-pumped OPOs with very low thresholds, low enough to operate CW, rather than in pulsed mode. For the 1000-nm and longer pump wavelengths, there appears to be minimal if any degradation due to photorefractive effects, especially if the material is held above room temperature. PPLN based optical parametric devices provide efficient wavelength generation out to wavelengths around 4 μm, beyond which the absorption in LN becomes a limiting issue. Variants on PPLN, such as periodically poled KTP (PPKTP) and stoichiometric $LiTaO_3$ (PPSLT) provide some variation in OPO performance but no extension to longer wavelengths.

The more recent development of GaAs-based QPM structures, called orientation-patterned GaAs (OP-GaAs) provides a material equivalent to ZGP in terms of a very high figure of merit (also around 200), longer-wavelength transparency to 17 μm, and all the advantages of QPM. As with ZGP, the material cannot be pumped with Nd-doped lasers, and 2000-nm pumps provide the best option. OP-GaAs-based OPOs are now under active investigation, and have shown thresholds low enough for CW operation,[31] with 5.3 W of combined signal (3.8 μm) and idler (4.7μm) power when pumped by a 24.7 W

[29] M.S. Webb, P.F. Moulton, J.J. Kasinski, R.L. Burnham, G. Loiacono, and R. Stolzenberger, 1998, "High-average-power $KTiOAsO_4$ optical parametric oscillator," *Opt. Lett.* 23: 1161.

[30] A. Hemming, J. Richards, A. Davidson, N. Carmody, S. Bennetts, N. Simakov, and J. Haub, 2013, "99 W mid-IR operation of a ZGP OPO at 25 percent duty cycle," *Opt. Express* 21: 10062.

[31] L.A. Pomeranz, P.G. Schunemann, S.D. Setzler, C. Jones, and P.A. Budni, 2012, "Continuous-wave optical parametric oscillator based on orientation patterned gallium arsenide (OP-GaAs)," in Conference on Lasers and Electro-Optics 2012, OSA Technical Digest (Optical Society of America), paper JTh1I.4.

Ho:YAG laser. In addition, the materials can be readily pumped by the relatively low peak powers available from pulsed fiber lasers.[32] Given the limited aperture size (1-2 mm height) to date of the material compared to ZGP, it remains to be seen if OP-GaAs can provide the high-energy pulses needed for some applications.

Supercontinuum Generation

The development of high-peak-power, picosecond-duration pulses from mode-locked solid-state lasers in the late 1960s allowed new nonlinear optical effects to be observed. When tightly focused into transparent solids, the green output from a frequency-doubled, mode-locked Nd:glass laser was observed to generate white-light—or "supercontinuum (SC)"—output in an essentially continuous spectra from 400 to 700 nm.[33] Because the pulses were destructive in nature, the source was of little practical use. The physics behind white-light generation were explained as a combination of the nonlinear processes, self-phase modulation, and four-wave mixing.

Further developments examined SC generation in optical fibers, where significant nonlinearities can be readily generated over long distances with modest powers. As the understanding of the process in fibers improved, it became evident that the formation of optical solitons in fibers was key to the efficient, fiber-based generation of broadband SC.

A soliton is a pulse that balances the temporal stretching effect of self-phase modulation from the nonlinear refractive index of the material with the compressing effect of dispersion. The product of the soliton pulsewidth and energy is a constant, with typical pulsewidths in the 0.1-10 ps region. Solitons only appear (assuming the nonlinear index is positive) for soliton central wavelengths in the "anomalous dispersion" region for the fiber, where the group velocity for light increases with the frequency of the light, rather than in the more conventional dispersion rgion, where it decreases. For silica-based fibers, anomalous dispersion starts for wavelengths beyond 1,300 nm; the latter value is referred to as the zero-dispersion wavelength (ZDW).

Conventional step-index silica fibers with a ZDW of >1,300 nm are not well-matched to high-power sources such as Q-switched or mode-locked Nd-doped lasers around 1,060 nm and mode-locked Ti:sapphire lasers at 800 nm. The development of photonic-structure silica fibers allowed reduction of the ZDW to match the more powerful short-wavelength sources and was applied to continuum generation with Ti:sapphire mode-locked lasers[34] to allow control of the carrier-offset frequency of the comb, paving the way for precision frequency standards[35] and the nanometer-level ranging accuracies described in Chapter 3. Commercial photonic fibers with ZDWs in the 1,040 nm region allow SC generation with low-cost microchip, 0.5 ns duration, Q-switched Nd:YAG lasers.

With silica fibers the long-wavelength end of the SC is limited by the IR absorption in the material, and extension of SC generation to longer wavelengths requires different fiber glasses. Glasses with high mid- and long-wavelength IR tend to have longer ZDW wavelengths, but, as with silica, this can be shortened through the use of photonic structures or fiber tapers down to micron diameters. Work in this area is active and likely to produce significant SC powers in the IR in the next decade. One

[32] C. Kieleck, M. Eichhorn, A. Hirth, D. Faye, E. Lallier, and S.D. Jackson, 2009, "OP-GaAs OPO pumped by a Q-switched Tm, Ho:silica fiber laser," in Conference on Lasers and Electro-Optics/International Quantum Electronics Conference, OSA Technical Digest (CD) (Optical Society of America), paper CWJ2.

[33] R.R. Alfano and S.L. Shapiro, 1970, "Observation of self-phase modulation and small-scale filaments in crystals and glasses," *Phys. Rev. Lett.* 24: 592.

[34] J.K. Ranka, R.S. Windeler, and A.J. Stentz, 2000, "Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm," *Opt. Lett.* 25: 25.

[35] D.J. Jones, S.A. Diddams, J.K. Ranka, A. Stentz, R.S. Windeler, J.L. Hall, and S.T. Cundiff, 2000, "Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis," *Science* 288: 635.

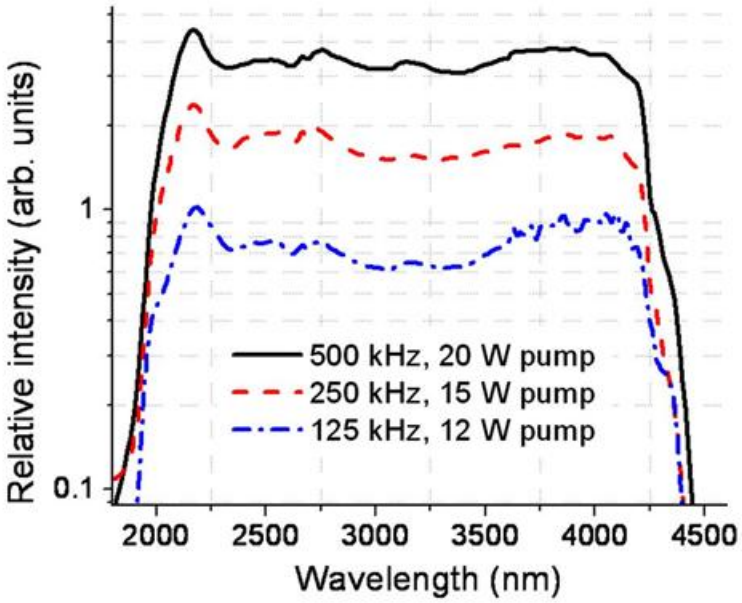

FIGURE 4-16 Intensity as a function of wavelength for SC source based on a zirconium barium lanthanum aluminum sodium fluoride (ZBLAN) fiber driven by a Tm:fiber-based source, for different pump pulse formats. SOURCE: O.P. Kulkarni, V.V. Alexander, M. Kumar, M. J. Freeman, M.N. Islam, F.L. Terry, M. Neelakandan, and A. Chan, 2011, "Supercontinuum generation from ~1.9 to 4.5 μm in ZBLAN fiber with high average power generation beyond 3:8 μm using a thulium-doped fiber amplifier," *J. Opt. Soc. Am.* B28: 2486.

example of a mid-IR SC source employs the glass material ZBLAN, which has a predicted ZDW of 1,650 nm, driven by the pulsed output of a Tm:fiber laser around 2,000 nm.[36] The generated spectra appear in Figure 4-16, showing generation out to 4,500 nm.

Regardless of the complexity of the process, SC generation in fibers leads to a remarkable result, essentially the generation of extremely broadband light in a diffraction-limited beam, with a power per nanometer greater by many orders of magnitude than other broadband sources, such as the blackbody-like radiation from incandescent lamps. Efforts to scale up the average power are underway, and silica-based sources have generated average powers as high as 39 W total power in the 400-2,250 nm region.[37] At least two companies, Fianium (Southampton, England) and NKT Photonics (Birkerød, Denmark) offer commercial, silica-based SC sources.

Conclusion 4-1: Laser technology is bifurcating into: (1) high-cost Q-switched, solid-state lasers and (2) quasi-continuous-wave fiber lasers, pumped by semiconductor diode lasers. The latter are becoming readily available and will further reduce costs for active EO sensing.

[36] Ibid.

[37] K.K. Chen, S. Alam, J.H.V. Price, J.R. Hayes, D. Lin, A. Malinowski, C. Codemard, D. Ghosh, M. Pal, S.K. Bhadra, and D.J. Richardson, 2010, "Picosecond fiber MOPA pumped supercontinuum source with 39 W output power," *Opt. Express* 18: 5426.

Conclusion 4-2: Solid-state lasers seem favorable for active EO due to their combination of output format, operating wavelength, relatively high efficiency, ruggedness, compact size, and reliability. They can be considered coherency converters of their laser diode pump lasers.

Conclusion 4-3: Diode lasers are by far the most widely used commercial lasers and are projected to lead to lower costs in source technology for active EO sensors; they are used as both stand-alone devices for low-cost, short-range sensors and as optical pumps for solid-state lasers used for long-range sensors

Conclusion 4-4: Mode-locked fiber lasers can generate femtosecond pulses with modest energies and a variety of techniques combining bulk solid-state lasers, and nonlinear optics can be used to provide broadband amplification of stretched pulses at high energies. Femtosecond lasers are useful for advanced active sensing techniques such as interactions of matter or formation of filaments.

Conclusion 4-5: Cascade lasers have emerged as an important source of mid- and long-wave infrared emission. Cascade lasers can operate without the need for cryogenic cooling, which is an advantage over other lasers operating at similar wavelengths.

DETECTORS/RECEIVERS

The detector, integrated with appropriate electronics, is the critical photon sensing device in an active receiver. In addition to needing high sensitivity for direct detection ladar, the detectors need to be able to detect short laser pulses without large attenuation. A wide variety of detectors are now available for active receivers. They range from single-element detectors for laser rangefinders and scanning ladar to staring focal plane arrays with 320 × 256 pixels for high spatial and range resolution 3-D imaging flash ladar systems and even larger framing arrays that can be used with spatial heterodyne or polarization-based imaging.[38] The spectral range of these detectors stretches from the visible to the longwave IR (10 μm), which makes them sensitive at the commonly available laser wavelengths and applicable for the various modalities of ladar operation.

The most commonly employed single-element detectors are Si and InGaAs PIN photodiodes that operate at the widely available laser wavelengths of 0.905 μm, 1.06 μm, and ~1.5 μm. For moderate- to high-sensitivity applications, semiconductor avalanche photodiode detectors (APDs) are most useful because they provide internal gain on the detectors.[39-47] According to Beck et al:

[38] P.F. McManamon, 2012, "Review of ladar: A historic, yet emerging, sensor technology with rich phenomenology," *Optical Engineering* 51: 060901.

[39] R.J. McIntyre, "Multiplication noise in uniform avalanche photodiodes," 1966, *IEEE Trans. Electron Devices*, ED-13: 164; R.J. McIntyre, 1999, "A new look at impact ionization – Part I: A theory of gain, noise, breakdown probability and frequency response," *IEEE Trans Electron Devices* 46: 1623.

[40] P. Bhattacharya, 1997, *Semiconductor Optoelectronic Devices*, Prentice Hall, Upper Saddle River, N.J.

[41] X. Sun, 1992, "Photon counting with silicon avalanche photodiodes," *J. Lightwave Technol.* 10: 1023.

[42] K.A. McIntosh, J.P. Donnelly, D.C. Oakley, A. Napoleone, S.D. Calawa, L.J. Mahoney, K.M. Molvar, E.K. Duerr, S.H. Groves, and D.C. Shaver, 2002, "InGaAs/InP avalanche photodiodes for photon counting at 1.06," *Appl. Phys. Lett.* 81: 2505.

[43] J. Beck, C. Wan, M. Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, and J. Campbell, 2004, "The HgCdTe electron avalanche photodiode," *Proc. SPIE* 5564: 44.

[44] M. Jack, J. Asbrock, S. Bailey, et al., 2007, "MBE based HgCdTe APDs and 3-D LADAR sensors," *Proc. SPIE* 6542, A5421.

[45] J.C. Campbell, et al., 2004, *IEEE J. of Selected Topics in Quantum Electronics*, 10: 777.

[46] A.S. Huntington, M.A. Compton, and G.M. Williams, 2007, "Linear-mode single-photon APD detectors," *Proc. SPIE* 6771: Q7710.

> "The APD is the solid-state equivalent of the photomultiplier tube in that both create many electrons from each incident photon. The number of electrons created per absorbed photon is the internal detector gain. In the typical APD, the noise associated with the gain process is called excess noise. The result is that noise increases with gain faster than the signal, and the signal to noise ratio in the detector is actually degraded. For this reason APDs are used in applications where the noise is limited not by the detector noise, but by a fixed system noise. Typically these are high-bandwidth systems that must detect very short laser pulses.
>
> The gain in the detector directly improves the signal-to-noise ratio (SNR) in situations where system noise limits the performance. The improvement continues with increasing detector gain until the noise of the detector becomes equal to the fixed system noise, at which point the SNR begins to degrade. In APDs, high-bias dark current and excess noise limit the useful gain and hence the SNR of the detection system. Considerable effort is therefore expended in reducing the excess noise in APDs, because being able to operate at twice the gain, for example, enables the selection of a laser that provides half the power output while maintaining the same SNR. In many cases, it is the laser power requirement that makes or breaks a system concept or application"[48].

The band structure of the semiconductor has the most fundamental impact on the excess noise factor.[49,50] The following paragragh is based upon text from Beck et al "The gain process in the APD is due to the ionization and multiplication of carriers in a high-field region created in the device with applied bias voltage. The main reason many APDs are noisy is because both the electron and the hole tend to ionize, making the multiplication process less predictable and leading to the unstable situation of infinite gain. In addition, optical phonon interactions with electrons and holes in the multiplication region remove kinetic energy from the carriers and further randomize the multiplication process. The ratio between the relative ionization rate of holes and that of electrons is the ionization coefficient ratio, k. If only one carrier ionizes ($k = 0$ or ∞), the process is far less noisy. If both carriers ionize and gain is turned up with increasing bias, a point is reached where the regenerative nature of the two-carrier gain process results in infinite gain and zero bandwidth. This infinite gain point is called avalanche breakdown. The $k = 0$ device on the other hand does not exhibit avalanche breakdown. Reduction in excess noise is achieved by optimizing the device geometry"[51].

APDs are most commonly used for communication and active sensing applications. With the rise of fiber optics-based telecommunications, high bandwidth near infrared (NIR) InGaAs PIN and APD detectors are used to detect the 1.55 μm laser pulses that carry the signal information.

As described in Chapter 2, 3-D imaging systems have traditionally operated with either a single detector element, or a linear array of detectors, with scanned laser systems. 3-D flash imaging can be performed with 2-D arrays of APD detectors or focal plane arrays (FPAs). According to Beck et al "The FPAs require special integrated circuit readouts that provide circuitry for each pixel; they are used to detect the laser return pulse for signal peak amplitude and time delay (range). For 2-D arrays, this circuitry needs to be located underneath each pixel. Finally, because the laser needs to illuminate a large area", laser power is critical[52]. For long operating ranges, the ultimate sensitivity of a detector is thus highly desirable.

---

[47] J. Beck, C. Wan, M. Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, and J. Campbell, 2006, "The HgCdTe electron avalanche photodiode," *IEEE LEOS Newsletter*, 8 (October).

[48] Ibid

[49] M.A. Kinch, 2007, *Fundamentals of Infrared Detector Materials*, SPIE Press, Bellingham, WA.

[50] M.A. Kinch, 2008, "A theoretical model for the HgCdTe avalanche photodiode," *J. Electron. Mater.* 37: 1453.

[51] J. Beck, C. Wan, M. Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, and J. Campbell, 2006, "The HgCdTe electron avalanche photodiode," *IEEE LEOS Newsletter*, 8 (October).

[52] Ibid

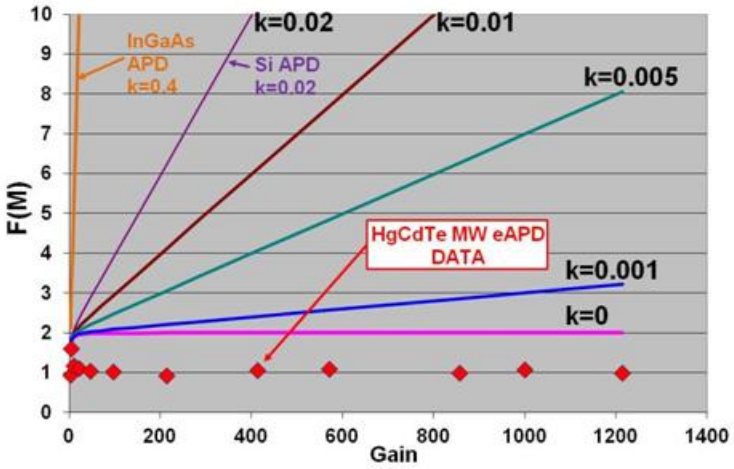

FIGURE 4-17 Excess noise factor plot for commonly used semiconductors for APDs. SOURCE: J. Beck, C. Wan, M. Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, and J. Campbell, 2004, "The HgCdTeelectron avalanche photodiode," *Proc. SPIE* 5564: 44.

For operation with eye-safe laser wavelengths of around 1.5 µm, InGaAs APDs with spectral sensitivity out to 1.7 µm are an attractive choice. They readily meet the bandwidth requirements needed for operation with pulses 1-10 ns wide. However, InGaAs APDs have a relatively high excess noise factor due to a hole to electron ionization coefficient ratio, *k*, of about 0.4, meaning that both holes and electrons will ionize significantly. For InAlAs, $k = 0.2$. Si, by comparison, has an ionization ratio of 0.02 and therefore much less excess noise, however it is not sensitive to wavelengths greater than 1.0 µm.

Figure 4-17 shows a plot of excess noise factor $F(M)$, calculated using McIntyre's equation,[53] as a function of detector gain for various *k* values and for the common APD semiconductors. The $k = 0$ line with $F(M) = 2$ is based on McIntyre's original theory for thick multiplication regions.[54] HgCdTeelectron-initiated APD (e-APD) devices exhibit $F(M) \sim 1$,[55] as shown for the measured data in Figure 4-17. Current theoretical modeling based on HgCdTe band structure explains the measured data.[56]

The excess carrier noise factor imposes a separate limit on the usable gain of an APD. In the limit of high avalanche gain, the sensitivity of an ideal noiseless APD is limited by the shot noise on the optical signal itself. However, most $k \neq 0$ APDs generate multiplication noise in excess of the shot noise already present on the optical signal; this excess multiplication noise intensifies with increasing avalanche gain, such that for any given level of downstream amplifier noise, there is a limit to how much avalanche gain is useful. Increasing the avalanche gain beyond the optimal value increases the shot noise faster than the amplified signal photocurrent, degrading the signal-to-noise ratio.[57]

[53] R.J. McIntyre, 1966, "Multiplication noise in uniform avalanche photodiodes," *IEEE Trans. Electron Devices*, ED-13:164.; R.J. McIntyre, 1999, "A new look at impact ionization – Part I: A theory of gain, noise, breakdown probability and frequency response," *IEEE Trans Electron Devices* 46: 1623.

[54] Ibid.

[55] J. Beck, C. Wan, M. Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, J. Campbell, 2004, "The HgCdTe electron avalanche photodiode," *Proc. SPIE*, 5564: 44.

[56] M.A. Kinch, 2008, "A theoretical model for the HgCdTe avalanche photodiode," *J. Electron. Mater.* 37: 1453.

[57] G.M. Williams, D.A. Ramirez, M.M. Hayat, and A.S. Huntington, 2013, "Time resolved gain and excess noise properties of InGaAs/InAlAs avalanche photodiodes with cascaded discrete gain layer multiplication regions," *J. Appl. Phys.* 113: 093705.

The tunable bandgap semiconductor $Hg_{1-x}Cd_xTe$ is the only commonly used semiconductor that exhibits $k = 0$ behavior and therefore the electron initiated avalanche multiplication exhibits close to ideal APD characteristic of excess noise factor $F(M) = 1$. Electron-initiated APD operation in $Hg_{1-x}Cd_xTe$ has been demonstrated for a broad range of compositions from x = 0.7 to 0.21, corresponding to cutoff wavelengths from 1.3 to 11 μm.[58] Thus, HgCdTe at gain =100 provides 10 to 20 times less noise than InGaAs or InAlAs APDs and 4 times less noise than Si APDs.

In addition to APD operation in the linear mode, beyond the breakdown voltage Geiger-mode operation is realized and forms the basis for Geiger-mode APDs (GM-APDs), a technology widely used today for active sensing. GM-APDs are commonly based on Si, InGaAs and InGaAsP.[59,60] They are being used as photon counting detectors in ladar systems and have been highly successful in 3-D imaging applications at 1.06 and 1.55 μm. Both Geiger and linear-mode APDs are discussed in greater detail below.

Superconducting niobium nanowire based single photon counting detectors have been demonstrated recently.[61] This detector technology has been successfully used for some active sensing applications at 1.56 μm.[62]

In the following sections the various detector technologies in use are discussed as well as those that are expected to be developed in the next 5-10 years.

Single Element, Small Format Arrays, and Large Format Arrays

The choice of a detector for active EO sensing is strongly dependent on the kind application. Imaging 2-D lidar applications are discussed in Chapters 2 and 3.

Single-element Si PIN diodes and APDs suitable for the spectral range from 0.255 μm to 1.1 μm are commercially available and used for ranging applications such as with a pulsed 1.06 μm laser.[63] The active area of a PIN ranges from diameters of 230 μm to 3.0 mm, and APDs have larger diameters, 5-16 mm. The Si APDs have a quantum efficiency (QE) of ≤40 percent at 1.06 μm and can be operated either in normal linear mode with an internal gain of 250 or more, or in Geiger-mode.[64] These low-cost detectors are robust, immune to magnetic fields, and can be operated at or near room temperature.

[58] J. Beck, C. Wan, M. Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, and J. Campbell, 2006, "The HgCdTe electron avalanche photodiode," *J. Electron. Mater.* 35: 1166.

[59] B.F. Aull, A.H. Loomis, D.J. Young, R.M. Heinrichs, B.J. Felton, P.J. Daniels, and D.J. Landers, 2002, "Geiger-mode avalanche photodiodes for three-dimensional imaging," *MIT Lincoln Laboratory Journal* 13: 335.

[60] S. Verghese, K.A. McIntosh, Z.L. Liau, C. Sataline, J.D. Shelton, J.P. Donnelly, J.E. Funk, R.D. Younger, L.J. Mahoney, G.M. Smith, J.M. Mahan, D.C. Chapman, D.C. Oakley, and M. Brattain, 2009, "Arrays of 128 × 32 InP-based Geiger-mode avalanche photodiodes," *Proc. SPIE* 7320: 73200M.

[61] R.H. Hadfield, 2009, "Single photon detectors for optical quantum information applications," *Nature Photonics* 3(12): 696.

[62] A. McCarthy, N.J. Krichel, N.R. Gemmell, X. Ren, M.G. Tanner, S.N. Dorenbos, V. Zwiller, R.H. Hadfield, and G.S. Buller, 2013, "Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection," *Optics Express* 21(7): 8904.

[63] See http://voxtel-inc.com/products/.

[64] M.A. Krainak, X. Sun, G. Yang, L.R. Mikoc, and J.B. Abshire, 2009, "Photon detectors with large dynamic range and at near infrared wavelength for direct detection space lidars," *Proc. of SPIE* 7320: 732005.

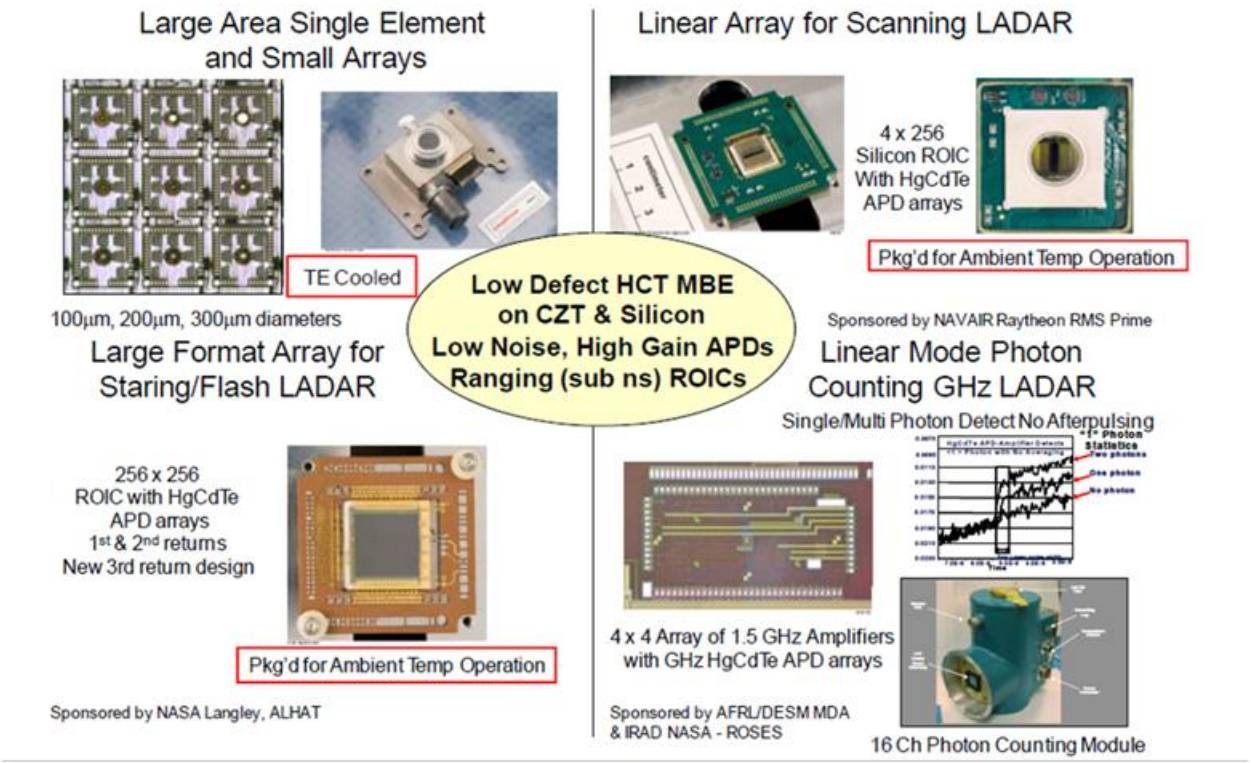

FIGURE 4-18 Family of HgCdTe APD detectors in a variety of formats. SOURCE: M. Jack, G. Chapman, J. Edwards, W. McKeag, T. Veeder, J. Wehner, and F. Amzajerdian, 2012, "Advances in LADAR Components and Subsystems at Raytheon," *Proc. of SPIE* 8353: 83532F-1.

A variety of pixel format detector arrays have been developed for a wide range of applications. Figure 4-18 shows detectors in multiple formats from single pixels to 256 × 256 arrays for various ladar applications. No single format is suitable for all classes of applications. For example, for gas-sensing lidar applications, small detector formats are well suited, but they need to be sensitive in the wavelength region for the specific gases.[65] For 3-D topographic mapping, larger formats are preferred to maximize ground surface area per unit time and to reduce reliance on mechanical scanners.

There are two main types of detectors that are sensitive to single photons, Geiger-mode (GM) and linear mode (LM). Both rely on avalanche multiplication of the carrier generated by the incident photon so that noise in ROIC can be overcome. Avalanche of carriers occurs under high electric field when accelerated carriers impact the lattice to create more carriers, which are subsequently accelerated and have the opportunity to impact ionize. Both electrons and holes can ionize at rates that are dependent on the fundamental properties of the semiconductor material. The two modes are different in the magnitude of the avalanche multiplication that is created.

[65] J. Beck, J. McCurdy, M. Skokan, C. Kamilar, R. Scritchfield, T. Welch, P. Mitra, X. Sun, J. Abshire, and K. Reiff, 2013, "A highly sensitive multi-element HgCdTe e-APD detector for IPDA lidar applications," *Proc. of SPIE* 8739: 87390V.

Geiger-Mode APDs

Geiger-mode avalanche photodiode detectors (GM-APDs) are biased at voltages several volts above the breakdown voltage, typically much higher than biases of linear-mode detectors. At this high voltage a metastable state is reached where carriers generated through photon absorption (or dark) entering the depletion region begin avalanche multiplication and cause a sustained chain of impact ionizations, resulting in a measurable current pulse. Materials with high probabilities of electron or hole ionization are usually good for Geiger-mode detectors. The current generation of GM-APDs with single-photon-counting sensitivity were developed at MIT's Lincoln Laboratory (MIT-LL).[66,67] Figure 4-19

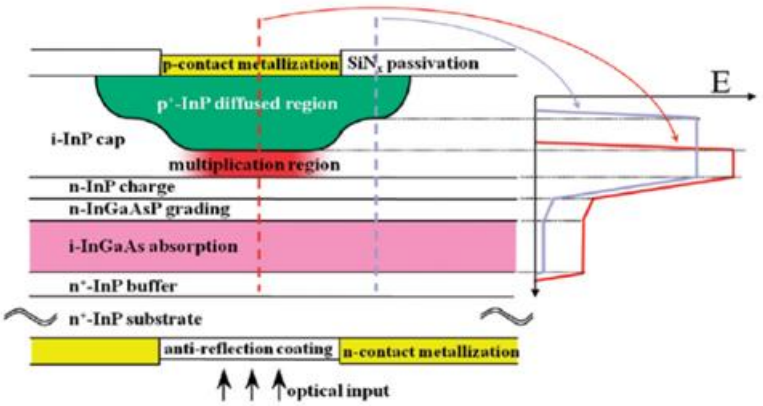

FIGURE 4-19 Geiger-mode detector architecture from Princeton Light wave.
SOURCE: M.A. Itzler, M. Entwistle, M. Owens, X. Jiang, K. Patel, K. Slomkowski, T. Koch, S. Rangwala, P.F. Zalud, Y. Yu, J. Tower, J. Ferraro, 2009, "InP-based Geiger-mode avalanche photodiode arrays for three-dimensional imaging at 1.06 µm," *Proc. of SPIE* 7320, 73200O.

shows the architecture of a GM-APD device based on lattice matched $In_{0.53}Ga_{0.47}As$, with a bandgap of ~0.75 eV (1.65 µm cutoff at 300 K), grown on InP substrate, that is suitable for detection at ~1.5 µm.[68,69] In this backside-illuminated detector architecture the laser photon absorption occurs in the narrower bandgap InGaAs layer, and the resulting photo excited holes are injected into the high-field InP multiplication region[70]. Operating at temperatures ≥200 K the dark current in this device is dominated by the generation-recombination current in the InGaAs absorber layer.

The GM-APD is a noiseless gain detector with nearly infinite gain and a well-designed device is essentially shot noise limited. The detectors show 1 ns or less timing performance. The dark count rates are typically ~10 kHz and photon detection efficiency (PDE) is ~30 percent. These devices have to be

[66] B.F. Aull, A.H. Loomis, D.J. Young, R.M. Heinrichs, B.J. Felton, P.J. Daniels, D.J. Landers, 2002, "Geiger-mode avalanche photodiodes for three-dimensional imaging," *MIT Lincoln Laboratory Journal* 13: 335.

[67] S. Verghese, K.A. McIntosh, Z.L. Liau, C. Sataline, J.D. Shelton, J.P. Donnelly, J.E. Funk, R.D. Younger, L.J. Mahoney, G.M. Smith, J.M. Mahan, D.C. Chapman, D.C. Oakley, and M. Brattain, 2009, "Arrays of 128x32 InP-based Geiger-mode avalanche photodiodes," *Proc. SPIE* 7320: 73200M.

[68] M.A. Itzler, M. Entwistle, M. Owens, X. Jiang, K. Patel, K. Slomkowski, T. Koch, S. Rangwala, P.F. Zalud, Y. Yu, J. Tower, and J. Ferraro, 2009, "InP-based Geiger-mode avalanche photodiode arrays for three-dimensional imaging at 1.06 µm," *Proc. of SPIE* 7320: 73200O.

[69] M. Entwistle, M.A. Itzler, J. Chen, M. Owens, K. Patel, X. Jiang, K. Slomkowski, and S. Rangwala, 2012, "Geiger-mode APD camera system for single photon 3-D LADAR imaging," *Proc. of SPIE* 8375: 83750D.

[70] M.A. Itzler, M. Entwistle, M. Owens, X. Jiang, K. Patel, K. Slomkowski, T. Koch, S. Rangwala, P.F. Zalud, Y. Yu, J. Tower, and J. Ferraro, 2010 "Design and performance of single photon APD focal plane arrays for 3-D LADAR imaging". Proc. SPIE 7780:77801M, Detectors and Imaging Devices: Infrared, Focal Plane, Single Photon.

reset after photon detection because of a holdoff time, also called the dead time, during which the detectors are blinded. This issue can be mitigated by spreading the signal over multiple pixels to reduce the probability of a photon striking a blinded pixel.

A GM-APD FPA is illustrated in Figure 4-20. They are fabricated by In bump bonding to ROICs designed to enable 3-D imaging. A lenslet array is mounted on the detector array to focus the incident light onto an array of APDs and thereby improve the detector fill factor. They also increase the signal when small junctions are used to reduce leakage current. Geiger-mode operation of the APD involves momentarily biasing the diode voltage above its avalanche breakdown voltage[71]. In this metastable state, a single photon can trigger a signal large enough to trigger complementary metal oxide semiconductor (CMOS) circuitry in the ROIC unit cell. Digitization of the signal is performed in the unit cell. The unit cells also accurately clock the time of arrival for each pixel and then disarms the APD. After the CMOS circuit disarms the APD, it must wait for a reset time of 1-2 μs until the trapped holes are thermally ionized and can escape. Only at that point is the device ready to be re-armed, without it immediately after-pulsing, and to receive the next photons.

In the GM-APD, regardless of the number of photons incident at a given time, the signal level is the same. When intensity information is required, the histogram of hits over a period of time is measured. For 3-D imaging, multiple range returns on a pixel are handled using multiple pulses. The ROICs can be designed for linear-mode ladar or for continuous sensing for coherent mode operation.

GM-APD devices are commercially available in the United States from Princeton Lightwave and from Boeing Spectrolab in 32 × 32 and 32 × 128 pixel formats. Recently 64 × 256 formats were demonstrated at MIT-LL.[72] Overall, GM-APDs are the more mature detector technology currently in use for 3-D imaging with photon counting sensitivity. They have been used for 3-D imaging applications in a variety of environments but primarily in airborne settings.

The layout for a typical GM-APD camera system consists of FPAs packaged on thermoelectric coolers (TECs) and ceramic carriers with microlenses above the FPA, as shown in Figure 4-21. Frame rates of 186 kHz have been reported and are scaleable up to 300 kHz with appropriate range gates. Range gates are selectable in increments of 2, 4, 8 and 10 μs. Timing is captured digitally with a phase-locked-loop master clock and 13 bits of resolution.

Linear/Proportionate-Mode APDs

Linear-mode APD devices have been most commonly fabricated from Si, InGaAs and HgCdTe. Linear Si APD technology is mature and widely available. Sensitivities to single-photon counting have been demonstrated below the breakdown voltage of ~ 250 V.[73]

[71]G. M. Williams, "GHz-rate Single Photon Sensitive Linear Mode APD Receivers" Available at http://voxtel-inc.com/files/2012/07/GHz-rate-Single-Photon-Sensitive-Linear-Mode-APD-Receivers-SPIE-Winter-2009-.pdf.

[72] S. Verghese et al., 2007, "Arrays of InP-based avalanche photodiodes for photon counting," *IEE Journal of Selected Topics in Quantum Electronics* 13(4): 870.

[73] X. Sun, 1992, "Photon counting with silicon avalanche photodiodes," *J. Lightwave Technol*. 10: 1023.

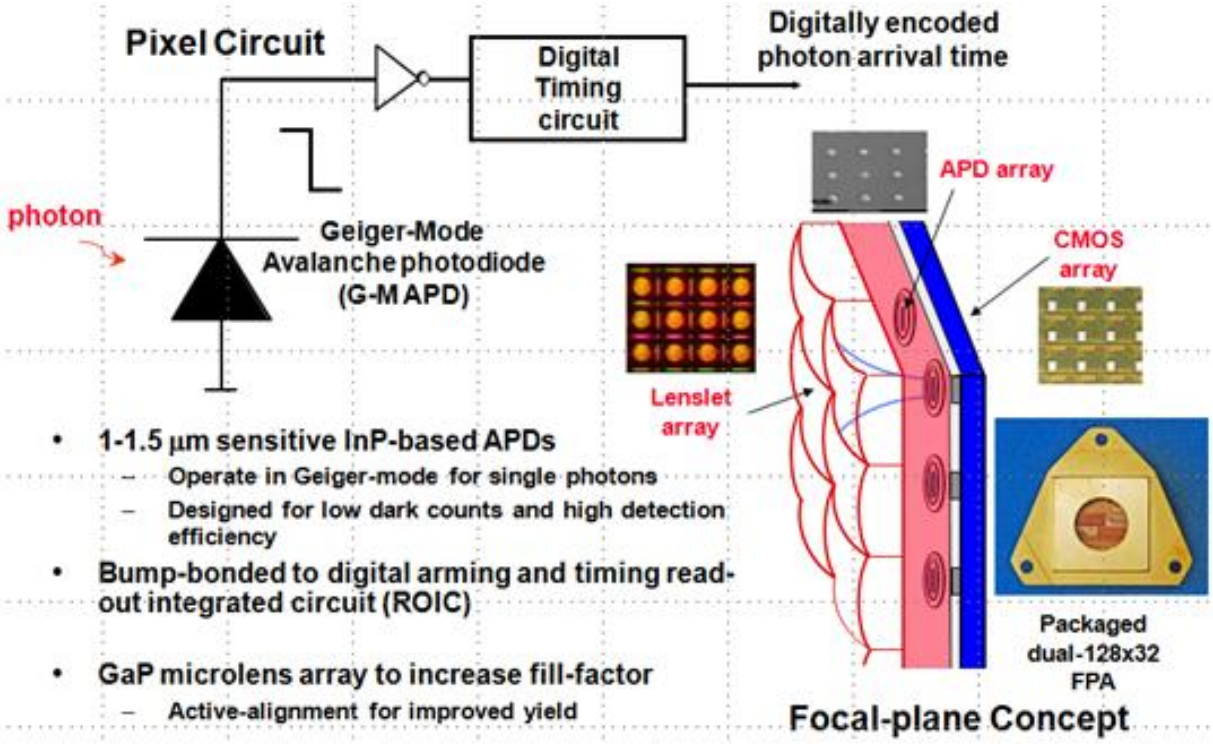

FIGURE 4-20 Geiger-mode sensing concept. SOURCE: Melissa Choi, MIT-LL.

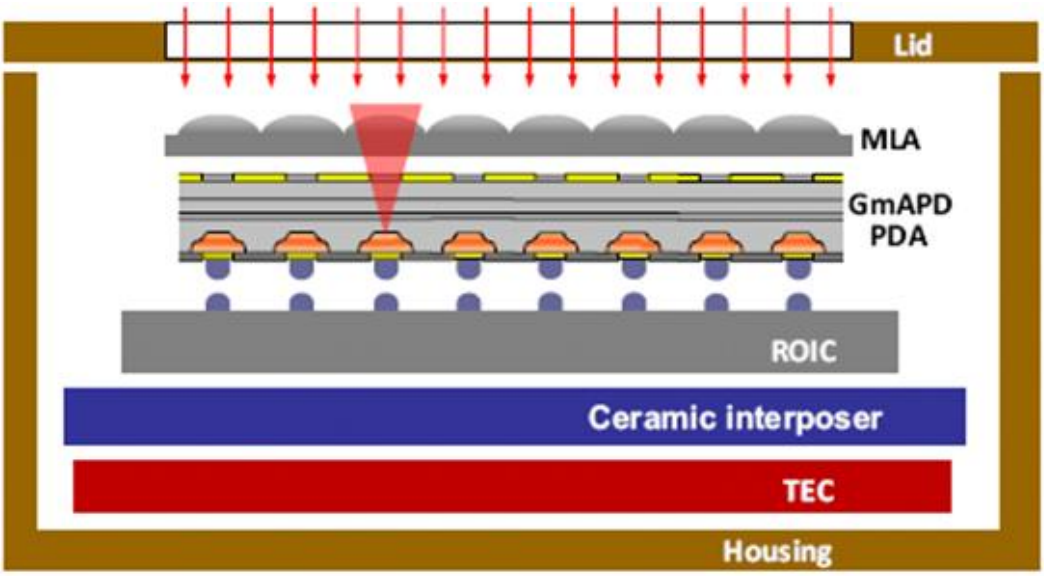

FIGURE 4-21 GM-APD camera packaging. SOURCE: M. Entwistle, M.A. Itzler, J. Chen, M. Owens, K. Patel, X. Jiang, K. Slomkowski, and S. Rangwala, 2012, "Geiger-mode APD Camera System for Single Photon 3-D LADAR Imaging," *Proc. of SPIE* 8375: 83750D.

InGaAs-based APDs

Linear-mode InGaAs APDs are commercially available in a variety of formats ranging from single pixel to 128 × 128 FPAs. As discussed above, due to the high *k* value of 0.4 the usable gain of conventional InGaAs is limited to ~10. To overcome this limitation, a multi-gain-stage linear-mode InGaAs single-carrier multiplication (SCM) APD has been developed at Voxtel, Inc.,[74,75] that allows single photon noise equivalent input (NEI) sensitivity when coupled to a custom ROIC.[76] The key innovation here is a method to increase the gain at which low-noise multiplication can be achieved.

According to Williams et al. "the APD is grown by molecular beam epitaxy on InP substrates from lattice-matched InGaAs and InAlAs alloys. Avalanche multiplication inside the APD occurs in a series of asymmetric gain stages whose layer ordering acts to enhance the rate of electron-initiated impact ionization and to suppress the rate of hole-initiated ionization when operated at low gain. The multiplication stages are cascaded in series, interposed with carrier relaxation layers in which the electric field is low, preventing avalanche feedback between stages. These measures result in lower excess multiplication noise and stable linear-mode operation at higher avalanche gain than are characteristic of APDs fabricated from the same semiconductor alloys in bulk. The noise suppression mechanism is analyzed by simulations of impact ionization spatial distribution and gain statistics. The devices employing this design have been demonstrated to operate at linear-mode gain in excess of 6,000 without avalanche breakdown. The usable APD gain achieved is 10-45. Excess noise characterized by an effective impact ionization rate ratio below 0.04 has been reported at gains over 1,000".[77]

The detector arrays are fabricated as backside illuminated (through the InP substrate), mesa-etched diodes with 7 to 10 gain stages to provide a QE of 85 percent across the 950 to 1,650 nm spectral band. The ROIC operates at a 10 kHz frame rate, enables less than 200 ps time jitter and 3 ns pulse pair resolution by using a low-noise GHz-class preamplifier. The hybridized FPAs, in 32 × 32 and 128 × 128 pixel formats with 36-42 μm pixel pitch, are capable of time stamp and amplitude sample of the first three or the first and last pulse returns. The device operates at 230 K. Figure 4-22 shows an example.

HgCdTe APDs

The $Hg_{1-x}Cd_xTe$ material system is a solid solution of CdTe (Eg=1.6 eV at ~0 K) and HgTe (Eg= -0.3 eV at ~0 K), where the bandgap is tunable between these extremes asxis varied from 0 to 1[78]. Despite the wide tunability in bandgap and therefore the detector cutoff wavelength, the lattice constant changes little with the x value. As a consequence HgCdTe structures with different bandgaps can be grown on the same lattice-matched substrate, which is typically $Cd_{0.96}$ $Zn_{0.04}Te$.

---

[74] G.M. Williams, D.A Ramirez, M.M. Hayat, and A.S. Huntington, 2013, "Time resolved gain and excess noise properties of InGaAs/InAlAs avalanche photodiodes with cascaded discrete gain layer multiplication regions," *J. Appl. Phys.* 113: 093705.

[75] See http://voxtel-inc.com/products/.

[76] A.S. Huntington, M.A. Compton, and G.M. Williams, 2007, "Linear-mode single-photon APD detectors," *Proc. SPIE* 6771: Q7710.

[77] Ibid

[78] J. Beck, C. Wan, M.Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, J. Campbell, 2004, "The HgCdTe Electron Avalanche Photodiode," *Proc. SPIE*, 5564: 44.

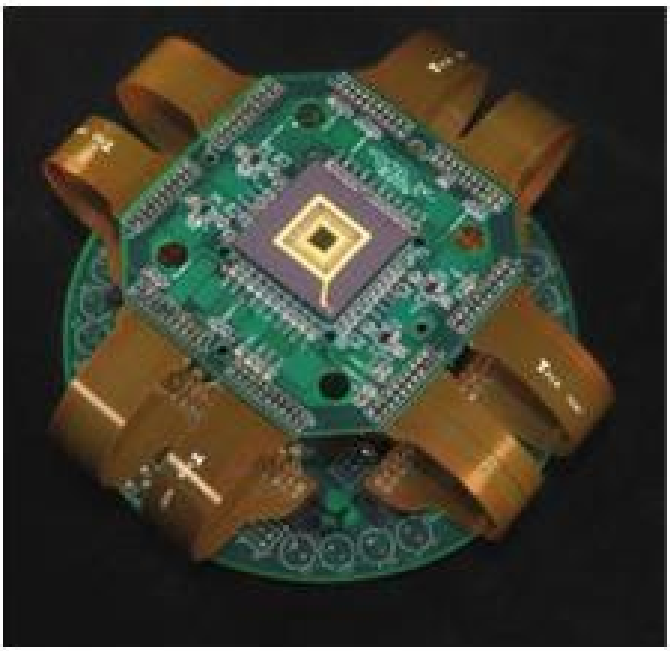

FIGURE 4-22 A 32 × 32 element InGaAs SCM-APD shown mounted with drive electronics. SOURCE: http://voxtel-inc.com/products/.

One of the commonly employed HgCdTe electron-initiated APD architectures, the front-side-illuminated, cylindrical geometry, lateral collection, p-around-n device geometry is illustrated in Figure 4-22. This particular device architecture used at DRS Technologies is called High Density Vertically Integrated Photodiode (HDVIP®).[79,80] The HDVIP structure with HgCdTe grown by liquid phase epitaxy (LPE) is currently employed in MWIR and LWIR staring FPAs in production at DRS. The key features of this architecture are: "(1) interdiffused CdTe passivation of both surfaces for low 1/f noise, (2) thermal cycle reliability that is detector and array size independent, (3) low defects due to diode junction orientation being orthogonal to the direction of threading dislocations, and (4) front-side illumination for high quantum efficiency from the IR into the visible region as shown in Figure 4-23, and good modulation transfer function (MTF)"[81].

For APD operation the reverse bias is increased until an electric field sufficient for avalanche multiplication is achieved. At a few volts bias, the n-doped region becomes fully depleted. Photoelectrons generated in the p-doped region diffuse to the n region, which is the multiplication region. Multiplication occurs as the electrons traverse the high-field depletion region shown Figure 4-24. A unique feature of the HgCdTe e-APD is that holes generated in the multiplication region, for all practical purposes, do not

[79] J. Beck, C. Wan, M.Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, J. Campbell, 2004, "The HgCdTe Electron Avalanche Photodiode," *Proc. SPIE*, 5564: 44.

[80] J. Beck, C. Wan, M. Kinch, J. Robinson, P. Mitra, R. Scritchfield, F. Ma, and J. Campbell, 2006, "The HgCdTe Electron Avalanche Photodiode," *IEEE LEOS Newsletter*, 8 (October).

[81] P. Mitra, J. D. Beck, M. R. Skokan,; J. E. Robinson,; J. Antoszewski,; K. J. Winchester,; A. J. Keating,; T. Nguyen, K. K. M. B. D. Silva, C. A. Musca, J. M. Dell, L. Faraone, 2006, "SWIR hyperspectral detection with integrated HgCdTe detector and tunable MEMS filter" *Proc. SPIE* 6295:62950G, Infrared Detectors and Focal Plane Arrays VIII.

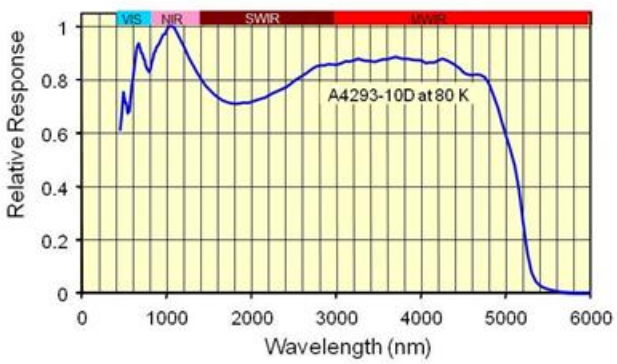

FIGURE 4-23 Broad spectral response from visible to MWIR is characteristic of front-side-illuminated HDVIP e-APD. SOURCE: DRS Technologies, Dallas Texas.

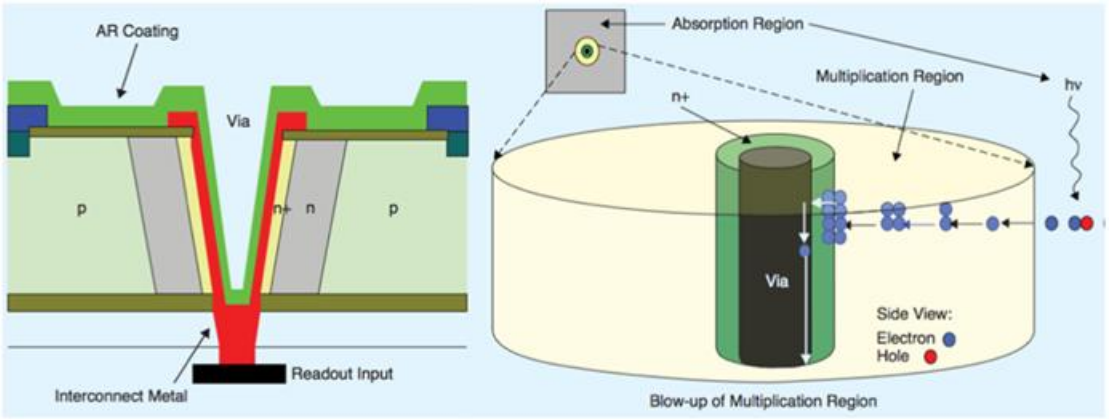

FIGURE 4-24 Schematic of a single-element HgCdTe e-APD based on the lateral collection p-around-n geometry. SOURCE: DRS Technologies, Dallas, Texas.

multiply, which results in an ideal $k = 0$ APD, where k is the hole to electron ionization coefficient ratio. Another unique feature of the HgCdTe e-APD is that the multiplication process is largely deterministic, which results in an excess noise factor near 1.0. The geometry favors high mobility electron injection into the central high-field region, which helps achieve high bandwidth. The geometry and the 3-5 µm thick HgCdTe layer makes the shunt capacitance of each diode extremely small, typically 6 to 12 fF per diode, which reduces the ROIC preamplifier noise and maintains a wide electrical bandwidth.

The noiseless gain behavior of the HgCdTe e-APD is illustrated in Figure 4-25. It is a 4.5 µm cutoff e-APD operating at 85 K with a 1.2-1.8 µm cold filter and illuminated with a 1.55 µm, 10 ns pulsewidth laser with 800 photons incident on the 64 µm pitch detector. The plot shows the effect of the detector gain achieved as a function of bias voltage on the left. At bias voltages of <1 V, where most HgCdTe photodiodes for passive mode are operated, there is no signal corresponding to the incident laser photons. As the bias voltage is raised to 4 V corresponding to an APD gain of ~5 the signal is barely discernible, while at 8 V the gain is ~180 and the SNR is ~45. At 10 V it reaches a gain of ~700 and a SNR of 90 and the detector becomes shot noise limited.

These characteristics of HgCdTe have spawned a large effort in the development of e-APDs. Detector gains of well over 1000 have been demonstrated. Detectors across the 1.7-11 µm spectral band have been fabricated, some in array formats as large as 640 × 480 pixels with 25 µm pitch and integrated in active imaging systems. Linear-mode HgCdTe e-APD devices have been realized in several device

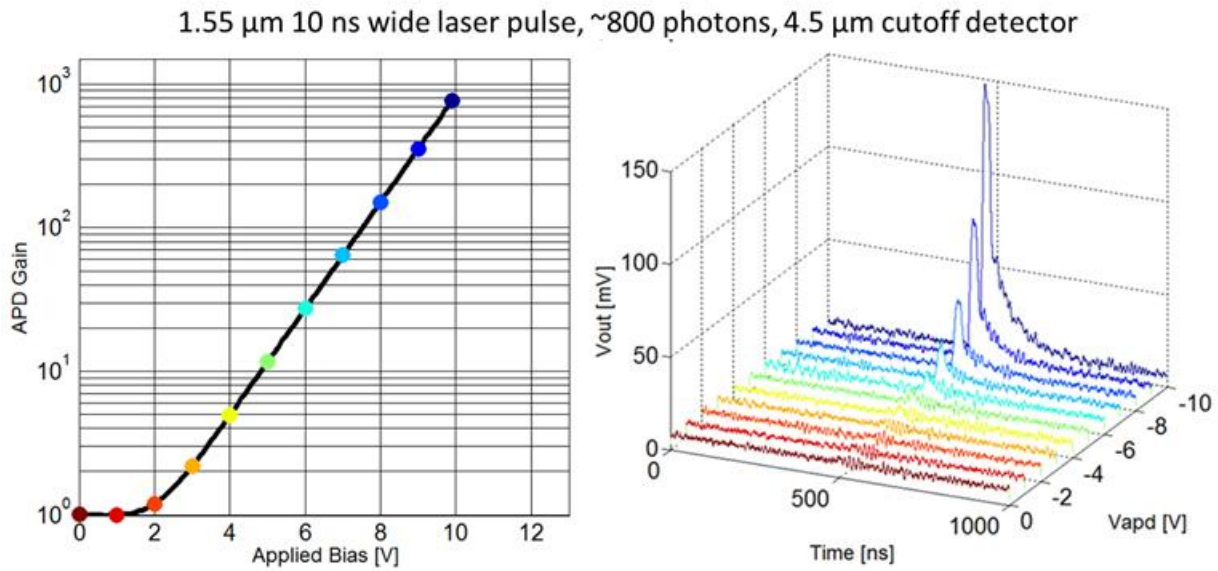

FIGURE 4-25 Effect of noiseless gain. Left: APD gain vs. applied bias on a HgCdTe detector at 80 K incident with 800 photons in 10 ns, 1.55 µm pulses. Right: oscilloscope traces of the signal output taken at various applied bias voltages. The colored data points on the left correlate with the same colored traces on the right. SOURCE: DRS Technologies, Dallas, Texas.

architectures, grown by liquid phase epitaxy (LPE) and molecular beam epitaxy (MBE), in the United States and Europe.[82-87] Detectors and FPAs have been demonstrated for single-photon counting,[88,89,90] active-passive imaging,[82] and 3-D imaging applications.[91]

[82] W. McKeag, T. Veeder, J. Wang, M. Jack, T. Roberts, T. Robinson, J. Neisz, C. Andressen, R. Rinker, T.D. Cook, and F. Amzajerdian, 2011, "New developments in HgCdTe APDs and LADAR receivers," *Proc. of SPIE* 8012: 801230.

[83] J. Beck, M. Woodall, R. Scritchfield, M. Ohlson, L. Wood, P. Mitra, and J. Robinson, 2008, "Gated IR imaging with 128x128 HgCdTe electron avalanche photodiode FPA," *J. Electron Mater.* 37: 1334.

[84] A. Ashcroft and I. Baker, 2010, "Developments in HgCdTe avalanche photodiode technology and applications," *Proc. of SPIE* 7660: 76600C.

[85] J. Rothman, "APD Development at CEA Leti-Minatec, 2009, "ESO Detectors for Astronomy Workshop proc. Available at http://www.eso.org/sci/meetings/dfa2009/program.html.

[86] E. deBorniol, J. Rothman, F. Guellec, G. Vojetta, G. Destéfanis, 2012, "Active three-dimensional and thermal imaging with a 30-µm pitch 320 × 256 HgCdTe avalanche photodiode focal plane array," *Optical Engr.* 51: 061305.

[87] O. Gravrand, G. Destefanis, S. Bisotto, N. Baier, J. Rothman, L. Mollard, D. Brellier, L. Rubaldo, A. Kerlain, and M. Vuillermet, 2013, "Actual issues in HgCdTe research and expected progress in IR detector fabrication," *J. Electron. Mater.* 42(11):3349.

[88] J.D. Beck, R. Scritchfield, P. Mitra, W. Sullivan III, A.D. Gleckler, R. Strittmatter, and R.J. Martin, 2011, "Linear mode photon counting with the noiseless gain HgCdTe e-APD," *Proc. of SPIE* 8033: 80330N.

[89] M.L. Bryan, G. Chapman, D.N.B. Hall, M.D. Jack, S.M. Jacobson, and J. Wehner, 2012, "Investigation of linear-mode, photon-counting HgCdTe APDs for astronomical observations," *Proc. of SPIE* 8453: 84532F.

[90] G. Vojettaa, F. Guelleca, L. Mathieua, K. Fouberta, P. Feautrierb, and J. Rothman, 2012, "Linear photon-counting with HgCdTe APDs," *Proc. of SPIE* 8375: 83750Y.

[91] E. deBorniol, J. Rothman, F. Guellec, G. Vojetta, and G. Destéfanis, 2012, "Active three-dimensional and thermal imaging with a 30-µm pitch 320 × 256 HgCdTe avalanche photodiode focal plane array," *Optical Engineering* 51: 061305.

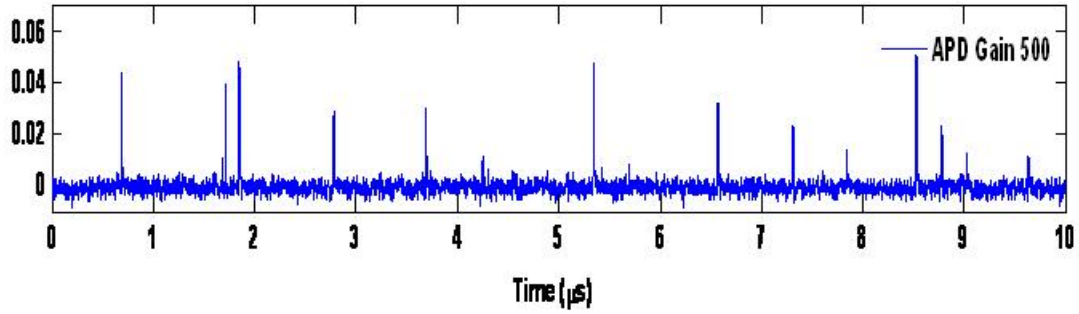

FIGURE 4-26 Analog output showing single photon initiated events captured with a 4,330 nm cutoff HgCdTe detector at 84 K. SOURCE: J.D. Beck, R. Scritchfield, P. Mitra, W. Sullivan III, A.D. Gleckler, R. Strittmatter and R.J. Martin, 2011, "Linear mode photon counting with the noiseless gain HgCdTe e-APD," *Proc. of SPIE* 8033: 80330N.

The linear gain and low excess noise not only allow detecting single photon events, they also enable proportional photon counting with a greatly enhanced dynamic range. Gain above 500 is readily achieved, which is high enough for single photon detection when integrated with a low noise ROIC.[92] The minimum time between distinguishable events less than 10 ns has been reliably demonstrated. Figure 4-26 shows high-speed oscilloscope traces of spikes of single and multiple photon-initiated events obtained from the analog output of the detector at a gain of 500 and SNR ~14. The spectral response of the detector shows a high QE from 400 nm in the visible to the 50 percent cutoff wavelength at 4,330 nm in the MWIR.

The key attributes of the linear-mode HgCdTe, photon-counting, electron-initiated APD may be summarized as follows:

- No after-pulsing effects
- Ability to resolve photons that are closely spaced in time
- Ability to measure the number of photons in a multiphoton pulse
- Ultrahigh dynamic range: ability to run in photon counting mode at low flux levels and to transition seamlessly to ordinary linear mode at higher flux levels
- Broad and tunable spectral range offered by HgCdTe, and
- Photon detection efficiency of ≥50 percent

Linear-mode APDs performance data for several device architectures are tabulated in Table 4-2. The table compiles data from various sources taken under different conditions and therefore cannot be directly compared. Instead each column should be considered in a stand-alone manner. This table is by no means a complete or an exhaustive list of data from all the organizations; rather it presents the state of the art for specific device architectures.

Comparison of Linear-Mode and Geiger-Mode APDs

A direct apples-to-apples comparison of the performance of linear-mode and Geiger-mode APDs for active sensing applications is relatively complex and is best addressed based on individual applications and at the sensor level. Each approach has its own advantages and disadvantages, so in designing a receiver, the beneficial characteristics of each approach are leveraged to maximize sensor

[92] J.D. Beck et al., op. cit.

TABLE 4-2 Performance Parameters of Various Linear-mode APD Photon Detectors

Device Parameter	Si Arp[ad]	HgCdTe[b]	HgCdTe[c,d]	InGaAs[a]	InAlAs[e]
QE (percent)	35 at 1.06 µm	>90 at 1.06 µm (no ARC)	83 at 1.55 µm[f]	80	80 at 1.06 µm
Detector cutoff (µm)	1.1		4.3		1.2/1.2
Bandwidth (MHz)	140	500 (est.)	120[g]	100-1000	>1000
k (ionization ratio)	0.008	0	0		0.15
Excess Noise factor	3	1	1.3-1.4	TBD	8 (est. for *k*=0.15; M=45)
Gain M	120	200-350	468	30	45
Dark Current (pA)	50	0.5	1		
Bias (V)		16	13		71/82.5
NEP (fW/rt-Hz)	30	7-25 (at gain =100)	0.5	81	150/250
Dynamic Range (dB)	>20		> 30	22	
Pixel format	Single pixel	4 × 4	2 × 8 64 × 64 µm pix	Single pixel	Single/16 channel
Pixel area µm	700 (dia.)		64 × 64	200 (dia.)	75/100 (dia.)
Operating temperature (K)	300	101-120	84	300	300

[a]Krainak et al., 2009, "Photon detectors with large dynamic range and at near infrared wavelength for direct detection space lidar," *Proc. of SPIE* 7320: 732005-1.
[b]Wang et al., 2011, "n-on-p vertical architecture SAM design," *Proc. of SPIE* San Diego, CA, Aug.21-25.
[c]Beck et al., 2011 "Linear mode photon counting with the noiseless gain HgCdTe e-APD," *Proc. SPIE*, 8033: 80330N.
[d]Gleckler et al, 2012, "Application of an end-to-end linear mode photon counting (LMPC) model to noiseless-gain HgCdTe APDs," *Proc. Of SPIE* 8033: 1.
[e]Bai et al., 2012, "16 channel GHz low noise SWIR photoreceivers", *Proc. of SPIE,* 8353: 83532H-3.
[f] At unity gain flood illuminated. "QE percent" = QE*FF*eCE (quantum efficiency in active area 98 percent × fill factor 97 percent × electron collection efficiency 86.8 percent). Responsivity(e/ph) = QE*Gain.
[g] BW of Detector/ROIC.

performance. For example, detector or sensor performance comparison at the "per pulse" level could be misleading because in the linear mode, a single pulse to "detect" at low pulse repetition frequency (PRF) would compare with multiple pulses to "detect" at high PRF in the Geiger-mode.

For the case of photon-counting ladar, a systematic comparison between the two approaches based on statistical modeling has been reported.[93] The key conclusions that can be drawn from the study of Holmos et al. are the following:

- For the same low noise levels obtained in LM-APD as in GM-APD, a comparable ladar performance is expected.
- Under conditions of obscuration, GM-APD approaches the performance of LM-APD by increasing the number of pulses.

[93] M.J. Halmos, R.A. Reeder, B. Boland, J.P. Bulot, and M.J. Klotz, 2011, "Photon counting ladar," available at http://sti.usra.edu/clrc2011/presentations/Session%204/Summary%20M%20Halmos.pdf.

- For fast-moving targets or when the measuring time or the measuring number of pulses is limited, LM-APD may have an advantage over GM-APD, depending on the speed of the target.
- Intensity measurements in LM-APDs can be performed with fewer pulses than in GM-APDs.
- The ROIC unit cell circuit for LM-APD is considerably larger than for GM-APD, which may limit the pixel pitch in future arrays.

For fast moving targets, the linear mode can collect the image in one pulse with no concern about smear, but one concern for the Geiger-mode system is that the image will smear over the multiple pulses required for the "frame." In the aggressive case where 1000 pulses are needed to make an image (such as in an intensity image with multiple ranges per posting) at 100 kHz, the time required to collect an image is 10 msec. During this time, the angular rate that smears one pixel is 1 deg/sec, which can easily be corrected in processing using the output of an inexpensive gyro. The target velocity that smears one pixel (calculated as the resolution divided by the collection time (20 cm/10 ms)) is 20 m/s or 44 mph, so the Geiger-mode system could effectively image most ground vehicles in this regime. In more typical cases requiring 10 to 100 pulses to make an image, the image acquisition time would be even faster.

Note that different assumptions about the characteristics of the APDs (e.g., higher frame rate for the linear-mode APD, different sensitivity assumptions, etc.) and the sensors will change the results described above. A complete discussion of the trades between the two types of systems requires a detailed technical analysis and robust framework for comparison, which is beyond the scope of this report.

Both linear-mode and Geiger-mode APDs have been used for coherent-mode ladar. Coherent-mode detectors, which work by beating a local oscillator with the returning pulse in an active sensor, can provide information on the target reflecting the laser pulse. Some of the comparisons between the various technologies are illustrated in Table 4-3.

Geiger-Mode Detector Summary

- Dark count rate 1-10 kHz for thermoelectric (TE) cooled operation (263-300 K), array results reported with 33-37 percent photon detection efficiency (PDE).
- 64 × 256 array is currently the largest.
- Jitter of 135 ps.
- Primarily 1.06 μm, can perform at 1.55 μm and 2.4 μm with a trade off in DCR.
- Crosstalk <1 percent to adjacent pixels.
- <1 ns time resolution.
- Pros
  - Single photon sensitivity,
  - Sub-nanosecond timing information,
  - Direct photon-to-digital conversion on the focal plane, low noise readout,
  - Scalable to larger arrays,
  - High bandwidth (~GHz),
  - TE-cooled operation,
  - Compatible with direct and coherent sensing,
  - Extensive 3-D imaging results, and
  - Mature, fielded and commercially available technology.
- Cons
  - 1 μsec reset time required to avoid afterpulsing, and
  - Saturation flux ~1 x 10^9 photons/sec for typical arrays.

TABLE 4-3 Performance Parameters Comparison of Linear Geiger and Coherent Mode Detectors

Parameter	Linear	Geiger	Coherent
Dark count rate	Critical target $<10^4$	Critical target $<10^4$	Not critical I_{dark} overcome by LO
Spectral range	NIR – LWIR	NIR – SWIR	NIR – LWIR (Linear)
Bandwidth	GHz	GHz	GHz
Multipulse resolution on a single pixel	Nanoseconds	Microseconds Dead time from after-pulsing	Nanoseconds
Intensity	Yes (single pulse)	No (single pulse) Yes with multipulse statistics	Yes (single pulse linear)
Sensitivity limit	Subphoton	Subphoton	One photon
Probability of return detection	QE x P_{gain} x P_{amp} > 80%	QE x P_{gain} x P_{amp} 30-50% trade with I_{dark}	QE x fLO x P_{amp} > 80% with enough LO
Radiation hardening	High	Low	High
Focal plane power	Moderate to high	Low	High (LO on FPA)
Laser constraints	Minimal Odd pulse shapes and variable amplitudes OK	High rep rate required Need to adjust power to achieve subphoton laser returns	Severe High beam quality and alignment needed
Operating temperature	Ambient – 80 K trade DCR	Ambient – 200 K	Ambient – 80 K
ROIC complexity	High	Moderate	High

SOURCE: M. Jack, J. Wehner, et al., personal communication (2013).

Linear-Mode Detector Summary[94]

- Single photon sensitivity demonstrated.
- 6 nW MDS at 300 K 256 × 256 demonstrated (NEP~ 0.4 nW).
- Operability >95 percent.
- NEI < 0.5 photons at 80 K.
- NEP < 0.1 nW at 80 K.
- 1.5 GHz bandwidth demonstrated.
- Gains 100-1000 (at 80 K) for SWIR and up to 8000 for MWIR (at 80 K) demonstrated.
- DCR of 10-20 kHz at 150 K demonstrated.
- Fex <1.05 measured.
- Gain uniformity 3 percent.
- 12 e/s dark current at M=24.
- 320 × 256, 30 µm pitch ROIC 1.5 kHz Passive/Gated active readout.

[94] Based on results from different groups.

- Jitter < 100 ps.
- 1024 × 768, 20 μm pitch gated mode ROIC, 256 × 256 w/ 300 kV/A 100 MHz flash ROIC
- Pros
  - HgCdTe excess noise factor ~1.
  - MBE growth or LPE for HgCdTe.
  - Trades QE against J_{dark} (DCR).
  - Bandwidth.
  - Tunable gain.
  - Design flexibility with cutoff wavelength.
  - Compatible with linear mode and coherent sensing.
  - High bandwidth designs to 4.5 GHz.
- Cons
  - Sensitive to surface passivation.
  - MBE typically has higher defect rate than LPE.
  - ROIC more complex.
  - Under development.

Linear-mode APD arrays have very low detector reset time hence much less dead time, about 2-3 orders of magnitude shorter than current GM-APD technology. Since the aperture efficiency (number of photons detected versus the number of photons at the aperture) of GM-APDs is low, the target information is accumulated statistically from a multitude of laser shots.

High-gain linear-mode APDs have the ability to measure signal amplitude information, providing additional target information to support image processing and analysis. These traits reduce the laser power and number of shots and maximize the aperture efficiency of the system.

Geiger-mode APD arrays form the basis of 3-D imaging systems that are far more mature than those based on linear-mode staring arrays. GM-APDs have been successfully deployed in multiple missions as part of the Airborne Ladar Imaging Research Testbed (ALIRT) and High Altitude Lidar Operational Experiment (HALOE) systems.

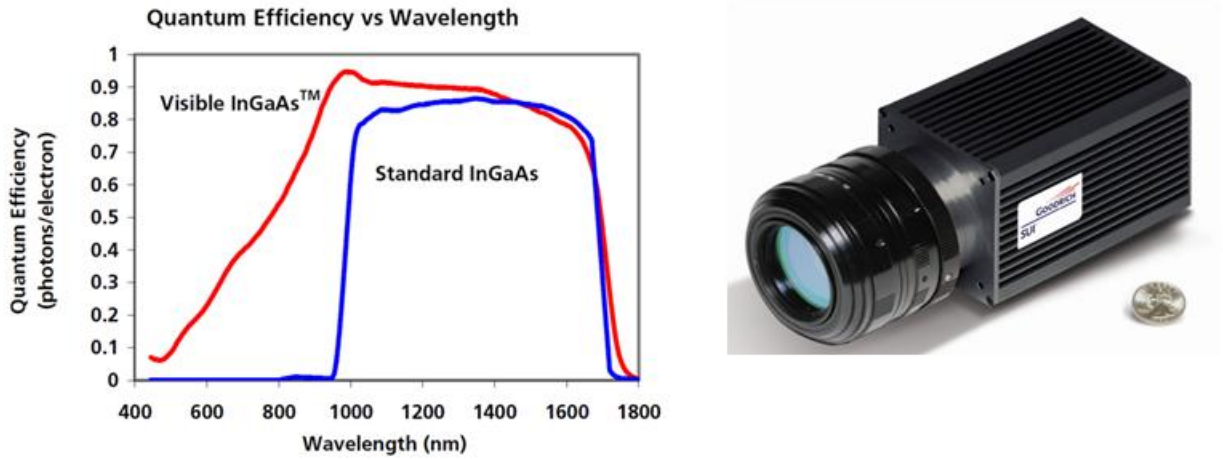

FIGURE 4-27 *Left:* Spectral response of $In_{0.53}Ga_{0.47}As$ photodiodes grown on InP. *Right:* Camera based on a 640 × 480 FPA. *Right:* Camera based on a 640 × 480 FPA. SOURCE: UTC-Goodrich.

FRAMING CAMERAS

Framing cameras are used for digital holography and polarization-based flash 3-D ladar. These are conventional passive FPAs, built with gated imaging capability in the ROIC. The FPAs are widely available from the visible to the LWIR wavelengths. In the visible both charge-coupled device(CCD) and CMOS-based imagers are available in the multimegapixel to tens of megapixel formats.

For commonly used ladar laser wavelengths of 1.06 and 1.57 μm cameras using InGaAs FPAs are available. The highest performing FPAs in this class are based on lattice matched $In_{0.53}Ga_{0.47}As$ grown on InP substrates. These backside illuminated (through the InP substrate) FPAs have sensitivity over the 0.9-1.7 μm band. The long wavelength end is limited by the bandgap of the InGaAs, while the short wavelength end is limited by the absorption in InP. The latter can be extended into the visible by removing the InP substrate. The quantum efficiency, however, falls off rapidly between 0.9 and 0.5 μm as shown in Figure 4-27.

Standard versions of the InGaAs FPAs are commonly in the 640 × 512, 25 μm pitch pixel format. Higher resolution versions are available in 1.3-megapixel format with 1,280 × 1,024 pixel and 12.5 μm pitch that operate at frame rates of 30 Hz. Depending on the required sensitivity, the FPAs are TE-cooled or operated at room temperature. The sensitivity achieved in these cameras in terms of NEI is between 1 and 2×10^9 photons/cm^2-s.

As longer wavelength lasers become more available and more popular for ladar applications such as those at ~2 μm and longer, detector arrays with longer cutoff wavelength are needed. Such arrays with varying cutoff wavelengths from the MWIR band to the LWIR in large array formats are increasingly available. The arrays with high sensitivity and integrated with high f-number optics require cooling to ≤150 K.

In the MWIR band, InSb FPAs as large as 8 K × 8 K have been reported (L-3 Cincinnati Electronics) with 10-12 μm pixel pitch.[95] In the LWIR and MWIR bands HgCdTe FPAs 1280 × 720 format with 12 μm pitch are available, and a 4K × 4K HgCdTe array has been demonstrated by Raytheon.

Owing to wide spectral response of the front-side-illuminated HgCdTe detectors (Figure 4-21), detector arrays fabricated with MWIR cutoff can also be used in the SWIR band by employing appropriate cold filters. An example of MWIR and SWIR imagery obtained with a 640 × 480, 12 μm pixel pitch format HgCdTe MWIR FPA is shown on the left in Figure 4-28. The camera operates at a frame rate of 60 Hz with the FPA maintained at 150 K. The noise equivalent temperature difference (NETD) measured in the MWIR band with 4 degree horizontal FOV is ~25 mK.

Most recently, cameras with 1,280 × 720 HgCdTe LWIR FPAs have been demonstrated with pixel pitches of only 5 μm at DRS with DARPA sponsorship.[96] This is the first detector array with a pixel pitch smaller than the detection wavelength (Nyquist limit). Similar format arrays have been developed in the MWIR band as well. These front-side-illuminated devices have broad spectral response and can be used in the MWIR and SWIR bands.

Gated Cameras

Range-gated imaging is performed primarily in the SWIR band with a 1.55 μm or 1.06 μm laser. In Chapter 2 this was described as 2-D imaging. This section describes two range-gated imaging technology approaches: one that has been in use for a few years now and another that was developed more recently.

[95] M.R. Kruer, 2012, "Infrared sensors for navy applications," keynote address presented at 2012 U.S. Workshop on the Physics and Chemistry of II-VI Materials, Nov. 27-29, Seattle, Washington.

[96] N. Dhar and R. Dat, 2012, "Advanced imaging research and development at DARPA," *Proc. SPIE* 8353: 835302.

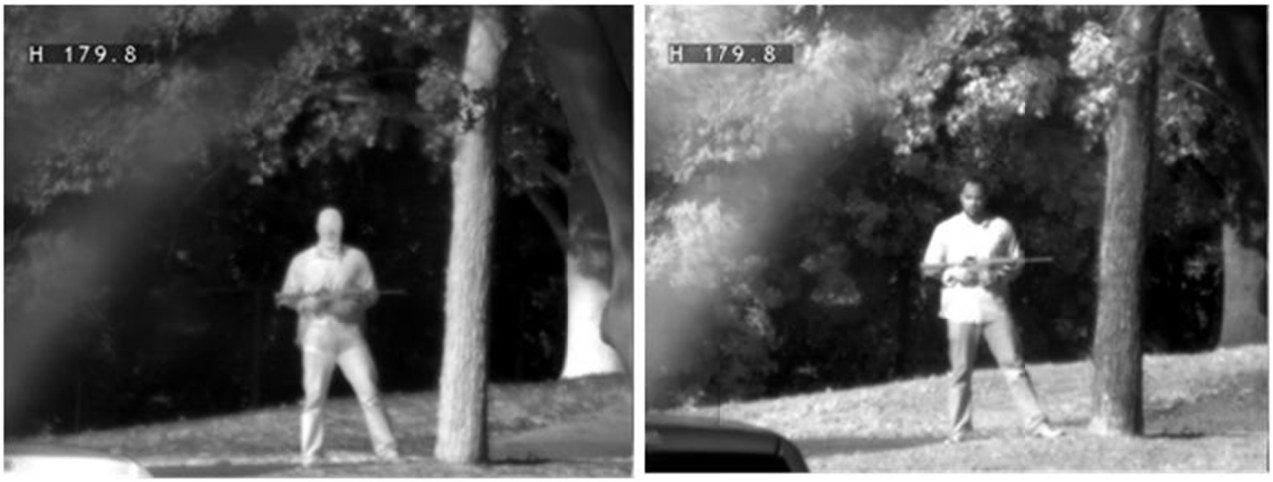

FIGURE 4-28 MWIR (*left*) and SWIR (*right*) imagery taken with a 640 × 480, 12 μm pixel camera operating at 150 K. SOURCE: DRS Technologies, Dallas, Texas.

The first of these is the Electron Bombarded Active Pixel Sensor (EBAPS), based on a GaAs or InGaAs photocathode in proximity focus with a high resolution, CMOS Active Pixel Sensor (APS) imager anode.[97] The SWIR version of this TE-cooled sensor has a spectral response that ranges from 950 to 1,650 nm. Electrons emitted by the photocathode are directly injected in the electron bombarded mode into the anode. Low noise gain is achieved in the CMOS anode by conversion of the high energy photoelectron (resulting from 1-2 keV bias applied between the photocathode and CMOS anode) to electron-hole pairs in the anode via the electron bombarded semiconductor (EBS) gain process. The electrons are collected in the anode pixel and subsequently read out. The EBS gain process is inherently low noise with an excess noise factor of less than 1.1.

Intevac's gated imaging sensor uses a 640 × 480 format, 13.4 μm pitch (with pixels binned at 2 × 2) FPA with QE of ≥20 percent at 1.55 μm. It has a minimum gate width of ~70 ns and a gate rise and fall time of 65 ns. SOURCE: http://www.intevac.com/wp-content/uploads/2012/01/LIVAR-M506-Datasheet1.pdf.

A second, more recent gated imaging system developed at DRS Technologies for the U.S. Naval Research Laboratory performs both gated 2-D active and passive functions in two spectral bands. It uses an MWIR HgCdTe e-APD FPA with 640 × 480 format, 25 μm pixel pitch. The MWIR APD FPA is fabricated using the HgCdTe technology described above. In the MWIR band the NETD at gain = 1 is ~18 mK (with f/4 optics and 16 ms integration) while in the SWIR band used for gated imaging the noise equivalent photons (NEPh) are ~2 photons for an integration time of 100 ns. The principle of range-gated imaging is illustrated in Figure 4-29. MWIR e-APD HgCdTe FPAs with on-demand detector gain along with the broad spectral response, such as shown in Figure 4-23, enables MWIR passive imaging, SWIR passive imaging and gated imaging at ~1.55 μm or 1.06 μm, commonly available laser wavelengths. Owing to the high ambient flux available in the MWIR band, no APD gain is generally needed, while in the SWIR gated mode, imaging with enhanced SNR is possible with detector gain. Figure 4-30 shows an example of passive imaging in the MWIR and SWIR bands as well as range gated imaging at 1.57 μm with 83 ns gate width all imaged with a single MWIR FPA. The SWIR gated imagery raw data show laser

[97] V.W. Aebi, K.A. Costello, P.W. Arcuni, P. Genis, and S.J. Gustafson, 2005, "EBAPS®: Next generation, low power, digital night vision," OPTRO 2005 International Symposium May 10, 2005, Paris, France.

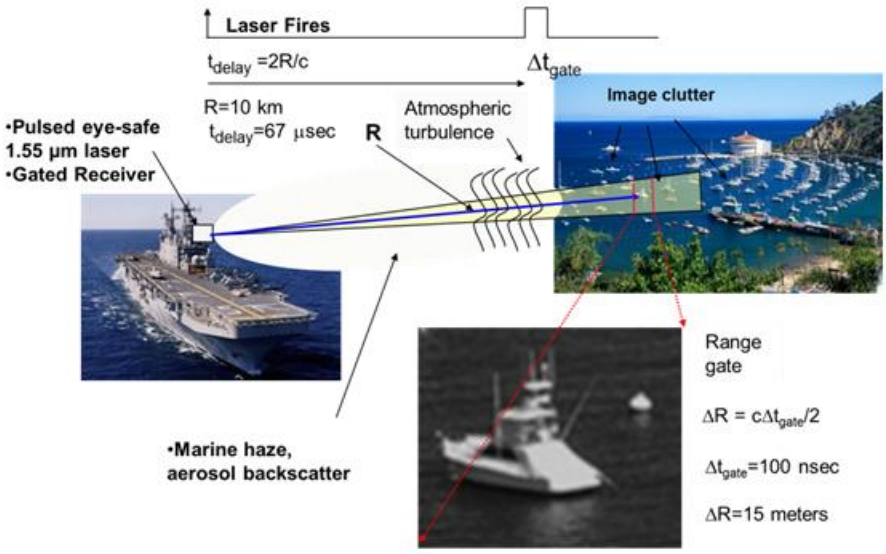

FIGURE 4-29 Illustration of the principle of range-gated imaging applied to a target in the midst of large background clutter. The use of 1.57 µm laser can significantly reduce the haze common in the marine environment. SOURCE: J. Waterman, Naval Research Laboratory.

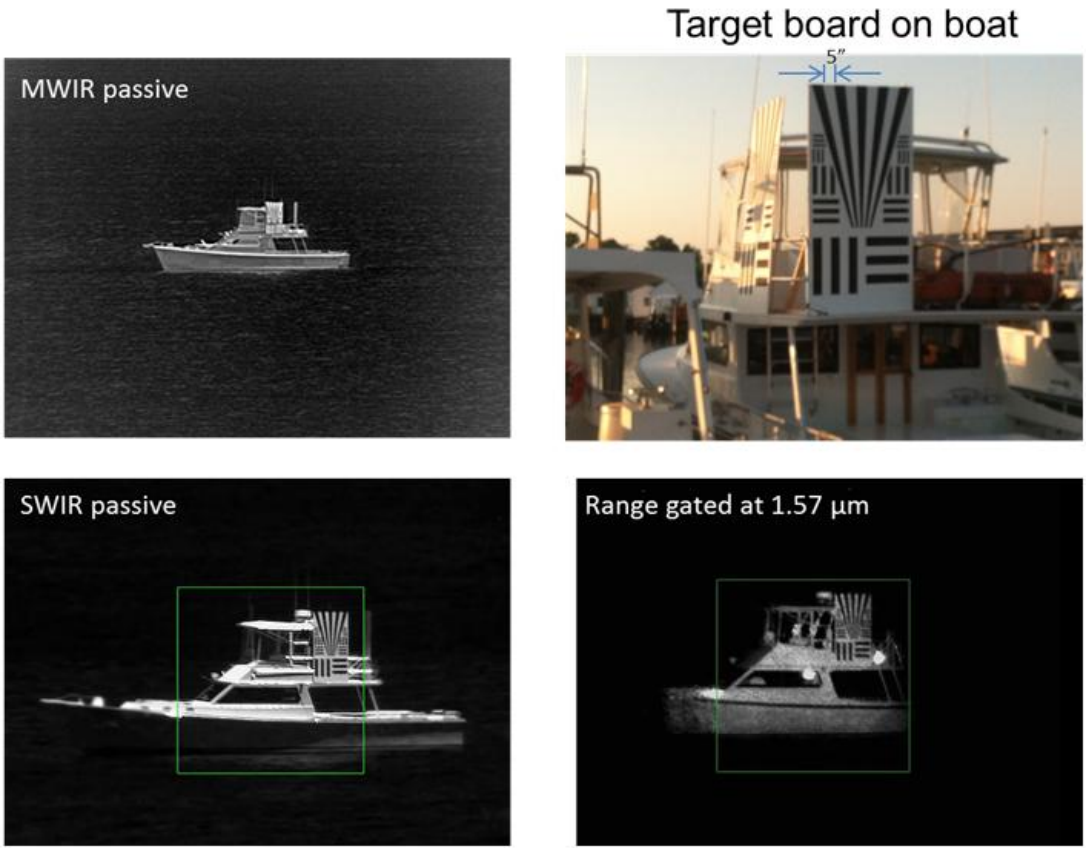

FIGURE 4-30 Passive imaging of a boat in the MWIR (*top left*) and SWIR (1-1.8 µm band), and range gated imaging with 1.57 µm laser all with the same FPA. Only the part of the boat that is illuminated with the laser is imaged (*bottom right*). The green squares in the SWIR imaging were image processed. SOURCE: DRS Technologies, Dallas, Texas.

speckle as is common due to interference effects. The image processed region within the green square drastically reduces the speckle effects, improving the image quality relative to the region outside the square.

A summary of both types of gated imaging FPAs is given in Table 4-4.

Conclusion 4-6: Geiger-mode avalanche photodiodes based on InGaAs/InP structures with single photon counting sensitivity in the 1-1.6 µm band are now available in 32 × 32, 32 × 128, and 64 × 256 pixel formats. Recently, this detector technology was extended to 2.2 µm operation using InGaAs/Gauss Type II superlative structure grown on InP. It is expected that this approach will be developed into array formats in the next few years.

Conclusion 4-7: Linear-mode avalanche photodiodes based on InGaAs/InP with single-photon counting sensitivity are available in pixel formats as large as 128 × 128 for wavelengths ≤1.6 µm. HgCdTe-based APDs with photon counting sensitivities in this wavelength band have been demonstrated in pixel formats up to 256 × 256. HgCdTe APD arrays with sensitivity from the visible to the midwave infrared (MWIR) wavelengths of 5 µm have been developed that are capable of real time 2-D and 3-D imaging. The latter arrays enable passive MWIR, passive SWIR, and active 2-D SWIR, all from the same 640 × 480 pixel format imaging array.

TABLE 4-4 Performance Characteristics of 2-D Gated Imaging IR Arrays

Array Technology	FPA Format and Pitch (µm)	Frame Rate (Hz)	Gate Width (ns)	ROIC Noise and Operating Temperature	Spectral Range and QE	Comments
Intevac[a,b] InGaAs/InP TE EBCMOS	640 × 480, 13.4 µm	<30	≥70 ns	~1 e^- per pixel/ µsec at 20°C; -40 to +20°C	0.95 to 1.65 µm QE≥ 20% at 1.55 µm	Gated 2-D, in production; NEPh=9 photons
SELEX[c] HgCdTe APD	320 × 256, 24 µm					
CEA/Leti[d] HgCdTe APD	320 × 240, 30 µm	1,500	≤1 µs	3-4 e^- normalized with detector gain of 10; ≤200 K	0.2–3.0 µm QE > 50%	Gated 2-D, 3-D SW
DRS[e] HgCdTe APD	640 × 480 25 µm	<120 Hz	≥25 ns	1-2 e^- normalized with detector gain of 70; 80 K	0.5-4.5 µm QE >50%	FPA capable of passive MW, SW and gated SW; Median NEPh~2 photons

[a]See http://www.intevac.com/intevacphotonics; LIVAR® M506.
[b]S. Demiguel, 2009, *Proc. of SPIE* 7298: 729836.
[c]I. Baker et al., 2008, *Proc. of SPIE* 6940: 69402L.
[d]J. Rothman, et al., 2012, *J. Electronic Materials* 41: 2928.
[e]P. Mitra DRS Technologies, Dallas, Tex.
SOURCE: P. Mitra, DRS Technologies, Dallas, Tex.

Conclusion 4-8: In the next few years it is expected both Geiger and linear-mode avalanche photodiode technologies will continue to develop to similar array sizes with comparable sensitivity. As the formats grow the pixel pitch will need to shrink to keep the size of the collection optics practical and from growing unreasonably large. Reducing the pixel pitch to sizes comparable to those being developed for passive imaging arrays will be a challenge for the ROIC technology, much more so than for the detectors. The sensitivity of linear-mode APDs is improving and is approaching that of GM-APDs.

Conclusion 4-9: The most difficult challenge for active sensing detectors going forward is the read out integrated circuit technology, regardless of which detector technology is used. While significant advances have been made in recent years, getting the signal out at high data rates, particularly for large-format arrays, remains a challenge and needs to be prioritized for future development. It is also an important area to watch in monitoring foreign developments.

Conclusion 4-10: For polarization sensing active imaging and digital holography (spatial heterodyne), a wide variety of formats is now available with standard detector technologies based on InGaAs, InSb, or HgCdTe all the way to 2K × 2K pixels. It is expected that these arrays will grow into larger formats and with pixel pitches of 5 μm or less. HgCdTe arrays with LWIR and MWIR cutoffs and response extending into the near IR have been demonstrated in 1,280 × 780, 5 μm pitch pixel format.

REMOTE ULTRA-LOW-LIGHT IMAGING

The imaging detector arrays discussed thus far are arrays of individual detectors the sensitivities of which range from many photons per pulse (e.g., "conventional" flash) to single photons (e.g., GM-APDs). In such array detectors, each pixel performs a timing measurement, and position information is derived from which detector fires. The Remote Ultra-Low Light Level Imaging (RULLI)[98] detector, on the other hand, is a single large-area position-sensitive detector and is sensitive at the level of a single photon over the entire area.

Originally developed for ultra-low light passive imaging (e.g., moonless nights[99]), the operating principle of RULLI (below) is inherently well-suited for active 3-D imaging as well, so long as dynamic range constraints are recognized and dealt with. A single photon incident upon a photocathode produces a single electron, which is accelerated to a microchannel plate stack that provides a gain of about 10^7. This macroscopic pulse of charge (localized in both time and space) then passes through a planar crossed delay line, as in Figure 4-31. This produces four current pulses that then propagate in both directions along both of the crossed serpentine delay line wires. By measuring the difference in delay time for the arrival of each current pulse at the ends of the delay lines, the *X*-*Y* position of the incident electron cloud can be localized, and by measuring the arrival time with respect to the outgoing laser pulse, a time-of-flight range measurement is performed.

Thus the key technological components of a RULLI detector are the photocathode, the microchannel plate stack, the crossed delay line, and the electronics. Figure 4-32 shows the RULLI crossed delay line and assembled sensor head for the Nocturnal Camera (NCam) imaging camera built by Los Alamos National Laboratory. The active area has a 25 mm diameter and an effective number of resolution elements of 500 × 500, resolving the position of the photon arrival with a full width at half maximum (FWHM) of about 80 μm. The timing accuracy for ranging is about 1 ns, and the system can detect up to 2 million counts/s. NCam operates in the visible with a quantum efficiency of 19 percent in the green. Dark noise is quite low: <100 counts/s/mm^2 at 22°C and <1 count/s/mm^2 at 0°C. The packaged

[98] D. Thompson, 2013, "Imaging one photon at a time," Los Alamos report number LA-UR-13-22617.

[99] Under starlit conditions, a visible remote sensing system at a range of 10 km with a 1 cm aperture and 100 percent detection efficiency will detect 1 photon per second per square meter from a target of albedo 1.

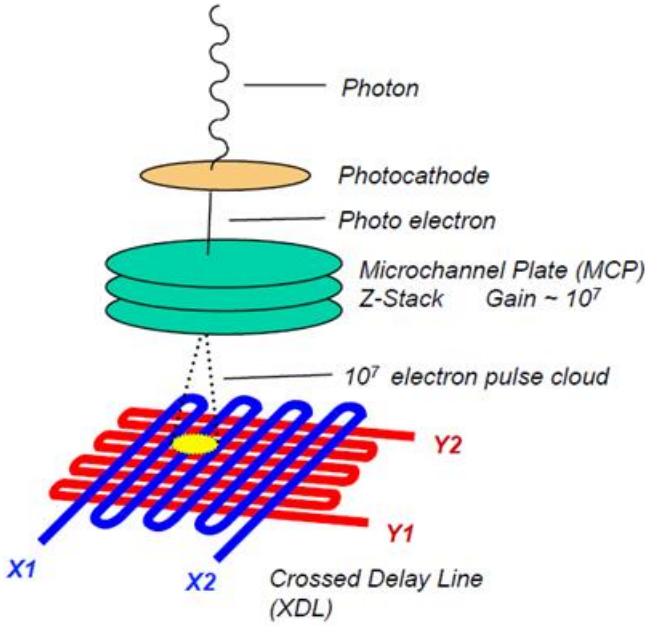

FIGURE 4-31 RULLI detector elements and concept.
SOURCE: D. Thompson, 2013, "Imaging one photon at a time,"
Los Alamos Report Number LA-UR-13-22617.

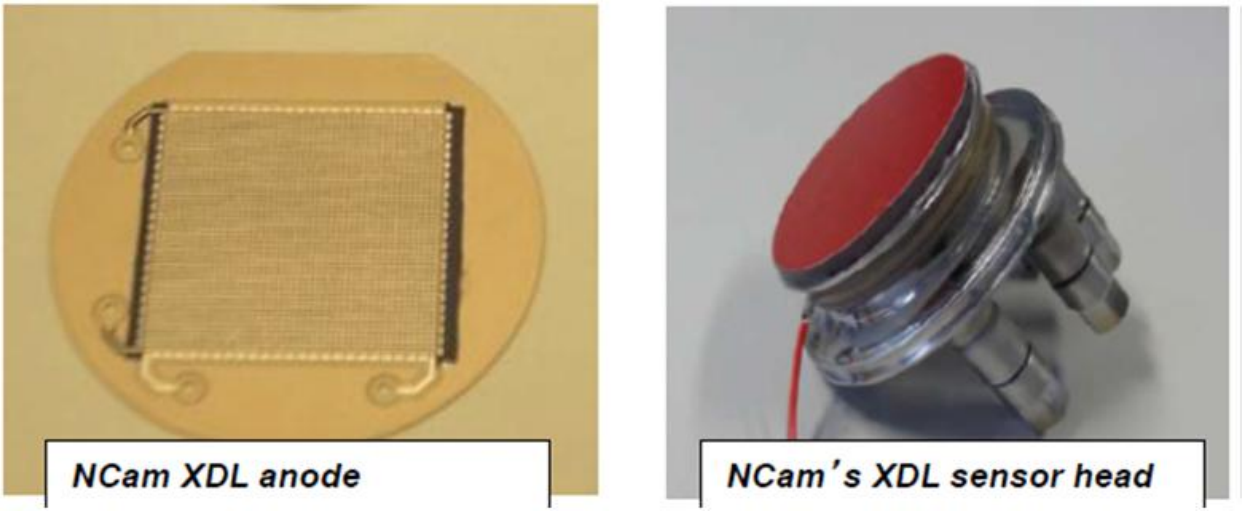

FIGURE 4-32 RULLI crossed delay line (*left*) and assembled sensor head (*right*). SOURCE: D. Thompson, 2013, "Imaging one photon at a time," Los Alamos Report Number LA-UR-13-22617.

NCam is 5 in. × 5 in. × 13 in. and weighs 10 lb. It uses ~30 W of power from a 28 V buss or battery and is suited for airborne operation.

RULLI can operate in spectral bands other than the visible, depending on the photocathode used. Figure 4-33 shows the spectral dependence of the QE for several available photocathode materials.

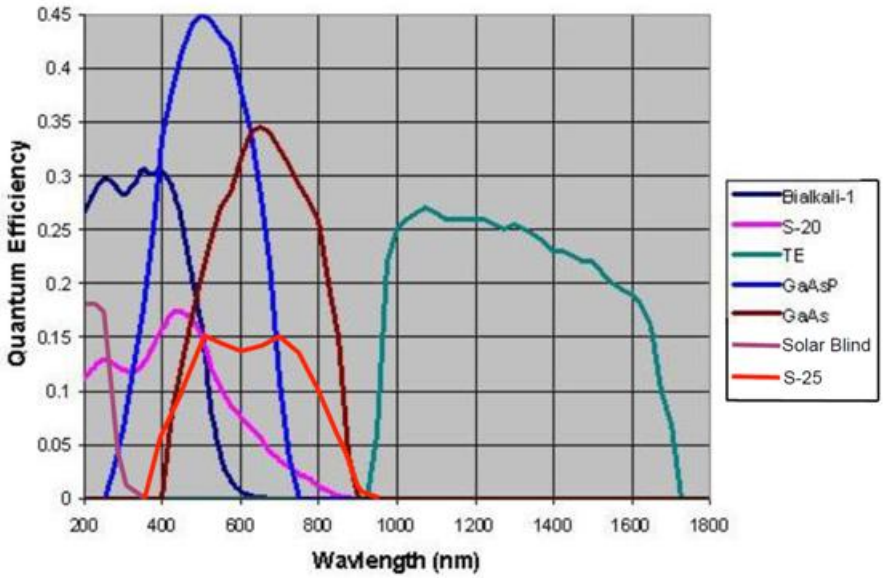

FIGURE 4-33 Spectral dependence of the quantum efficiency for several available photocathode materials. SOURCE: D. Thompson, 2013, "Imaging one photon at a time," Los Alamos Report Number LA-UR-13-22617.

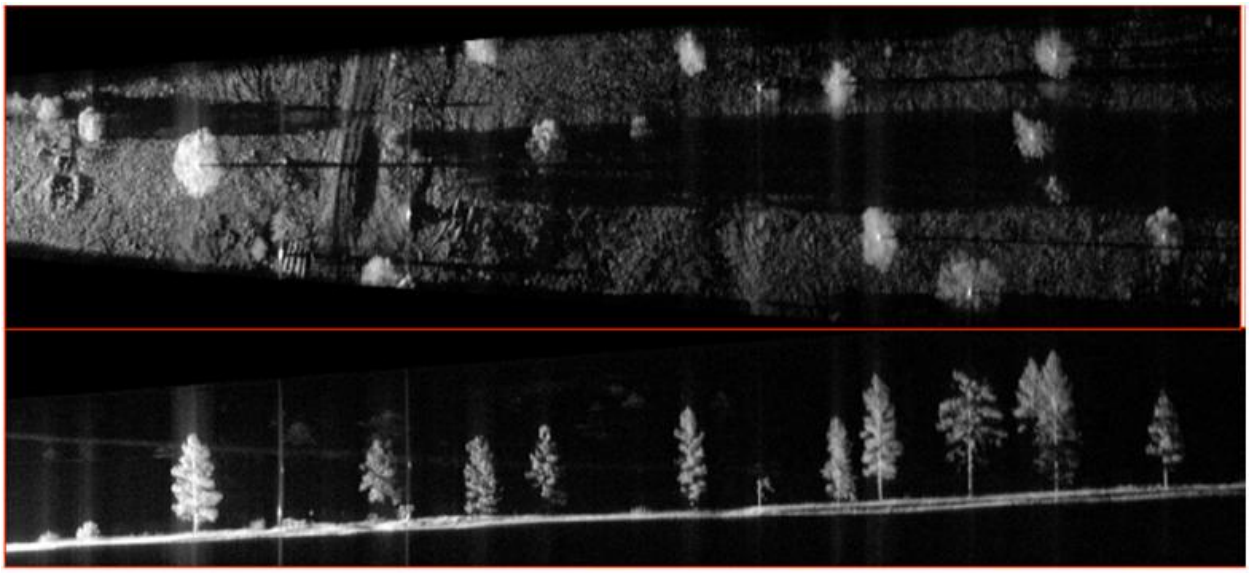

FIGURE 4-34 A typical RULLI/NCam active 3-D imaging product. SOURCE: D. Thompson, 2013, "Imaging one photon at a time," Los Alamos Report Number LA-UR-13-22617.

When operated in an active imaging mode, RULLI can produce high-quality 3-D products but operates in a rather different regime than some of the ladar systems described elsewhere in the report. Since multiple photoelectron pulses incident upon the crossed delay line can give rise to positional ambiguities, RULLI is limited to receiving only one photon per "frame." Due to the time necessary to read out the delay line, this imposes a maximum detection rate of about 2 million photons/s. In addition, this extreme sensitivity of RULLI means that it is very sensitive to background photons, favoring nighttime moonless operation and the use of very narrow spectral filters.

Because the scene, whether in passive imaging or ladar mode, is built up one photon at a time with a maximum count rate of 1-2 MHz, the imaging time for a scene can be rather long. RULLI/NCam employs both position and attitude knowledge as well as motion compensation and registration algorithms to build the scene images as well as identify movers within a scene.

Figure 4-34 shows a typical RULLI/NCam active 3-D imaging product. This image was taken from a ground perspective (note the shadowing) using a 1 MHz pulse repetition rate from a mode-locked laser over a 10 s exposure time. The range to target in this image went from 180 m to 395 m. One limitation of RULLI is due to the high PRF required to image scenes in a reasonable measurement time. For a constant PRF, this implies an ambiguity range (determining from which transmitted pulse a given return photon came and, hence, the range to target) of $R_a = c/(2 \cdot \text{PRF})$. This limitation can potentially be overcome, however, with PRF modulation techniques to resolve ambiguities.

RULLI typically operates with a short laser pulse, ~100 ps. Combined with the timing precision of the crossed-delay readout, this results in a high range resolution suitable for foliage penetration (FOPEN) scenarios. Active RULLI was recently used for performing topographic lidar in the marshy riverine environment near Savannah River,[100] which proved a very powerful technique compared with manpower-intensive GPS surveying and commercial airborne lidar (which typically operate with longer pulses not well suited to the marsh-grass environment.)

The low-count-rate limitation of the crossed-delay-line RULLI design can be mitigated by parallelizing the readout using an array of microstrips rather than a serpentine delay line. This is the concept behind RULLI-XS, which is being developed jointly by LANL and the University of California at Berkeley. Figure 4-35 shows the design concept, which has a goal of 10^8 photon counts per second over a megapixel array with <1 ns timing resolution.

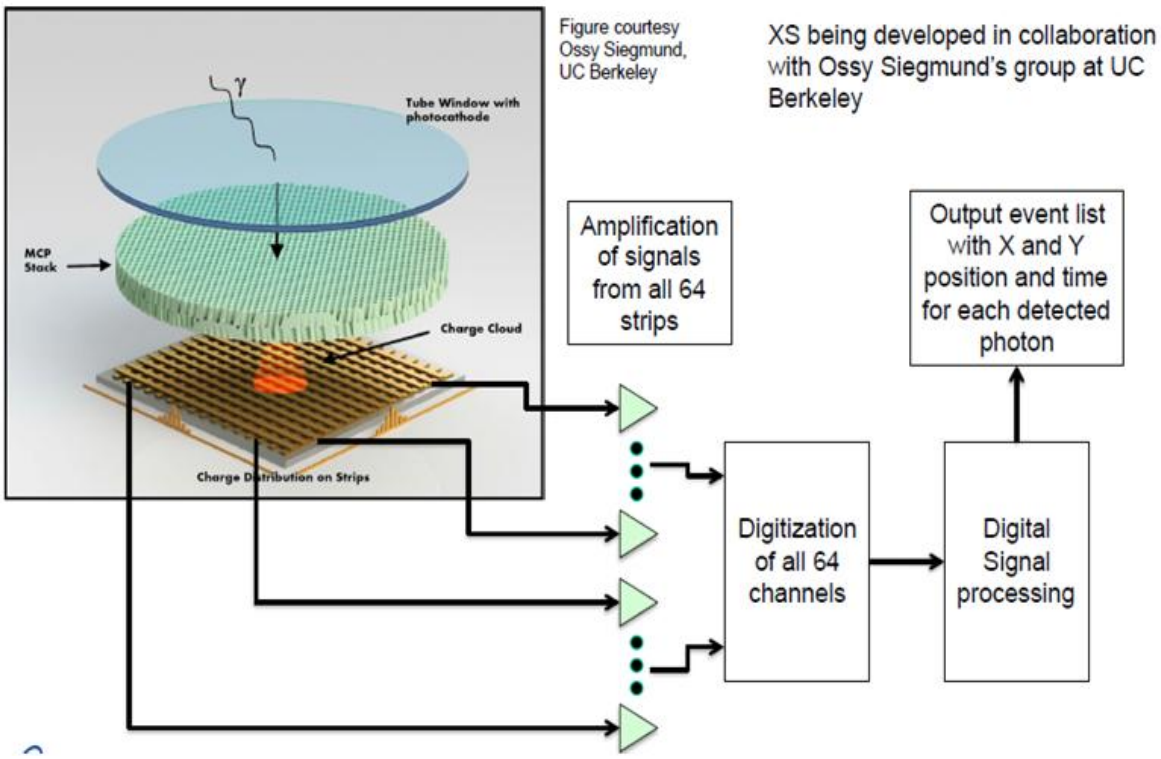

FIGURE 4-35 Design concept for RULLI-XS with parallelized readout to overcome count-rate limitations. SOURCE: D. Thompson, 2013, "Imaging one photon at a time," Los Alamos Report Number LA-UR-13-22617.

[100] D. Thompson, 2013, *Imaging One Photon at a Time,* Los Alamos Report Number LA-UR-13-22617.

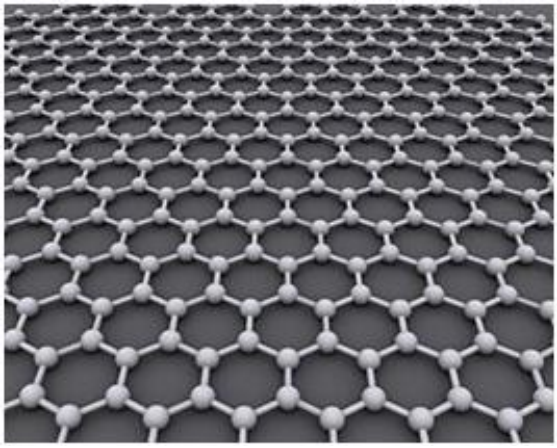

FIGURE 4-36 Graphene's honeycomb lattice of carbon atoms. SOURCE: By AlexanderAlUS (Own work) [CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], via Wikimedia Commons. Available at http://en.wikipedia.org/wiki/Graphene.

GRAPHENE

The electronic properties of graphene, one or more layers of a 2-D lattice of carbon atoms (Figure 4-36), are stimulating interesting research that may lead to important applications.[101] The field is burgeoning worldwide and results suggest graphene will have an important impact on active electro-optic systems.

The electronic band structure of a single layer of graphene is simple and was analytically calculated as early as 1947.[102] Because of its structural simplicity and unusual theoretical electronic properties, single layers of carbon were the basis for many early theoretical studies but the material was presumed not to exist in the laboratory.[103] In defiance of its presumed nonexistence, about a decade ago surprisingly robust single-atomic-layer graphene samples became available in the laboratory. Soon the unusual properties began to be published, stimulating an expanding interest in graphene.[104] The Nobel prize in physics for 2010 was awarded to Andre Geim and Konstantin Novoselov at the University of Manchester for experiments on graphene. In 2013, graphene researchers from Sweden's Chalmers University, led by Prof. Jari Kinaret, secured an unprecedented € 1 billion[105] from the European Union.[106]

[101] A.K. Geim and K.S. Novoselov, 2007, "The rise of graphene," *Nature Materials* 6: 183.

[102] P.R Wallace, 1947, "The band structure of graphite," *Physical Review* 71: 622.

[103] More than 70 years ago, Landau and Peierls showed that 2-D crystals were thermodynamically unstable and could not exist. R.E. Peierls, 1935, "Quelquesproprietestypiques des corpses solides," *Ann. I.H. Poincare* 5:177, and L.D. Landau, 1937, "Zurtheorie der phasenumwandlungen II," *Phys. Z. Sowjetunion* 11: 26.

[104] K.S. Novoselov et al., 2004, "Electric field effect in atomically thin carbon films" *Science* 306: 666.

[105] The Graphene Flagship will coordinate over 125 academic and industrial research groups in 17 European countries. Initial projects will be funded with $72 million for 30-months in addition to 20-to-30 additional targeted projects that will be announced.

[106] See http://europa.eu/rapid/press-release_IP-13-54_en.htm and http://www.eetimes.com/design/medical-design/4405947/EU-funds-research-on-graphene--brain-like-chips.

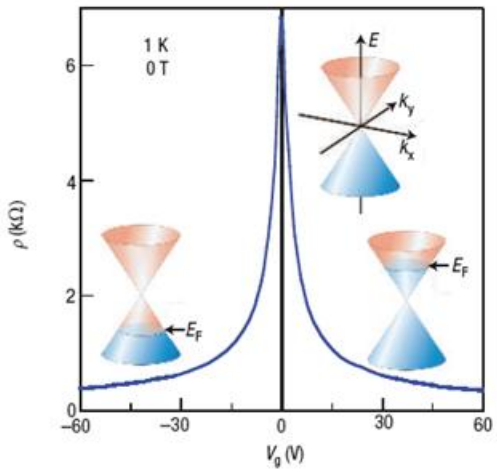

FIGURE 4-37 For a single layer of graphene the Fermi energy E_F changes with gate voltage V_g; positive V_g induce electrons and negative V_g induce holes. The carrier concentrations is $n = \alpha V_g$, where the coefficient $\alpha \approx 7.2 \times 10^{10}/cm^2/V$ for field-effect devices with a 300 nm SiO_2 layer used as a dielectric. The rapid decrease in resistivity ρ as charge carriers are added indicates their high mobility and does not noticeably change with increasing temperature to 300 K. SOURCE: Reprinted by permission from Macmillan Publishers Ltd: Nature Materials. A. K. Geim and K. S. Novoselov, 2007, "The rise of graphene," *Nature Materials* 6: 183, copyright 2013.

The electronic properties of a single layer of graphene derive from its conical symmetric energy bands, which produce a zero bandgap (Figure 4-37), making it a semimetal. The curve of energy versus wave vector *k* at low energy is linear, and that could lead to infinite charge mobility; in practice, however, the mobility is limited by impurity scattering or other defects. The symmetric valence and conduction band structure makes the electron and hole mobility equal, an unusual property.

Perhaps uniquely there is an electric field ambipolar effect whereby the Fermi level changes and charges carriers tune between electrons and holes as seen in Figure 4-37. Also, the charge mobility shows surprisingly little change with temperature, with measured mobilities that exceed 15,000 $cm^2/V/s$ and with predicted mobilities being an order of magnitude higher, likely exceeding that of semiconductors.

Additional discoveries show that graphene band structure also changes[107] with the following:

- Number of atomic layers of graphene,
- Details of the layer's edge boundaries, and
- Addition of dopants or defects.

Other unusual properties are extreme physical flexibility, a very wide photon absorption spectrum (from ultraviolet to multiple terahertz) and fast photo-induced carriers (subnanosecond response time.)

[107] See http://en.wikipedia.org/wiki/Graphene. The site shows an informative display of graphene band structure as a function of ribbon size and edge type. The zigzag edge structure is metallic while the "arm-chair" orientation can be semiconducting or metallic depending on the details of the edge and ribbon size.

Not all graphene properties are optimal. The band structure yields a constant value for the conductance and optical absorption over a wide spectral range, from visible to infrared.[108,109] The resultant probability of a photon absorption in a single layer of graphene is about 2.3 percent.[112,110] For high quantum efficiency detectors, multiple layers of graphene are needed.

The optoelectronic properties of graphene are a basis for a variety of photonic and electronic applications,[111] including fast, wideband photo detectors,[112] optical modulators,[113] and fast lasers.[114] A brief discussion of detector-related discoveries and applications follows.

Graphene Detectors

Many have recognized that graphene could make extremely wide-spectrum, highly sensitive, low-noise detectors at ambient-temperature. In 2012, Japanese, Russian, and U.S. researchers calculated the responsivity of dark-current-limited graphene photodiodes from the far IR to terahertz frequencies.[115] The researchers concluded that if multiple layers of graphene were used, the resultant detectors would have high QE and relatively low rates of thermo-generation—that is, the dark current and associated noise would be small. Because graphene-based detectors can exhibit high responsivity and detectivity at elevated temperatures in a wide radiation spectrum, the researchers say they can substantially surpass the performance of other detectors[116].

The benefits of graphene detectors are internationally appreciated. In June 2012 Iranian researchers published a theoretical analysis of an IR photodetector based on multiple layers of armchair-edge-boundary graphene nanoribbons (A-GNR) with a PIN structure[117] (Figure 4-38).

The goal was to extend the spectral capability and reduce the cost of HgCdTe and InSb photodetectors and imagers.

The Iranian researchers realized that armchair-edge graphene nanoribbons are semiconductors and that the electronically variable energy gap can be tuned from the visible through IR to THz. The study included three designs of detectors. They used the single-electron tight binding model to calculate the band structure and then derived the quantum efficiency, the temperature-dependent optical responsivity, and the variation of these properties with nanoribbon width and number of layers. They conclude A-GNRs have energy gaps in the IR range of electromagnetic spectrum and should be good IR photodetectors.

---

[108] A. Kuzmenko, E. Heumen, F. Carbone, and D. der Marel, 2008, "Universal optical conductance of graphite," *Physical Review Letters* 100.

[109] T. Stauber, N. Peres, and A. Geim, 2008, "Optical conductivity of graphene in the visible region of the spectrum," *Physical Review B* 78.

[110] R.R. Nair et al, 2008,"Fine structure constant defines visual transparency of graphene," *Science* 320:1308.

[111] F. Bonaccorso et al., 2010, "Graphene Photonics and Optoelectronics," *Nature Photon.* 4: 611.

[112] F. Xia et al., 2009, "Ultrafast graphenephotodetector," *Nature Nanotech.* 4: 839.

[113] Liu et al., 2011, "A graphene-based broadband optical modulator," *Nature* 474: 64.

[114] A.A. Lagatsky et al., 2013, "2μm solid-state laser mode-locked by single-layer graphene" *Appl. Phys. Lett.* 102: 013113.

[115] V. Ryzhii, N. Ryabova, M. Ryzhii, N.V. Baryshnikov, V.E. Karasik V. Mitin, and T. Otsuji, 2012, "Terahertz and infrared photodetectors based on multiple graphene layer and nanoribbon structures," *Opto-Electron Review* 20(1): 15.

[116] A. K.Geim and K. S.Novoselov, 2007, "The Rise of Graphene"*Nature Materials* 6(3).

[117] E. Ahmadi, A. Asgri, and K. Ahmadiniar, 2012, "The optical responsibility in IR-photodetectors based on armchair grapheme nanoribbons with PIN structure," *Superlattices and Microstructures*, 52: 605.

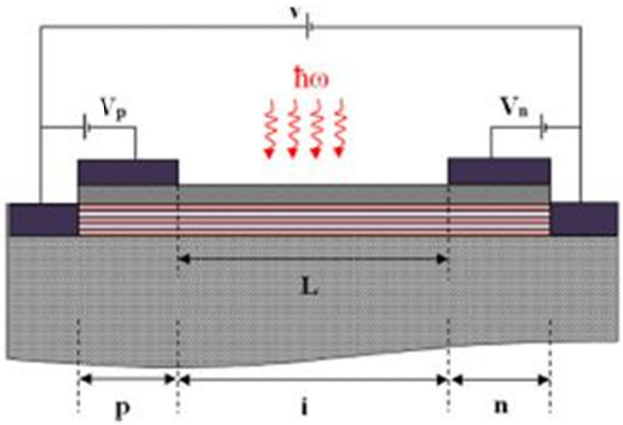

FIGURE 4-38 The schematic views of the A-GNR-based IR photodetector with electrically induced PIN junction. SOURCE: E. Ahmadi, A. Asgari, and K. Ahmadiniar, 2012, "The optical responsivity in IR-photodetector based on armchair graphene nanoribbons with p–i–n structure," *Superlattices and Microstructures* 52: 605, with permission from Elsevier.

Experimental verification of the broadband absorption and greatly improved sensitivity of a graphene quantum dot (GQD) enhanced detector is reported by Chinese researchers.[118,119] They report high photoresponsivity from monolayer graphene photodetectors with GQDs. The high photoresponsivity is partially attributed to internal gain of four times, called multi-excitation generation of carriers. This mechanism has been predicted theoretically[120] but has never been reported experimentally. It is caused by impact ionization, sometimes called inverse Auger recombination. The report includes a nearly simultaneous citation of the publication of another observation that bolsters the researchers' explanation and observations of gain in graphene.[121] They observed photoresponsivity about three orders of magnitude higher than that reported in the previous literature. In addition to utilizing gain, the increased performance is enhanced by the introduction of electron trapping centers and by optimizing a bandgap through band structure engineering.

The fabrication steps of the detector are relatively simple, as shown in Figure 4-39.

When tested over a large spectral range, there is broadband response shown from the visible (VIS) (532 nm), near-infrared (NIR) (1.47 μm) and mid-infrared (MIR) (~10 μm) ranges with photoresponses of: ~1.25 AW^{-1}, 0.2 AW^{-1} and 0.4 AW^{-1} in the visible, NIR, and MIR ranges, respectively.

A major limitation of the detector is slow speed, as seen in Figure 4-40.

[118] They have also filed for a patent through the university's Nanyang Innovation and Enterprise Office.

[119] Y. Zhang, T. Liu, B. Meng, X. Li, G. Liang, X. Hu, and Q.J. Wang, 2013, "Broadband high photo response from pure monolayer grapheme photodetector," *Nature Communications* 4:1811.

[120] T. Winzer, A. Knorr, and E. Malic, 2010, "Carrier multiplication in graphene." *Nano. Lett.* 10: 4839.

[121] K.J. Tielrooij et al., 2013, "Photoexcitation cascade and multiple hot-carrier generation in grapheme," *Nat. Phys.* 9: 248.

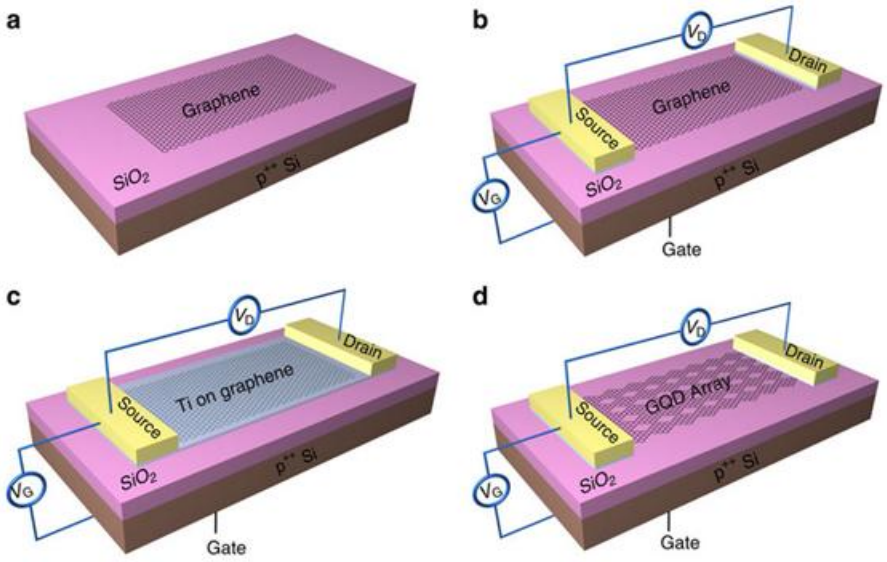

FIGURE 4-39 Fabrication of broadband graphene experimental detector: (a) a monolayer of graphene is placed on a SiO_2/Si substrate, (b) source and drain contacts are deposited, making a field effect transistor, (c) a Ti sacrificial layer is deposited for selective removal, and (d) depending on the Ti layer thickness, wet etching produces variable size regions of graphene quantum dots (GQDs). SOURCE: Reprinted by permission from Macmillan Publishers Ltd: Nature Communications. Zhang, T. Liu, B. Meng, X. Li, G. Liang, X. Hu, and Q. Jie Wang, 2013, "Broadband high photoresponse from pure monolayer graphene photodetector," *Nature Communications* 4:1811, copyright 2013.

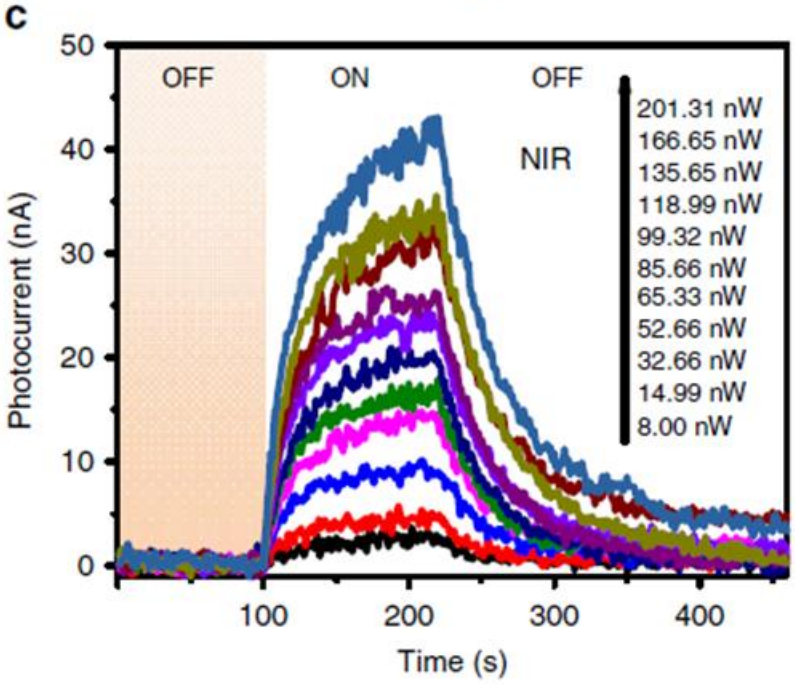

FIGURE 4-40 The slow response of a graphene dot detector to NIR light at 1.47 μm. Note the timescale is hundreds of seconds. SOURCE: Reprinted by permission from Macmillan Publishers Ltd: Nature Communications. Zhang, T. Liu, B. Meng, X. Li, G. Liang, X. Hu, and Q. Jie Wang, 2013, "Broadband high photoresponse from pure monolayer graphene photodetector," *Nature Communications* 4:1811, copyright 2013.

The slow response is attributed to traps, and there are multiple ways to reduce trap times and densities. It is well know that the boundary edges of graphene can change its character from semiconductor to metallic and even introduce defects. To better define the edge boundaries of a graphene dot Yongzhe Zhang et al. suggest an electron beam, a He ion beam, and focused ion lithography that should make more precise edges and reduce disturbances in other parts of the graphene dot. The report concludes the detectors are relatively simple to fabricate, and with defect engineering the performance will continue to improve.

Graphene detectors working in the terahertz range are also reported by Italian, French, and English researchers.[122] They measured a target at room temperature with a bilayer-graphene-based FET at $V_g = 3$. The detector was mounted on a *X*-*Y* translation stage having a spatial resolution of 0.5 μm. The THz image consists of 200 × 550 scanned points, collected by raster scanning the object in the beam focus, with an integration time of 20 ms/point. The experimental setup and results are seen in Figure 4-41.

Implications

Although graphene is an immature material technology, its potential for applications such as advanced lidar is significant. Imaging in all bands from visible to terahertz wavelengths with graphene detectors is possible, and several detector concepts are under intense development. Graphene-based, electronically-tuned, wide-spectrum imaging arrays operating at ambient temperature are possible future lidar system enablers.

Graphene edge boundaries (zigzag or arm chair) control its metallic and semiconductor properties but the boundaries require atomic accuracy and precision. Electron and ion beams are not now accurate and precise enough to make the desired boundaries because of collateral damage.

Conclusion 4-11: Electron beam processing is an important tool used in commercial semiconductor device production, but further advances will be required to develop the capability to make atomically correct graphene edge boundaries. Such techniques may be developed in the next 10-20 years, and if they are, the impact would be large.

Conclusion 4-12: Development of a scalable graphene detector array concept with appropriate fabrication tools is a noteworthy indicator that the graphene will enable new detector systems. Experimental single detectors are being theoretically explored and experimentally demonstrated for a variety of wavelengths. Advanced imaging systems likely require arrays of detectors. These arrays could be developed in the next 5-10 years and if this happens, they would have a large impact.

Conclusion 4-13: The United States has no significant lead in graphene research and development projects, and Europe has committed to the largest known graphene research and development program. Knowledge of the potential for greatly improved IR detectors and other applications is driving R&D on graphene worldwide.

QUANTUM DOT INFRARED DETECTORS

Quantum dots (QDs) are small particles with sizes on the border between chemical molecules and bulk materials. Like atoms or molecules, they have discrete electronic spectra, and like bulk materials, there are valence and conduction bands with gaps in between. In 1996, a theorist proposed to enhance IR

[122] L. Vicarelli et al., 2012, "Graphene field effect transistors as room-temperature terahertz detectors," *Nature Mater.* 11: 865.

detector performance by use of small 3-D embedded nanoparticles called quantum dots.[123] The electronic structure of a QD is modified compared with bulk materials due the relatively large surface boundaries and low volume, and the consequently confined size of electronic excitations, called excitons, within the dot. Excitons are bound electrons and holes. They are quasi-particles because the pair has less energy than the individual constituents, making the bound pair behave like a single particle. The average relative distance between an exciton's hole and its electron is called the exciton Bohr radius.

When the size of this hydrogen-like atom arrangement is limited by the dot volume, quantum confinement is produced. QDs are in the weak confinement if their effective radii are the size of the unconfined exciton Bohr radius or larger and in the strong confinement if their radii are smaller. If the QD is dominated by quantum confinement, typically at sizes less than 10 nm, the electronic and optical properties are highly tunable. The electronic properties of QDs have multiple degrees of freedom in addition to size such as the shape and distribution of the dots; these degrees of freedom include material composition, including dopants, and the nature of the surrounding matrix. QDs add to the topological quantum confinement options, which include zero-dimensional (a point), 1-D quantum wells (a line), 2-D quantum wires (a plane) and 3-D dots (a solid). Over the mid- and long-wavelength IR bands, QDs have predicted absorption coefficient of 10^4/cm; low dark current; spectral tunability by size, shape, composition, and surrounding matrix; and long photo carrier lifetimes.[124] The QD geometry produces a large density of states near the conduction bandedge, creating the large absorption coefficient and temperature independent carrier distributions. The 3-D quantum confinement suppresses low-energy thermal carriers, reducing dark current compared to that in bulk material. The effective carrier lifetime is extended by a phonon bottleneck caused when the bands of QDs do not match the phonon spectrum, and the long carrier lifetime should aid collection and increase detector efficiency.[125] Finally, unlike quantum wells, QDs absorb normal incident light because the geometry relaxes the polarization photon selection rules. There are claims that quantum dot infrared photodetectors (QDIPs) will outperform HgCdTe, InAs, and quantum well detectors because they are based on mature growth technologies and the low dark currents allow high temperature operation.[126] These claims have not been substantiated, but there is serious, multifaceted research and new detector designs are attempting to achieve the promise.

When an IR photon is absorbed by a QD, two types of electronic transitions occur. The first type is called bound to bound, where the ground state carrier goes to an excited state confined within the QD. The second is a bound to continuum, where the ground state carrier is excited directly into a state within the surrounding matrix of the QD. To be collected, the bound to bound excited carrier must escape the QD by thermionic emission or tunneling. Once the excited carriers are out of the QD, they must then drift or be accelerated to a collection contact. During the collection process, some carriers may be captured by another QD or the same original one. These effects slow the response time, and some carriers are lost as their energy degrades from multiple interactions. It is desirable to keep QDs close to collection contacts, however to prevent strain and maintain high absorption, QDIPs require thin heterogeneous stacks of 30 or more QD active regions separated by strain-reducing matrix materials. The photocurrent must transport the entire stack to produce a signal. An example structure that attempts to optimize all the QDIP requirements is shown in Figure 4-42.

[123] V. Ryzhii, 1996, "The theory of quantum-dot infrared phototransistors," *Semiconductor Science and Technology* 11: 759.

[124] H. Jiang and J. Singh, 1998, "Self-assembled semiconductor structures: electronic and optoelectronic properties," *IEEE J. Quantum Electron.* 34(7): 1188.

[125] There is controversy about the importance of phonon bottlenecks.

[126] C.F. de Souza, A. Alizadeh, S. Nair, I. Saveliev, M. Blumin, H.E. Ruda, D.C. Hays, V.H. Watkins, and K.R. Conway, 2010, "Mechanism of IR photoresponse in nanopatterned InAs/GaAs quantum dot PIN photodiodes," *IEEE Journal Of Quantum Electronics*, 46(5).

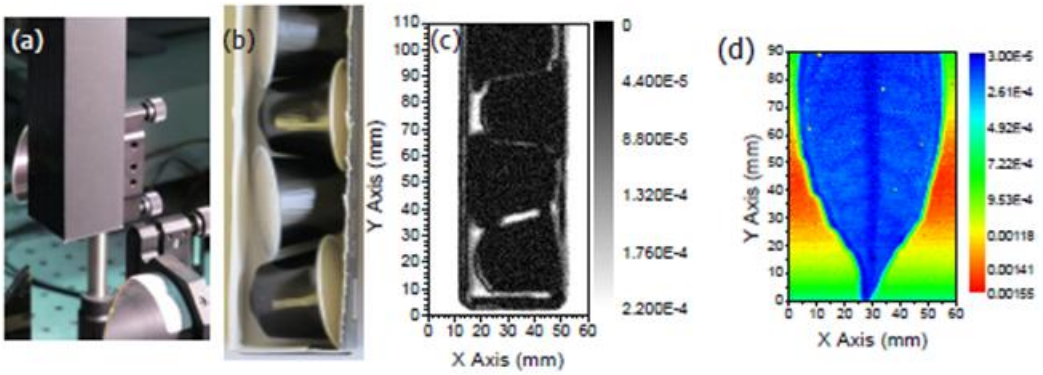

FIGURE 4-41 (a) Laboratory apparatus, (b) photo of a coffee-capsule box target, (c) 0.3 GHz transmission mode image of target, and (d) 0.3 GHz transmission mode image of a leaf measured under the same experimental conditions. SOURCE: Reprinted by permission from Macmillan Publishers Ltd: Nature Communications. Y. Zhang, T. Liu, B. Meng, X. Li, G. Liang, X. Hu, and Q.J. Wang, 2013, "Broadband high photoresponse from pure monolayer graphene photodetector," *Nature Communications* 4:1811, copyright 2013.

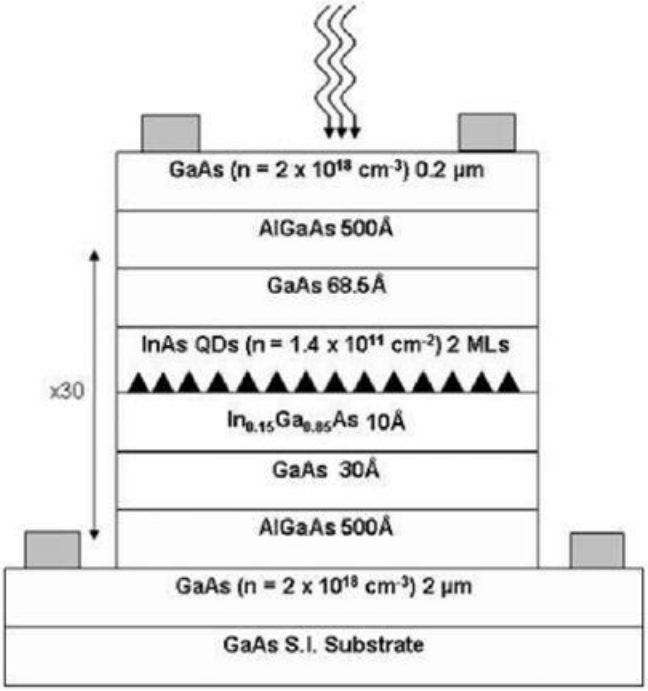

FIGURE 4-42 A 30-stack pixel element of a mid- and long-wavelength IR quantum dot in the well (DWELL) focal plane array. The design includes quantum wells that add quantum confinement to further reduce dark current. The dark current improvements have not yet been realized since reported DWELL arrays operate at 80 K. SOURCE: Copyright 2011 IEEE. Reprinted, with permission, from B.S. Paskaleva, W.-Y. Jang, S.C. Bender, Y.D. Sharma, S. Krishna, and M.M. Hayat, 2011, "Multispectral classification with bias-tunable quantum dots-in-a-well focal plane arrays," *IEEE Sensors Journal* 11(6), June.

Research Directions

The lack of theoretical performance from QDIPs could be related to QD growth mechanisms. The standard QD material is grown by MBE and materials such as InAs/GaAs. MBE produces QDs with lateral sizes from 15 to 25 nm, heights from 5 to 8 nm, and pyramidal shape. The density of InAs/GaAs QDs range from 10^9-10^{11} cm^{-2}. Dopants are added to produce two electrons in the filled ground state of the QD. For higher dopant densities, the dark current increases since the additional electrons occupy excited-state energy levels. These excited-state electrons can more easily contribute to the noise in the device by activating thermionic emission or field-assisted tunneling. For lower dopant densities, the absorption efficiency decreases since there are fewer electrons in the QD ground state available for photoexcitation. This trade-off is difficult to circumvent.

The nonuniformity of QDs grown by strained-layer epitaxy also limits realization of the predicted advantages of QDIPs. Materials improvements such as control of dopants, bandgap engineering such as DWELL, and the incorporation of resonant photonic structures are promising developments aimed at overcoming these limitations. In addition, fundamentally different QD material systems, such as colloidal QDs embedded in conducting polymers, could achieve inexpensive, room-temperature IR photodetectors. There are also proposals to add avalanche gain to QDIPs.

A recent innovative proposal suggests that photo avalanche gain can be added to QDIPs to achieve single-photon sensitivity.[127]

Calculations for this detector design predict that the self-quenching time is 150 ps; the dark current may allow operation at between 100 and 160 K and a detector efficiency of 8 percent at 6 μm.

Conclusion 4-14: Quantum dot infrared photodetectors promise to become an important technology in the field of infrared photodetection, particularly for the high operating temperature, high speed, multispectral imaging needs of the military and other sophisticated applications such as laser radar. Much research and development is needed to realize this promise.

OPTICAL ANTENNAS

Optical antennas can be thought of as extensions of microwave antennas into optical wavelengths. One definition of an optical antenna is this: "a device designed to efficiently convert free-propagating optical radiation to localized energy, and vice versa."[128,129] The approach is to use nanotechnology and plasmonics to create optical analogs of microwave antennas for detecting weak signals. Optical antennas exploit the unique properties of metal nanostructures, which behave as strongly coupled plasmas at optical frequencies and can focus electromagnetic waves into localized areas. To date, several reports have demonstrated surface plasmon-enhanced photodetectors using dipole antennas, metal hole arrays, and bulls-eye gratings. Figure 4-43 shows an example.

---

[127] M. Zavvari and V. Ahmadi, 2013, "Quantum-dot-based mid-IR single-photon detector with self-quenching and self recovering operation," *IEEE Electron Device Letters* 34(6): 783.

[128] P. Bharadwaj, B. Deutsch, and L. Novotny, 2009, "Optical antennas," *Adv. Optics Photonics* 1: 438.

[129] This is relatively new usage, in that in addition to being the receiving or transmitting element, it includes focusing light to be detected and presumably to assist in transmission. On p. 101 of an earlier NRC report *Seeing Photons*, an optical antennas is defined as a transduction device that is used to transmit or receive electromagnetic radiation. This is more restrictive usage, more like the definition used in the RF regime. *Seeing Photons* provides a discussion of optical antennas in the narrower definition. Because of the promise associated with focusing optical radiation, the expanded definition has been used lately, so additional discussion beyond that in *Seeing Photons* is required here.

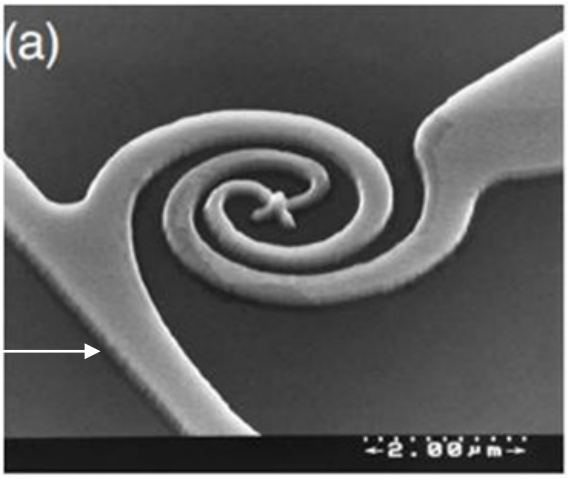

FIGURE 4-43 An example of optical antennas that enhance optical electric fields to a localized area for enhanced detection capability. Such tiny structures can be combined into sparse detector arrays. The scale in (a) is 2 μm. SOURCE: P. Bharadwaj, B. Deutch, and L. Novotny, 2009, "Optical antennas," *Advances in Optics and Photonics* 1: 438.

The antenna, using the definition above, can increase the optical energy density that falls on a photodetector, thereby increasing its efficiency. For a detector small compared with the wavelength λ, the received power is increased by the enhancement factor of the antenna (ratio of the intensity at the detector with and without the antenna). The expectation is that, as a result of an optical antenna, the optical intensity at the detector can be two to three orders of magnitude higher than the incident intensity. As was reported from Stanford, a half-wave Hertz dipole antenna (similar to those used for radio waves) of length ~ 380 nm for use at near-infrared wavelength of ~1.3 μm, was used to concentrate radiation into a nanometer-scale germanium photodetector. Antenna enhancement was demonstrated through a polarization contrast ratio of 20 in the photocurrent generated in the germanium element, which had an active volume of 0.00072 μm^3, two orders of magnitude smaller than previously demonstrated.[130]

While not as sensitive as other technologies, these detector arrays have the advantage of operating at room temperature and are valuable in the long-wave infrared (7-14 μm wavelengths). Much research on developing improved designs is under way. Of the many designs, only two more will be mentioned here. One is the promising use of surface plasmon antenna structures to concentrate light below the diffraction limit. Researchers at the University of California Los Angeles used nanopillars as absorbers for antennas in photodetector array applications.[131] According to Senanayake et al. "their small volume and high material quality results in a lower leakage current and junction capacitance, giving better noise performance and higher-speed operation than typical planar photodetectors... however, there is typically a tradeoff between the size of the absorbing volume and the quantum efficiency of the photodetector when the thickness of the semiconductor junction is reduced below the optical absorption length"[132]. Their solution takes advantage of antennas to boost the signal. Senanayake et al. determined that there was an

[130] L. Tang, S.E. Kocabas, S. Latif, A.K. Okyay, D.-S. Ly-Gagno, K.C. Saraswat, and D.A.B. Miller, 2008, "Nanometre-scale germanium photodetector enhanced by a near-infrared dipole antenna," *Nature Photonics* 2: 226.

[131] P. Senanayake, C.-H. Hung, J. Shapiro, A. Scofield, A. Lin, B.S. Williams, and D.L. Huffaker, 2012, "3D Nanopillar optical antenna photodetectors," *Optics Express* 20(23): 25489.

[132] Ibid.

18-foldenhancement of electric field intensity near the partial gold shell on the nanopillar sidewalls. A pitch of 580 nm gave resonances within the absorption spectrum of the $In_{0.35}Ga_{0.65}As$ material. Their experimental data indicated that absorption was enhanced by of a factor of 3 due to the localized surface plasmon supported by the metal cap.

In another demonstration of the ability of optical antennas to assist with measuring small optical signals, other Stanford researchers redesigned a metal-semiconductor-metal (MSM) photodetector consisting of a Ge nanowire so that its two metallic contacts served as optical antennas capable of concentrating light in the nanowire by a factor of 1.7.[133] The reported advantages of MSM photodetectors are high speed and low noise. The optical antenna was also designed to control the detector's polarization-dependent response.

Research in "optical antennas is currently driven by the need for high field enhancement, strong field localization, and large absorption cross sections in near-field applications."[134] This includes antennas for high-resolution microscopy and spectroscopy. Other suggested applications are increased efficiency in photovoltaics. The antenna can also be used as a light emitter, typically emitting controlled radiation from very small dimensions and used locally. In these ways optical antennas can make photonic processes more efficient and increase the specificity of gathered information.[135] Optical antenna arrays have not yet been demonstrated for long-range sensing, but this may happen in the future.

Some studies have predicted that intensity enhancements of 10^{4}-10^{6} are feasible. This means that optical antennas might reduce the area of discrete absorbers by 4-6 orders of magnitude without sacrificing any loss of absorption. One potential application might be mid-infrared sparse sampling detector arrays that would work at room temperature. The suggested increase of 10^{6} in signal could compete with Geiger-mode or other avalanche detectors.

Room-temperature microbolometer detector arrays have been developed for room-temperature sensitivity in the 7-4 μm wavelength range by BAE Systems, Lexington, Massachusetts, DRS Technologies, Dallas, Texas, Raytheon Vision Systems, Goleta, California, and others. In a commercial focal-plane array of 320 × 240 or 640 × 480 pixels, each microbolometer contains a thermally sensitive microbridge of vanadium oxide surrounded by silicon nitride.[136] The microbolometers are spaced at a 46 or 17 μm pitch respectively, and have a reported electrical NEP of 0.9 $pW/Hz^{1/2}$.

Researchers have integrated microbolometer arrays with optical antennas in order to increase the sensitivity. LWIR uncooled microbolometer detector arrays were reported in Russia in 2009.[137] The design was similar to the BAE arrays, but integrated antennas with the uncooled microbridge high-resistance microbolometers. The array was also 320 × 240 elements, with a reported threshold power of 160 pW/pixel in the 8-14 μm range and a response time of 12-16 ms. Their NEP appears to be 1.2 $nW/Hz^{1/2}$. This microbolometer array was also tested at terahertz frequencies. They measured a threshold power at 2.3 THz of 33 nW/pixel in a 160 × 120 format at a rate of 90 frames/s. This appears to be a NEP of 3.5 $nW/Hz^{1/2}$.

Recently, a sensitive camera integrating an array of THz optimized antenna-coupled microbolometers was demonstrated in France to spatially image a terahertz pulse.[138] A highly sensitive

[133] P. Fan, K.C.Y. Huang, L. Cao, and M.L. Brongersma, 2013, "Redesigning photodetector electrodes as an optical antenna," *Nano. Lett.* 13: 392.

[134] L. Novotny, 2009, "Optical Antennas: A New Technology That Can Enhance Light-Matter Interactions," *The Bridge* 39:4.

[135] Ibid.

[136] N. Butler, R. Blackwell, R. Murphy, R. Silva, and C. Marshall, 1995, "Dual use, low cost microbolometer imaging system," in Infrared Technology XXI, *Proc. SPIE* 2552: 583.

[137] M.A. Dem'yanenko et al., 2009, "Microbolometer detector arrays for the infrared and terahertz ranges," *J. Opt. Technol.* 76(12): 739.

[138] J. Oden, J. Meilhan, J. Lalanne-Dera, J.-F. Roux, F. Garet, J.-L. Coutaz, and F. Simoens, 2013, "Imaging of broadband terahertz beams using an array of antenna-coupled microbolometers operating at room temperature," *Optics Express* 21(4): 4817.

suspended membrane microbridge was developed at CEA-Leti, along with a terahertz antenna design based on crossed antennas and a resonant cavity, as shown in Figure 4-44. Arrays of 320 × 240 microbolometers were collectively processed at video frame rates above CMOS wafers by integration with a CEA-Leti designed application-specific integrated circuit. The estimated average power exciting a pixel was 0.26 nW, measured at 25 images/s. Their NEP appears to be $260/(25)^{1/2}$pW/Hz$^{1/2}$ = 52 pW/(Hz)$^{1/2}$, showing how well-designed antennas can increase the sensitivity compared to the Russian results. None appear yet to match the sensitivity of the BAE Systems detector.

Conclusion 4-15: None of the current antenna-coupled detectors is ready to be used in EO sensing applications, but in the next 5-10 years they may have a role to play in room-temperature, long-wave infrared detector arrays.

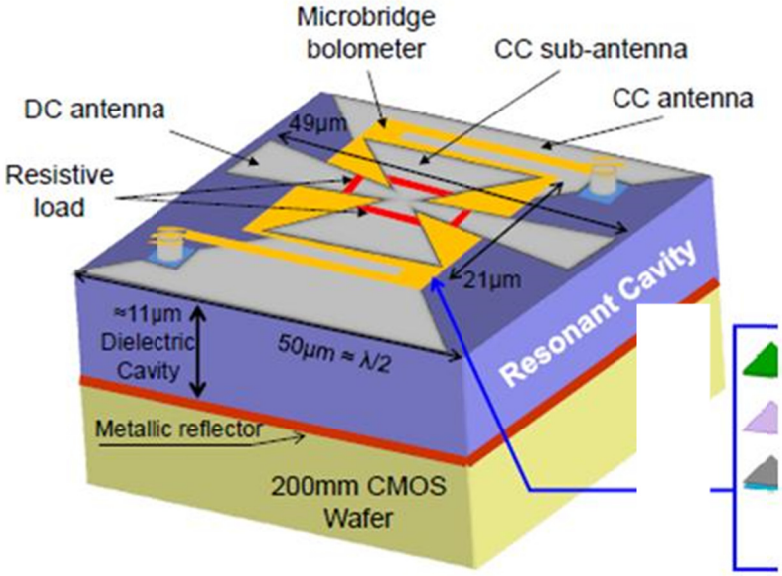

FIGURE 4-44 Geometry for microbridge microbolometer coupled to crossed microantenna. SOURCE: P. Bharadwaj, B. Deutsch, and L. Novotny, 2009, "Optical antennas," *Advances in Optics and Photonics* 1: 438.

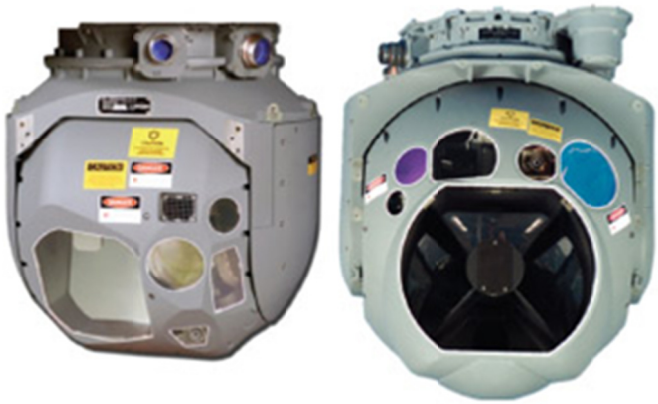

FIGURE 4-45 Two current optical gimbals: the Raytheon MTS-A (*left*) and MTS-B (*right*). SOURCE: http://www.raytheon.com/newsroom/technology_today/2012_i2/multispectral.html.

BEAM STEERING AND STABILIZATION

Active EO sensors need to point and stabilize the transmit beam and the receiver field of view. Gimbals are mechanically complex and expensive but can steer very accurately over wide angles. Except for obscurations, a three-axis gimbal can point in any direction. Figure 4-45 shows two current optical gimbals.

As can be seen, the mirror, which is the maximum size of the optical beam, is much smaller than the gimbal. A hole 2.5 to 3 times the size of the mirror must be cut in the vehicle on which the gimbal is mounted. Most military optical systems today use gimbal pointing. Gimbals can usually be pointed and stabilized very accurately but tend to cost hundreds of thousands of dollars.[139] Some estimates of gimbal cost are as follows: pilotage gimbals, perhaps 14 in. in diameter, cost on the order $75,000 to 100,000; medium altitude mission gimbals, perhaps 17 in. in diameter, on the order of $100,000 to 125,000, and high altitude gimbals, perhaps 22 in. in diameter, on the order of $150,000 to 175,000.[140]

Another very popular method of pointing is a fast steering mirror (FSM). An FSM is just a flat mirror that can be moved over small angles very quickly. Figure 4-46 shows an FSM. There is a trade-off with respect to the size of an FSM. The larger the mirror the more power it will take to move it quickly. For that reason one would like FSMs to be relatively small, so one wants to place them where the optical beam is small. At the same time the final steering angle is reduced by magnification, as shown below.

$$\vartheta_f = \frac{\vartheta}{M}$$

where θ_f is the final steering angle and θ is the actual angle steered by the steering mirror. M is magnification. If the beam is very small where one has the FSM in the system and the beam has to be magnified, then the steering angle is reduced by the magnification.

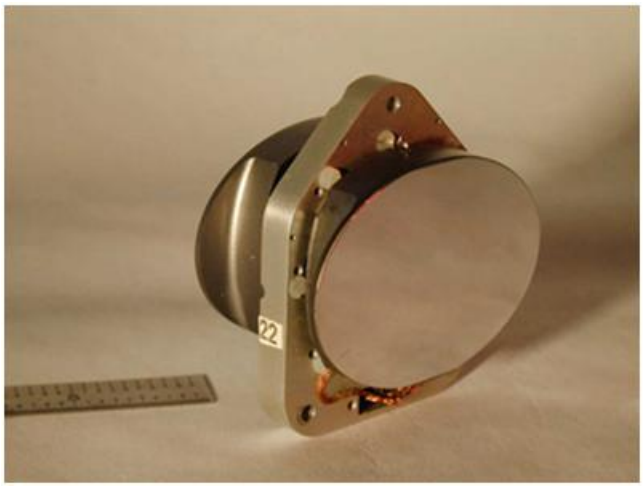

FIGURE 4-46 A fast steering mirror. SOURCE: Courtesy of NASA/JPL-Caltech.

[139] See http://www.churchillnavigation.com/gimbals, downloaded April 25, 2013.
[140] Personal communication from Jeanette Lurier, Raytheon, July 31, 2013.

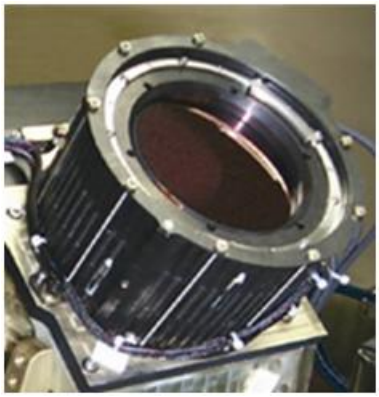

FIGURE 4-47 Early Lockheed Martin Risley prism. SOURCE: Courtesy of Dennis J. Adams, Lockheed Martin MST.

Steering flats look just like FSMs. The only difference is they are bigger and slower. One way to drive steering flats is using galvanometers. A "galvanometer is a type of sensitive ammeter, an instrument for detecting electric current."[141] "It is an analog electromechanical transducer that produces a rotary deflection of some type of pointer in response to electric current flowing through its coil in a magnetic field."[142]

Rotating polygonal mirrors have some advantages.

The only motion is simple rotation. Because the mirror has many sides, one does not have to rotate it fast to scan the angle rapidly. Of course, one can scan in only one dimension with a polygon.

Recently, Risley prisms,[143] and even Risley gratings,[144] have become more popular. Risley gratings are lighter and can be moved quicker but have dispersion. For narrow-band systems, Risley gratings have an advantage over Risley prisms. Figure 4-47 shows an early Lockheed Martin Risley steering stage. The OPTRA Risley prism steering stage, is an example of a commercial Risley steering device. Risley prisms work by rotating two or sometimes three steering prisms. It is possible to steer to large angles with Risleys using only rotation, a very simple motion. The OPTRA devices a linear fill factor of only a little more than 50 percent. Lockheed Martin has a Risley device with a linear fill factor of about 80 percent.[145]

Lockheed Martin increases the fill factor by placing the motor and much of the rotation mechanism in an area where the beam is smaller. Risley prisms have a zero crossing issue. If one tracks through zero one needs to rotate the wedges by 180 degrees. Risley gratings solve this problem by switching the rotation from one handedness to the other. One other advantage of Risley prisms is that they can be flush with the surface they are mounted in. Gimbals stick out into the wind stream but can point over very wide angles. Risley prisms are a flush, or conformal, steering approach. Of course any time one

[141] See http://en.wikipedia.org/wiki/Galvanometer, Feb 24.
[142] See http://en.wikipedia.org/wiki/Galvanometer, Feb 24.
[143] See http://www.optra.com/images/TP-A_New_Look_at_Risley_Prisms.pdf, December 6, 2013.
[144] C. Oh, J. Kim, J. Muth, S. Serati, M.J. Escuti, 2010, "High-throughput, continuous beam steering using rotating polarization gratings," *IEEE Photonics Technology Letters* 22(4): 200.
[145] D.J. Adams, A.F. Lindsay, B.L. Stamper, C.T. Knorr, A.P. Seibel, and M.R. Crano, "Risley Integrated Steering Module, Patent # 7898712 B2, March 12, 2011.

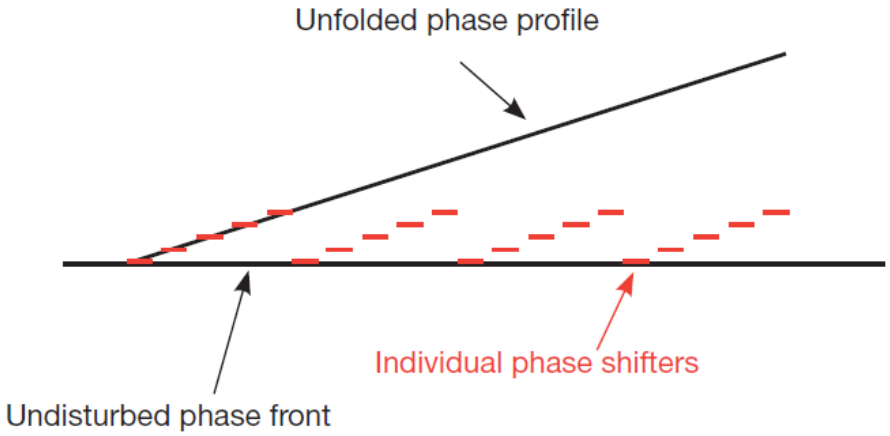

FIGURE 4-48 Modulo 2π beam steering at the design wavelength. SOURCE: Copyright 2013 IEEE. Reprinted, with permission from P.F. McManamon, P.J. Bos, M.J. Escuti, J. Heikenfeld, S. Serati, H. Xie, and E.A. Watson, 2009, "A review of phased array steering for narrowband electro-optical systems," *Proceedings of the IEEE* 97(6): 1078.

steers using a steering mechanism that is flush with a flat surface there is a cosine loss in power sent to a given angle, where the cosine is taken from a perpendicular direction. This cosine loss is due to the projected area of the aperture in the direction the beam is propagated.

Many nonmechanical approaches to beam steering were reviewed in a 2009 review paper.[146] A more informal discussion of was given in a magazine article a few years earlier.[147] The basic physics for any phased array beam steering approach is to modulate the optical path difference (OPD), or the phase, of the outgoing or incoming wave across an aperture to steer to a given angle. In order to minimize the amount of OPD or phase modulation required, often a modulo 2π phase modulation approach is used, subtracting 2π phase whenever the imposed phase become more than 2π. For single-frequency beams, this can be done with no effect on the steering angle, but 2π phase for one carrier frequency is not the same amount of OPD as for another carrier frequency, so this makes the steering approach dispersive. The basic approach to liquid crystal beam steering is to modulate the OPD and to use modulo 2π resets to limit the amount of OPD required. Basic liquid crystal beam steering using nematic liquid crystals at room temperature has about a 10 ms switching time. Modulating OPD is the same as modulating phase at the design wavelength. If the wavelength is not identical to the design wavelength, however, the beam will steer to a different angle, and usually sidelobes are developed. Figure 4-48 shows the unfolded phase for modulo 2π beam steering. As one can see in Figure 4-48, there also can be a digitizing error if the number of steps in a profile is limited.

The angle one can steer to at high efficiency using liquid crystal beam steering is limited to small angles primarily due to fringing field effects. When the 2π reset occurs, the OPD is supposed to drop sharply in position across the aperture from the 2π phase to the zero phase at the design wavelength. Because the OPD is generated by placing a voltage across a liquid crystal cell, and fringing fields limit how fast the voltage can change, this reset is not as sharp as it should be for liquid crystal beam steering. Figure 4-49 shows the effect. The net result is that a portion of the aperture steers the beam in the wrong

[146] P.F. McManamon, P.J. Bos, M.J. Escuti, J. Heikenfeld, S. Serati, H. Xie, and E.A. Watson, 2009, "A review of phased array steering for narrowband electro-optical systems," *Proceedings of the IEEE* 97(6): 1078.

[147] P.F. McManamon, 2006, "Agile nonmechanical beam steering," *Optics and Photonics News* 17(3): 21.

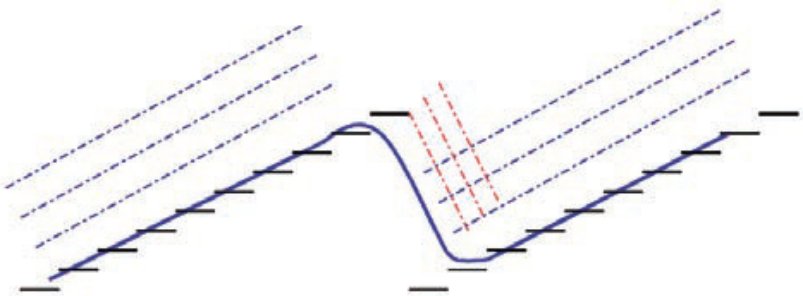

FIGURE 4-49 Reset limitations due to fringing fields. SOURCE: Copyright 2013 IEEE. Reprinted, with permission from P.F. McManamon, P.J. Bos, M.J. Escuti, J. Heikenfeld, S. Serati, H. Xie, and E.A. Watson, 2009, "A review of phased array steering for narrowband electro-optical systems," *Proceedings of the IEEE* 97(6): 1078.

direction, reducing efficiency, and high-efficiency steering is limited to small angles. Other efficiency losses using liquid crystal beam steering include the loss because each step is digital and also a loss due to spatial inertia in liquid crystals, which has an effect similar to the fringing field effect. Usually this second loss term is smaller than the fringing field effect, but not always.

Using the equations in McManamon et al.[148] it can be seen for 98 percent steering efficiency in one direction it is possible to steer to only about 0.2 degrees for a 1.5 μm wavelength and a liquid crystal birefringence of 0.35.

There are some new options for nonmechanical beam steering.[149] The basic physics of optical phased arrays are the same whether one uses a chip or uses birefringent liquid crystals to create OPD. The chip approach may not however have fringing field issues. The approach shown in the referenced paper does create OPD and does use modulo 2π beam steering. This particular chip is 64 × 64, with a pitch of 9 μm × 9 μm, making the array ~576 × 576 μm. The individual antenna radiators inside a pixel are 3.0 μm long, 2.8 μm wide. A second chip-scale optical phased array[150] uses random element pitch to reduce sidelobes but is also a very small chip. Often when phased-array individual elements are larger than a wavelength, people talk about the spurious sidelobes. Another way to consider this is that the largest FWHM angle that can be steered to without significant sidelobes is limited by the size of the individual phase shifters to[151]:

$$\vartheta = \frac{1.03\lambda}{d}$$

where λ is wavelength and d is the size of the individual phase shifter. Since steering is in only one direction and a higher efficiency than 50 percent (half max) is desired, it is necessary to restrict the steering angle to about one-fourth of that value. If 3 μm is used as the size of the radiator and 1.5 μm is

[148] P.F. McManamon, P.J. Bos, M.J. Escuti, J. Heikenfeld, S. Serati, H. Xie, and E.A. Watson, 2009, "A review of phased array steering for narrowband electro-optical systems," *Proceedings of the IEEE* 97(6): 1078.

[149] J. Sun, E. Timurdogan, A. Yaacobi, E.S. Hosseini, and M.R. Watts, 2013, "Large-scale nanophotonic phased array," *Nature* 493: 195.

[150] D. Kwong, A. Hosseini, Y. Zhang, and R.T. Chen, 2011, "1 × 12 unequally spaced waveguide array for actively tuned optical phased array on a silicon nanomembrane," *Applied Physics Letters* 99: 051104.

[151] J. Goodman, 1968, *Introduction to Fourier Optics*, San Francisco, Calif.: McGraw-Hill, 65.

the wavelength, steering can be done to about 1/8th radian, or about 7 degrees. In one example,[152] the chip is only about 576 μm—a very small aperture. Assuming a 10 cm beam, this implies a magnification of about a factor of 174, reducing the 7 degree steering angle to 0.04 degrees. Over time the steering arrays will become larger and the individual steering elements will became smaller. Sun et al. discuss only transmitting, not receiving, which is another required growth area. It is likely that nanofabricated phased arrays will become an interesting option for steering to small angles, using magnification after steering.

Another non-mechanical option for beam steering not covered in the McManamon et al., 2009 paper is called a Steerable Electro Evanescent Optical Refraction (SEEOR) prism. This device uses liquid crystals (LCs) "as an active cladding layer in a waveguide architecture where light is confined to a high index core and the evanescent field extends into the variable-index liquid-crystal cladding,"[153] This allows very large optical path delays, so for small apertures it eliminates the need for a 2π reset. Therefore, there is no fringing field issue. Also, because the liquid crystal layer is thin these devices can be relatively fast, responding in under 500 μs.[154,155] In-plane steering is accomplished by changing voltage on one or more prisms filled with liquid crystals, as shown in Figure 4-50.

Out-of-plane beam steering is based on the waveguide coupler designed by at Bell Laboratories in 1971 (see Figure 4-51).[156] In any waveguide, "if the cladding is too thin, light will leak out of the guided mode."[157] "In a planar slab waveguide Snell's law gives the propagation angle of the escaping light."[158] Since it is possible to tune the effective index of the waveguide, it is therefore possible to tune the angle of the escaping light.[159]

This waveguide-based liquid crystal beam steering can steer over relatively wide angles rapidly, but is limited to small apertures. In-plane steering can be up to 60 degrees or more in either direction, depending on the exact geometry. Out-of-plane steering can be 15 degrees or so in either direction. The main limitation on this technique is the size of the aperture, which will be limited to ~1 cm on a side or less.

Because of angle limitations on high-efficiency liquid crystal beam steering, and aperture size limitations on both the chip-scale steering and waveguide-based LC steering, there has been an interest in wide-angle step-stare steering approaches. Two techniques were initially used for wide angle beam steering in conjunction with liquid crystal steering: holographic glass and birefringent prisms.[160] From McManamon:[161]

---

[152] J. Sun, E. Timurdogan, A. Yaacobi, E.S. Hosseini, and M.R. Watts, 2013, "Large-scale nanophotonic phased array," *Nature* 493: 195.

[153] S.R. Davis, S.D. Rommel, G. Farca, B. Luey, N. Rebolledo, S. Selwyn, M.H. Anderson, 2011, "A new photonics technology platform and its applicability for coded aperture techniques," *Proc. SPIE 8165:*81651E, Unconventional Imaging, Wavefront Sensing, and Adaptive Coded Aperture Imaging and Non-Imaging Sensor Systems.

[154] S.R. Davis, S.T. Johnson, S.D. Rommel, M.H. Anderson, J. Chen, and T.-H. Chao, 2013, "New electro-optic laser scanners for small-sat to ground laser communication links," *Proc. SPIE* 8739: 87390H, Sensors and Systems for Space Applications VI.

[155] S.R. Davis, G. Farca, S.D. Rommel, S. Johnson, and M.H. Anderson, "Liquid crystal waveguides: New devices enabled by >1000 waves of optical phase control." *Proc. SPIE 7618:*76180E, Emerging Liquid Crystal Technologies V.

[156] R. Ulrich, 1971, "Optimum excitation of optical surface waves," *Journal of the Optical Society of America* 61: 1467.

[157] S.R. Davis, G. Farca, S.D. Rommel, S. Johnson, M.H. Anderson, "Liquid crystal waveguides: New devices enabled by >1000 waves of optical phase control," *Proc. SPIE 7618:*76180E, Emerging Liquid Crystal Technologies V.

[158] Ibid.

[159] Ibid.

[160] P. McManamon, 2006, "Agile Nonmechanical Beam Steering," *Opt. Photon. News* 17(3): 24-29.

[161] Ibid.

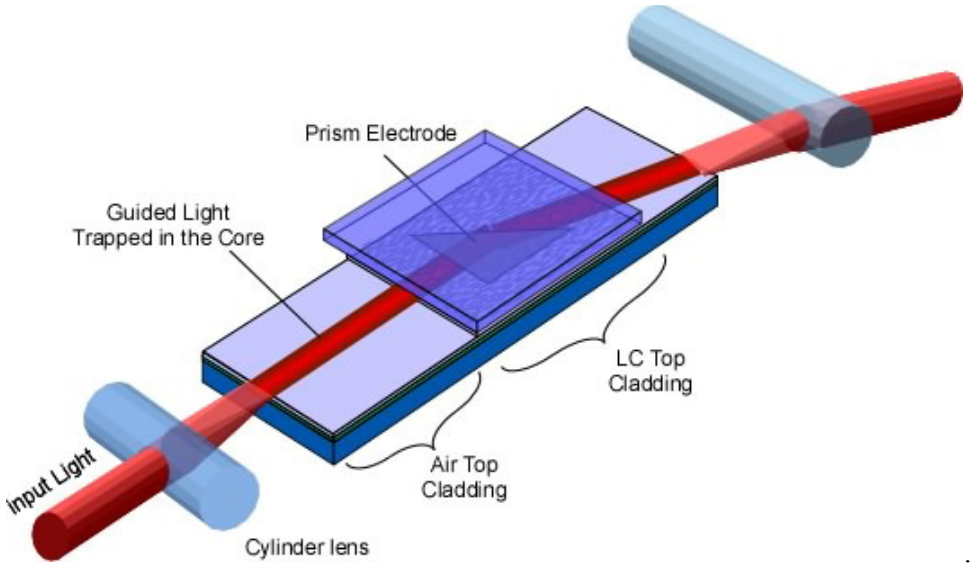

FIGURE 4-50 Voltage-tunable liquid crystal waveguide based in-plane steering. SOURCE: S.R. Davis, G. Farca, S.D. Rommel, S. Johnson, and M.H. Anderson; "Liquid crystal waveguides: new devices enabled by >1000 waves of optical phase control," *Proc. SPIE* 7618:76180E, Emerging Liquid Crystal Technologies V.

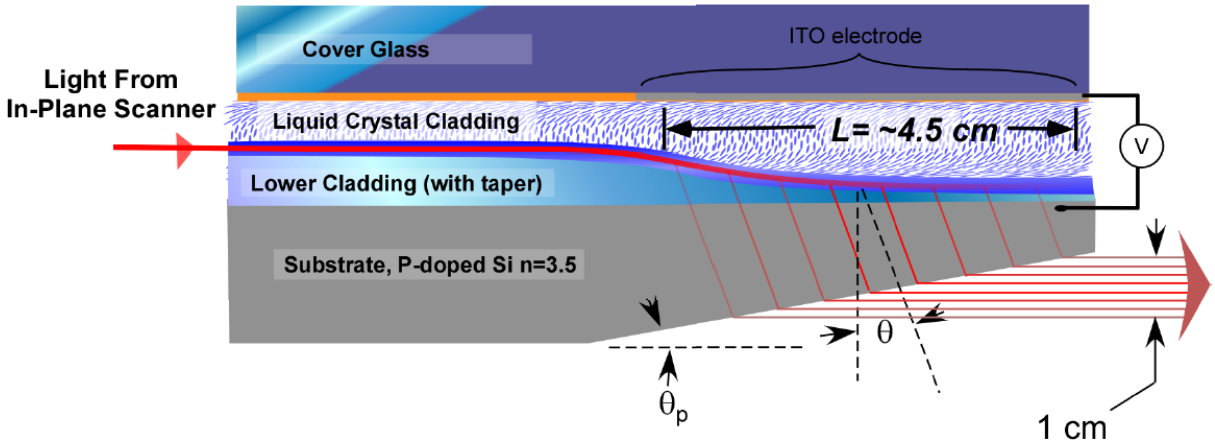

FIGURE 4-51 Liquid crystal waveguide for out-of-plane steering. SOURCE: S.R. Davis, G. Farca, S.D. Rommel, S. Johnson, M.H. Anderson, "Liquid crystal waveguides: new devices enabled by >1000 waves of optical phase control," *Proc. SPIE* 7618:76180E, Emerging Liquid Crystal Technologies V.

In the holographic glass approach, multiple holograms are angularly multiplexed in a single piece of glass. The holograms produce large angle deflections.

A liquid crystal beam steerer in front of the glass selects which hologram is addressed, resulting in the desired large angle of deflection.

Many layers of holographic glass can be used.

A second approach to wide-angle beam steering was birefringent prisms.

This technique also starts with an azimuth/elevation liquid crystal beam steerer for small- to moderate-angle beam steering.

Larger angles are reached using a set of binary birefringent prisms.

One polarization sees the prism deflecting the beam in one direction, while the other sees the prism deflecting it the other way.

More recently the most popular wide angle beam-steering approach is digital steering using polarization birefringent gratings. Fixed-angle digital versions of this steering are similar to birefringent prisms but use a set of gratings steering to fixed angles instead of using prisms. Such prisms are very efficient and very thin.[162] They are based on a fundamentally different method of nonmechanical beam steering,[163] in which the phase rather than the OPD is modulated. At the design wavelength this does not make any difference, but at any other wavelength there will be difference. The relatively mature approach to polarization birefringent steering uses set digital steering angles. There has been some work toward steering to adjustable angles that uses this approach, but it is a relatively immature at this time and very slow.[164]

Microelectromechanical systems (MEMS), lenslets, and electrowetting are alternative methods of nonmechanical beam steering.[165] At this time they are not as efficient as using the polarization birefringent gratings in combination with either an FSM or a liquid spatial light modulator for small angle steering.

Ideally one would like optical phased arrays similar to RF phased arrays, where it is possible to build a transmit/receive (T/R) module that is a fraction of a wavelength in size. In an ideal world, each of those T/R modules would be able to delay phase enough to provide true time delay steering. This means that a 45-degree steering angle, with a 20 cm diameter aperture would have a phase delay of 14.1 cm, or almost 100,000 wavelengths at 1.5 μm wavelength. This is ideal, but will be very challenging. The steering system can be much simpler if each dimension is steered separately because many fewer steering elements are required. Also, of course, one would like very rapid beam steering. If chip-scale approaches to wide angle continuous beam steering are used, chips will tend to be small, requiring magnification in the system.

One of the really interesting aspect of phased arrays, whether optical or RF, is the ability to track to a very small fraction of the diffraction-limited beam width.[166] In the classic 1968 paper on digital phased arrays it is shown that with modulo 2π beam steering one can adjust only one phase modulator in one phase ramp, and it has essentially the same effect as if much smaller modulators were used in every phase ramp. This means one can steer to 1/100th of a diffraction-limited spot size, or even better, using digital phased array steering technology.

Right now one can steer to large angles nonmechanically, but it requires many stages of beam steering. If the polarization birefringent grating approach to beam steering could be made variable and fast, this would be close to the grail for beam steering. If chip-scale optical phased arrays could be made with micron-class pitch and relatively large chips, this could be a viable continuous steering approach to be combined with fixed polarization birefringent gratings. Another viable continuous steering approach that can be used before beam expansion in combination with a stack of polarization birefringent gratings is the waveguide liquid crystal approach to nonmechanical steering.

---

[162] J. Kim, C. Oh, M. J. Escuti, L. Hosting, and S. Serati, 2008, "Wide-angle nonmechanical beam steering using thin liquid crystal polarization gratings," *Proc. SPIE* 7093, Adv. Wavefront Contr.: Methods, Devices, Applicat. VI.

[163] S. Pancharatnam, 1955, "Achromatic combination of birefringent plates," *Proc. Indian Acad. Sci.* XLI (4): 137.

[164] L. Shi, P.F. McManamon, and P.J. Bos, 2008, "Liquid crystal optical phase plate with a variable in-plane gradient," *J. Appl. Phys.* 104: 033109.

[165] P.F. McManamon, P.J. Bos, M.J. Escuti, J. Heikenfeld, S. Serati, H. Xie, and E.A. Watson, 2009, "A review of phased array steering for narrowband electro-optical systems," *Proceedings of the IEEE* 97(6): 1078.

[166] B.R. Hatcher, 1968, "Granularity of phased positions in digital phased arrays," *Proceedings of the IEEE* 56(11).

Pointing and stabilizing to very small angles is an expensive proposition, involving large optics. The newest methods of pointing and stabilization will be seen in the literature. Some very new nonmechanical approaches require lithographic, chip-scale processes.

Some form of beam steering and stabilization is required for active EO sensing. The issues considered here are the advantages and disadvantages of the new nonmechanical steering or other approaches.

The potential for random access, rapid steering is a significant advantage for optical-phased-array, nonmechanical steering and stabilization approaches. The ability to stabilize very precisely is a second very important feature of optical phased array approaches that steer based on writing phases using discrete electronics. Compactness, lower cost, and much higher reliability are all other possible advantages.

The relative immaturity of many nonmechanical approaches is a disadvantage. Also, at this writing, many stages are required to obtain continuous large-angle steering. Efficiency is also still an issue, and speed is another issue with some approaches. This is a developing area.

Active EO sensors need to point both the transmitter and the receiver at the object being sensed. Since active sensors bring their own photons often it is desirable to be cued to a small region to limit required illuminator power. If the transmitted laser beam is relatively wide and high-resolution imaging is desired, then a large receiver FPA is needed. So-called flash imagers use this approach, while scanning imagers use a narrow field of view quickly scanned. Scanned imagers put more stringent requirements on beam steering and stabilization because the beam must be pointed and stabilized to within a fraction of one angle/angle pixel. Beam steering and stabilization can be achieved many ways. Often, laser sensing systems are narrowband, which improves the beam steering options because one does not have to worry about dispersion in the beam steering mechanism. One significant advantage with some non-mechanical beam steering approaches is the ability to point in random directions over the field of regard rather than having to scan through all points between the current pointing location and the desired pointing location.

Europe and the United States both make gimbals and FSMs. Risley prisms do not have a large barrier to entry. In fact, the barriers to entry for almost all the optical phased array steering approaches are relatively small. While the United States is currently ahead in the development of this technology, the lead is fragile.

Conclusion 4-16: Polarization birefringent gratings will gain an increasing share of the beam steering for active EO sensors, especially when combined with flash imaging techniques.

Conclusion 4-17: Steerable electro-evanescent optical refraction prism waveguides will be used for smaller aperture, closer range ladars and to steer the beam before beam expansion in conjunction with polarization birefringent gratings.

THERMAL MANAGEMENT

Detectors

Cryogenically cooled IR detectors require complementary accompanying technologies of refrigeration and thermal isolation in order to produce the sensitivities and performance parameters that are desired. When IR sensors began to move out of the laboratory and into tactical applications in the 1960s, the first sensors to do so relied on adaptations of the laboratory equipment, primarily the refrigerators and vacuum systems. As such, these initial systems were large, heavy, and power hungry, better suited to lab conditions than the rigors of tactical environments. While providing much needed capability in the field, the systems required significant levels of support.

As the capability of the systems increased and the possible applications expanded, concerted design efforts were undertaken to provide cryogenic and vacuum packaging solutions that were better

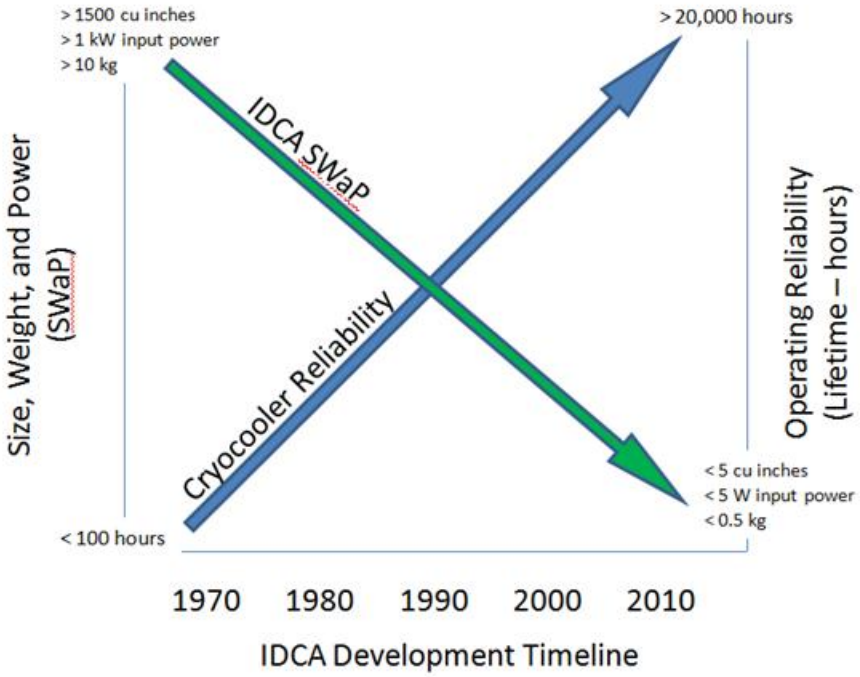

FIGURE 4-52 Evolution of cryocooler and IR detector packaging. IDCA, integrated detector/cooler assembly. SOURCE: D. Rawlings, DRS Technologies, Dallas, Texas.

suited to the tactical environments, were lighter and smaller, used less power, required less support, and offered high reliability. Figure 4-52 illustrates graphically the trend in size, weight, and power (SWaP) and reliability that has resulted from the focus of the user community and the sensor suppliers on removing the barriers to IR sensor adoption.

The first major push in SWaP reduction and reliability increase came with the development of the common module components and the forward looking infrared (FLIR) sensors in the 1970s. The concept broke the sensor down into basic building blocks and separated the cryogenic cooling function from the vacuum insulation packaging function, resulting in two components: the cryocooler and the vacuum Dewar.

Cryocoolers

The majority of cryogenic coolers for infrared sensors are based on the Stirling refrigeration cycle. This refrigeration process relies on pressure fluctuations within a reciprocating gas flow to extract energy from the refrigeration space; no liquid is produced and there is no refrigerant loop as is found in the more common commercial and consumer refrigerators. The working gas for Stirling cryocoolers is most commonly helium, used for its thermal properties. An early common module cryocooler is shown in Figure 4-53.

Cryocooler Classification

Most of the cryocoolers utilized in IR sensor systems can be classified as either rotary or linear drive.

FIGURE 4-53 Common module cryogenic cooler of 1 W.
SOURCE: DRS Technologies, Dallas, Texas.

Rotary Drive

The first of the common module cryocoolers used a rotary frameless AC motor as the primary driver of the compressor and expander pistons. A crank-slider mechanism converted the rotary motion to reciprocating piston action to provide the required pressure fluctuations. As electric motor technology has advanced, the rotary drive mechanisms have evolved and for the most parthave brushless DC motors that utilize high-performance permanent magnet materials to convert the electrical drive to rotary motion. Concurrent with the development of the high-efficiency motors, rotary-drive compressors have seen improvements in the basic compressor mechanisms to reduce wear on the compressor seals, reduce reliance on bearing lubricants, and improve overall reliability.

Linear Drive

In efforts aimed at improving cryocooler reliability for space mission applications, an additional class of cryocooler compressor was developed. These cryocoolers utilized a "linear" drive motor—either a moving magnet or moving coil approach—where the driving mechanism was directly coupled to the compressor piston. No conversion of rotary to linear motion was required. This approach allowed compressors to be designed without the crank-slider mechanisms of the rotary cooler, reducing mechanical complexity, eliminating bearings (and lubricants), and providing a straightforward path for implementation of contact-free "clearance" seals in the compressor. By producing the compressor as two moving pistons, linearly opposed and moving 180 degrees out of phase, virtually vibration-free compressors could be realized. Figure 4-54 illustrates this approach, and Figure 4-55 shows some examples.

Pulse-Tube Cryocooler

A variant of the Stirling cycle that has seen considerable development activity in recent years is the pulse-tube cryocooler. The pulse tube does away with the moving "displacer" component that is found in the standard configuration Stirling cycle cryocooler. By eliminating this moving element, the pulse tube offers reduced mechanical complexity and, hence, improved reliability. The potential wearing out of seals typically found in a displacer is mitigated and the lifetime of such a cooler can be increased. To

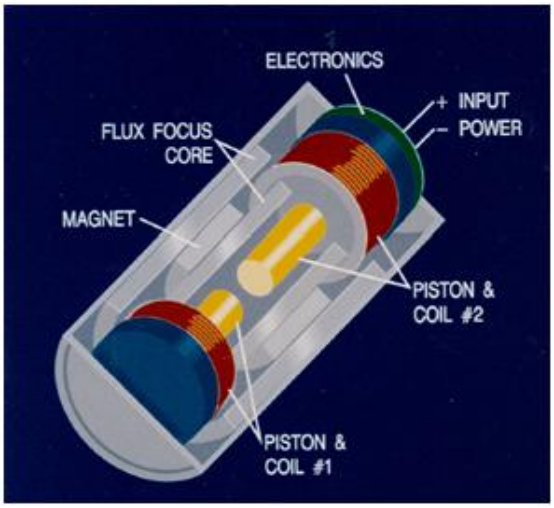

FIGURE 4-54 Linear drive compressor approach.
SOURCE: DRS Technologies, Dallas, Texas.

FIGURE 4-55 A family of linear drive cryocoolers.
SOURCE: DRS Technologies, Dallas, Texas.

date, pulse-tube cryocoolers have not seen application in tactical IR systems—SWaP, robustness in highly dynamic environments, and cost issues have been the primary inhibitors. In strategic applications, where absolute reliability is perhaps an overriding characteristic, pulse-tube cryocoolers are a common choice.

Dewars

Utilizing the IR detector's ability to convert incoming IR photons into electrical signal requires specialized packaging. As the detector needs to be cooled below temperatures at which frost might form, the array must be isolated from ambient air. Additionally, a variety of cooling, optical, and electrical requirements are imposed on the detector package. This detector package is commonly referred to as a Dewar, as it is usually a vacuum enclosure that thermally isolates the array from the environment. Common elements and features of an IR Dewar are shown schematically in Figure 4-56.

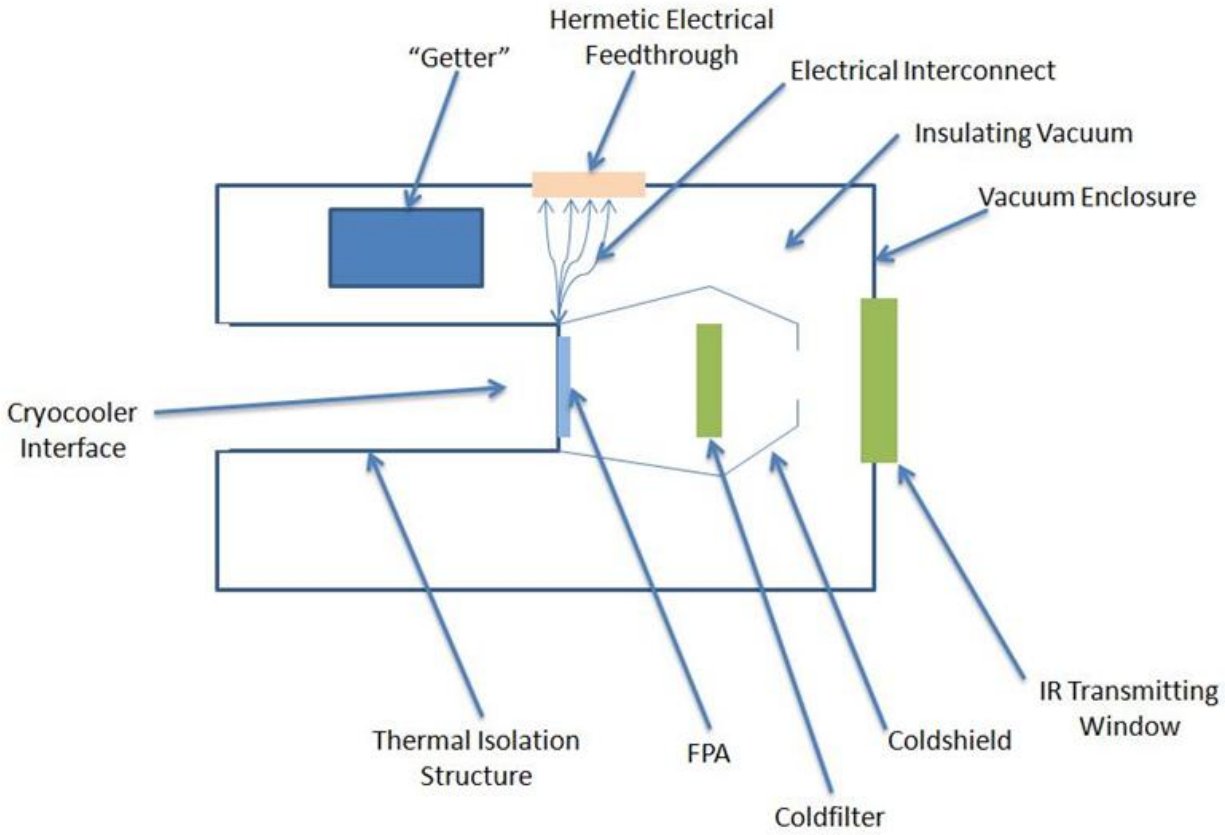

FIGURE 4-56 Components of a Dewar. SOURCE: D. Rawlings, DRS Technologies, Dallas, Texas.

The primary structure of the Dewar is the external enclosure that forms the boundary of the vacuum cell and the internal thermal isolation structure that functions to position the IR detectors. It is coupled to the refrigeration produced by the cryocooler, and the detector is thermally isolated from the external environment. The IR transmitting window serves as a path for the photons coming from the optical system through the vacuum enclosure to the detector. A hermetic electrical feedthrough serves as a path through the vacuum enclosure for the required electrical connections from the system electronics to the detector ROIC.

Inside the Dewar are several additional key components:

- *Coldshield*. The coldshield functions to limit the field of view of the detector to the desired optical bundle. By cooling the shield to temperatures at or near the array temperature, the energy emissions of its internal surfaces toward the detector are greatly reduced. The internal features and surfaces of the coldshield are designed so that stray light, coming from outside the desired optical bundle but coming through the aperture, is absorbed and prevented from impinging upon the detector. Modern coldshields typically eliminate more than 95 percent of the stray light.
- *Coldfilter*. The IR detector is often sensitive to a broad range of IR wavelengths, although the wavelengths of interest may only be a small subset of that broader spectrum. To eliminate the energy outside the region of interest, a coldfilter is often incorporated into the Dewar. The filter is inserted into the optical path and cooled to a temperature approaching the detector temperature. By cooling the filter, the self-emission of the filter element is minimized so that inclusion of the filter does not add another source of noise to the optical system. Coldfilters are fabricated from a variety of substrate materials, most

commonly silicon, BK7 glass, sapphire, and germanium. Key to coldfilter performance is high transmission in the range of interest and high rejection outside that band.

Impact of Operating Parameters

The operating efficiency of a Stirling cycle cryocooler can be defined by a figure of merit called the coefficient of performance (COP). For an ideal Stirling cycle refrigerator, the COP can be determined from the two temperatures at which the cycle is operating:

- T_H, the ambient temperature at which the cryocooler is rejecting heat
- T_C, the refrigeration temperature at which the cryocooler is maintaining the IR detector.

These two temperatures define the COP:

$$COP = \frac{TC}{T_H - T_C}$$

Inspection of this relationship reveals that the COP rises (the cryocooler efficiency increases) as the refrigeration temperature is increased and/or the ambient temperature at which the cooler is rejecting heat is decreased. For a fixed ambient temperature, this is illustrated in Figure 4-57.

This characteristic of the Stirling cycle cryocooler is being exploited by IR detector developers as they refine their detector arrays to operate at warmer and warmer temperatures. By increasing the operating point of a detector from 77 K to 110 K, as an example, the COP of a Stirling cryocooler operating in a 23°C ambient rises from 0.35 to 0.55, an increase of 68 percent. This increase in efficiency results in a corresponding drop in required input power.

In addition to the increased cycle efficiency brought about by the increase in detector operating temperature, the reduced temperature differential between the detector and the ambient temperature results in a reduced parasitic heat leak from the IDCA environment through the detector packaging (the Dewar), thereby reducing the refrigeration demand placed on the cryocooler. This effect is also illustrated in Figure 4-57.

This combination of increased efficiency and reduced refrigeration demand results in less power being dissipated by the cryocooler, which in turn reduces the heat that must be managed by the system integrator.

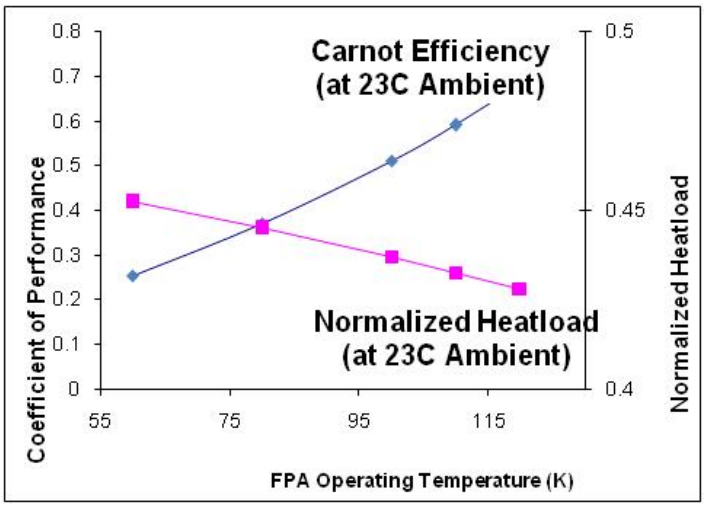

FIGURE 4-57 Impact of FPA operating temperature. SOURCE: D. Rawlings, DRS Technologies, Dallas, Texas.

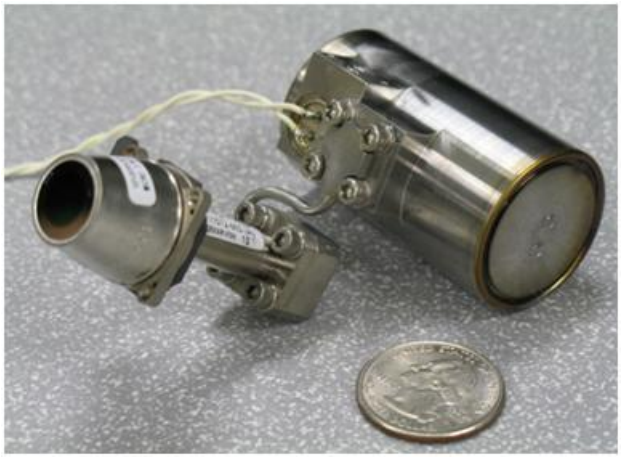

FIGURE 4-58 Advanced integrated Dewar cooler assembly. SOURCE: DRS Technologies, Dallas, Texas.

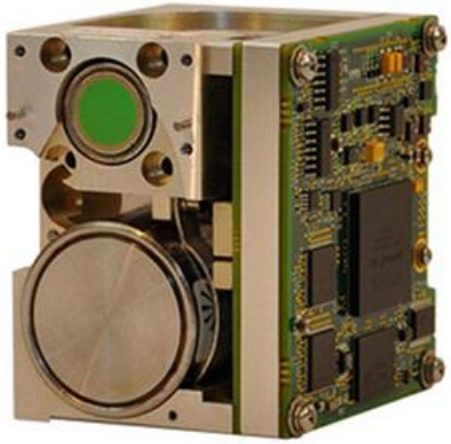

FIGURE 4-59 High-operating-temperature IDCA camera core. SOURCE: DRS Technologies, Dallas, Texas.

IDCA designers have taken advantage of these increasing array operating temperatures to develop lighter and lower power IDCAs. IDCAs with 640 × 480 pixel mid-wave IR (MWIR) detectors operating on as little as 2 W of input power at room temperature are being incorporated into IR sensor systems. One such IDCA is shown in Figure 4-58. Figure 4-59 shows a complete camera core based on a version of that IDCA. This module is about 2 × 2 × 2 inches, including the video processing electronics and runs on less than 5 watts of input power.

Conclusion 4-18: Cooling detector arrays such as those required for HgCdTe avalanche photodiodes operating in the mid-wave infrared band are not a major impediment from a size and power consumption standpoint. Already, state-of-the-art Dewar-cooler technologies, particularly those based on linear drive technology, are getting as small as 5 × 5 × 5 cm and power consumption of a few watts, comparable to those of thermoelectric coolers used for cooling detector arrays to 230 K. MEMS-based coolers are under development that will further reduce the size and power requirements.

TELESCOPES

Telescopes are often used on both transmit and receive. A single aperture for both transmit and receive, called a monostatic system, may be used. In a monostatic system, an approach must be developed for T/R isolation. If near-field reflections feed back into the receiver, it can be damaged or blinded. Alternately, in a bistatic system, separate apertures are used for transmit and receive. Bistatic systems do not have much of a T/R isolation issue since the apertures do not share common optics. Near-field reflections in theory could cause an isolation issue. For image plane imaging, optics are needed to transfer the captured field from the pupil plane to a focus at the image plane. Apertures can be various sizes, depending on need. No optics are required for pupil plane imaging, at least in theory. The received signal can be captured directly without using any telescope. In reality a telescope may be used to magnify the effective size of the receive focal plane array. While there is no need in this report to discuss the various types of telescopes, there is value in discussing the idea of arrays of subapertures. That approach is considered in the MIMO section in Chapter 3.

ADAPTIVE OPTICS

As will be described in more detail in Chapter 5, the atmosphere affects active EO sensors in three main ways: absorption, scattering, and refractive index variations (which can cause beam spreading or fluctuations). For active EO sensors these effects occur both for the illumination beam and for backscatter from the object being viewed by the receiving sensor. This section will describe methods to correct for distortion in active imaging caused by variations in the index of refraction of the atmosphere.

The effects of atmospheric turbulence are discussed in Chapter 5; they can be described as either thin turbulence or extended turbulence (see Figure 5-11). Adaptive optics systems can provide phase compensation for thin turbulence and some compensation for extended turbulence.

Adaptive Optics Systems

If the receive optics for a sensor are larger than the Fried parameter r_0 of the atmosphere,[167] an adaptive optics mirror can be put in the receive path to compensate for phase distortions. If the transmit optics are larger than the r_0 of the atmosphere, again adaptive optics can be used to compensate and therefore to form a beam with less divergence. There are two main parts to any adaptive optics system. One part is the device that actually imposes the phase shift. This can be a mirror with many actuators behind it, or it can be a liquid crystal device or some other device to modulate phase. Adaptive optics mirrors may use electro-strictive (lead magnesium niobate, or PMN) actuators, piezoelectric-driven actuators, or other approaches. The second part is the control system to decide what phase shifts to impose. To determine what phases need to be imposed, the incoming signal is first measured. While multiple measurement instruments may be used, the most popular is the Shack Hartman sensor, shown in Figure 4-60. A second method of determining the required phase shift is a metric that judges the quality of the compensation. One such approach is called stochastic parallel gradient descent, SPGD, where increments toward an ideal metric are chosen along a gradient.

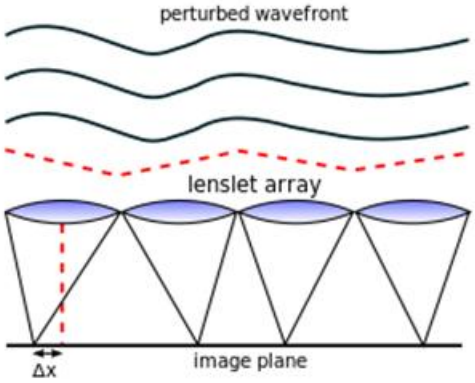

FIGURE 4-60 Shack Hartman wavefront sensor. SOURCE: By 2pem (Own work) [CC-BY-SA-3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons. http://en.wikipedia.org/wiki/Adaptive_optics, downloaded May 17, 2013.

[167] The Fried parameter r_0 is a measure of the effects on the quality of optical transmission through the atmosphere attributable to random inhomogeneities in its refractive index. It is defined as the diameter of a circular area over which the root-mean-square aberration due to passage through the atmosphere is equal to 1 radian. As such, imaging from telescopes with apertures much smaller than r_0 is less affected by atmospheric inhomogeneities than by diffraction due to the telescope's small aperture. However the imaging resolution of telescopes with apertures much larger than r_0 (i.e., all professional telescopes) will be limited by the turbulent atmosphere, preventing the instrument from approaching the diffraction limit.

Another issue for an adaptive optics system may be having a point source to use as a guide for finding the optimum phase corrections. Sometimes an artificial point source is created, called a guide star.[168]

Schemes for Countering Extended Turbulence

Extended turbulence phase changes away from the pupil plane will result in amplitude changes at the pupil plane. Adaptive optics mirrors do not exist that change both phase and amplitude. When amplitude is changed it only decreases, creating a loss to be avoided. An alternative method of compensating for extended turbulence can occur for coherent imaging, whether it is spatial or temporal heterodyne. Once the field is captured at the pupil plane, virtual phase screens can be inserted in the computer at various ranges, and a sharpness metric can be used to judge the influence of the phase screens. These phase screens can be changed until an image with the best sharpness is reached.

One limit is the amount of beam spreading that can be tolerated for the illuminator beam. If flash illumination is used, then this is not a difficult limit. Assume even a 32 × 32 pixel FPA on receive. This means the illuminator beam can be 32 times larger in angle than a pixel, and the illuminator aperture can be 32 times smaller than the receive aperture. For a 20 cm receive aperture the transmit aperture can be less than 1 cm. Another way to think of this is that the Fried parameter can be less than 1 cm without interfering with the illuminator beam.

Another limit is the turbulence on receive. This will limit angular resolution. The diffraction on receive will be limited to the diffraction associated with an aperture diameter equal to the Fried parameter unless some form of compensation is employed. If an adaptive optics device were invented that could not only impose an array of phase changes on an incoming beam, but could also amplify or decrease the signal level in each pixel, then it would be possible to physically compensate for extended turbulence.

Conclusion 4-19: Any long-range active EO sensor will be limited by the atmosphere. The amount of limitation will vary from hour to hour and day to day. Atmospheric compensation can extend the operational range.

Conclusion 4-20: The ability to compensate for extended turbulence will greatly aid long-range active EO sensors.

PROCESSING, EXPLOITATION, AND DISSEMINATION

Processing, exploitation, and dissemination (PED) refers to the set of steps required to transform sensed photons into usable information. The processing step refers to the initial transformation of sensed photons into an image that, in turn, can be exploited in the context of an application such as robot navigation. The dissemination step refers to delivery of the information to its ultimate consumer(s)—for example, analysts seeking to understand an environment using change detection on a time series of images.

There may be multiple types of processing involved, for example, image formation from ROICs (see Figure 4-61), followed by image compression to reduce bandwidth required for real-time dissemination of imagery from an airborne sensor. Table 4-5 provides some PED requirements for some selected ladar applications.

[168] See http://en.wikipedia.org/wiki/Laser_guide_star, download May 17, 2013.

TABLE 4-5 PED Requirements Versus Selected Application Domains

Application	Possible platform	Processing	Exploitation	Dissemination
Surveillance	UAV	Image formation and refinement, image/video selection and compression	Precise positions for objects and actors in real-time	High bandwidth desirable, but not easily achievable with RF
Mapping	Airborne imager	3-D image formation, registration to knit multiple swaths together	Precise 3-D maps of terrain, e.g., urban terrain	Can be gathered and stored until return to base
Local navigation	Driverless automobile	Image formation	Avoiding obstacles in real time	Not applicable; intended for local use

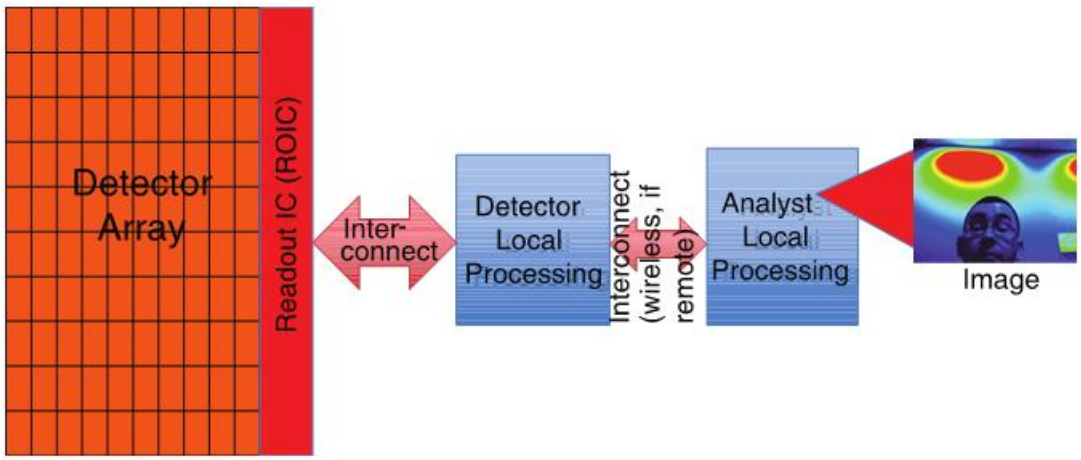

FIGURE 4-61 Abstract model of steps from detector to analyst. SOURCE: National Research Council, 2010, *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*, The National Academies Press, Washington, D.C., Figure 4-9.

Target Tracking

Target tracking is a general technique useful in a variety of application domains. An easy to understand example is tracking of pedestrians and other vehicles in autonomous vehicle navigation. Because ladar is much more precise than radar for this application, it could generate reliable trajectories.

Target tracking[169] first involves detection and acquisition. Once a target is acquired, the illuminator must somehow be periodically directed toward the target as its direction and velocity are tracked, to guarantee that sufficient reflected information is available for real time track maintenance.

[169] C. Grönwall, G. Tolt, D. Karlsson, H. Esteki, J. Rydell, E.E. Armstrong, and J. Woods, 2010, "Threat detection and tracking using 3D FLASH ladar data," *Proc. SPIE* 7696: 76960N.

Various processing schemes can be employed, from tracking classified objects (object classification is discussed in the section below) to simply tracking centroids of moving objects.[170]

Target Classification

Ladar target classification basically consists of image matches against a set or sets of known characteristics. The goal is often not an exact match, which would be difficult, but rather a best match or a match within a known threshold or error bound. A commercial use might be pedestrian detection for an autonomous vehicle.

Ladar target classification is an active area of research, with a variety of recent advances reported in the technical literature—for example, using ensembles of classifiers[171] or using machine learning techniques such as support vector machines(SVM).[172]

Developing 3-D Maps

Ladar has found a variety of mapping applications, some of which exploit the ability of the technology to produce 3-D imagery. Examples include archeology,[173] forestry,[174] and geology.[175] Mapping is typically performed as an aerial surveying activity; data are collected and partially processed, then stored on the aircraft. The recorded survey data are then post-processed; e.g., swaths are combined into a map and then analyzed.

3-D Target Metrics

For monochromatic 2-D imagery the National Imagery Interpretability Rating Scale, NIIRS, rating is used extensively to characterize image quality. Leachtenauer et al.[176,177] empirically determined a method to calculate the NIIRS rating using various image metrics, such as the ground sample distance (GSD) and the relative edge response (RER).

---

[170] P. Morton, B. Douillard, and J. Underwood, 2011, "An evaluation of dynamic object tracking with 3D LIDAR," Australasian Conference on Robotics and Automation (ACRA), Melbourne, Australia: Australian Robotics and Automation Association (ARAA).

[171] Z.-J. Liu, Q. Li, Z.-W. Xia, and Q. Wang, 2012, "Target recognition for small samples of ladar range image using classifier ensembles," *Opt. Eng.* 51(8): 087201.

[172] Z.-J. Liu, Q. Li, and Q. Wang, 2013, "Random subspace ensemble for target recognition of ladar range image," *Opt. Eng.* 52(2): 023203.

[173] F. Diep, 2012, "How lasers helped discover lost Honduras city," *TechNews Daily*, June 7, http://www.technewsdaily.com/5837-lasers-helped-discover-lost-honduras-city.html, accessed May 30, 2013.

[174] R.A. White and B.C. Dietterick, 2012, *Use of LiDAR and Multispectral Imagery to Determine Conifer Mortality and Burn Severity Following the Lockheed Fire,* U.S. Forest Service, General Technical Report PSW-GTR-238, pp. 667-675, http://www.fs.fed.us/psw/publications/documents/psw_gtr238/psw_gtr238_667.pdf.

[175] R.A. Haugerud, D.J. Harding, S.Y. Johnson, J.L. Harless, C.S. Weaver, B.L. Sherrod, 2003, "High-resolution lidar topography of the Puget Lowland, Washington," *GSA Today*, Geological Society of America, June.

[176] J.C. Leachtenaur, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, 1997,"General image quality equation: GIQE," *Applied Optics* 36(32): 8322.

Leachtenauer et al. later used the same empirical methodology applied to infrared images.[178] Thurman and Fienup provided physical explanations for many of the parameters in the General Image Quality Equation (GIQE).[179] Equation 1 is a generic version of the GIQE. Various constants can be used in this equation, representing somewhat different forms of the GIQE.

$$NIIRS = c_0 + c_1 Log_2(GSD) + c_2 Log_2(RER) + c_3 \frac{G}{SNR} + c_4 J,$$

where *NIIRS* is the image quality rating assigned in accordance with the National Image Interpretability Rating Scale, *RER* is the relative edge response, *J* is the mean height overshoot caused by edge sharpening, *G* is the noise gain resulting from edge sharpening, and *SNR* is the signal to noise ratio. Recently Kammerman wrote a draft of a paper developing an information theoretic approach to calculate the NIIRs rating for monochromatic 2-D images, as compared to an empirical method.[180] Ideally, this information-based theory would be a better predictor of NIIRS ratings for monochromatic visible imagery and would also provide a path to generate useful metrics for other imaging modalities, such as IR imagery, radar, and 3-D ladar imagery. At this time, there are no good 3-D metrics allowing a comparison of various 3-D imaging sensors, much less allowing them to be compared against other sensing modalities.

Conclusion 4-21: It would be very useful to be able to compare the value of various 3-D images, especially if they could also be compared against other sensing modalities.

Recommendation 4-1: Three-dimensional metrics should be developed in such a way that the ability of a given sensor to perform a given function can be compared against the ability of another sensor to perform the same function. 3-D imagery can be developed in multiple ways: for example, using stereo from passive sensors vs. using a 3-D active EO sensor, or using 3-D ladar vs. interferometric synthetic aperture radar. Metrics are needed to be able to compare the products of different sensing modalities.

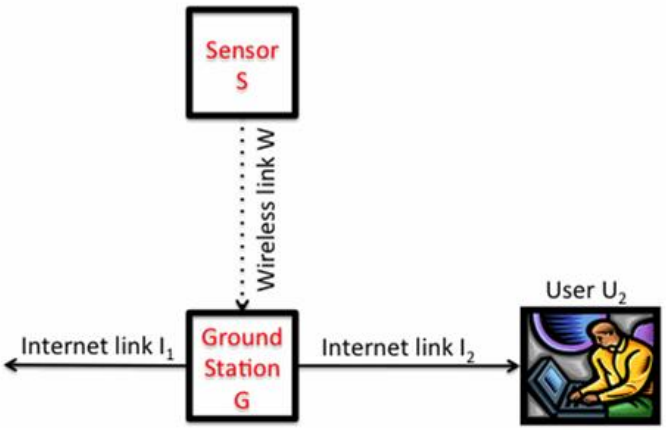

FIGURE 4-62 Communication model for ladar used for ISR on a UAV.

[178] J.C. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, 2000, "General image-quality equation for infrared imagery," *Applied Optics* 39(26): 4826-4828.

[179] S.T. Thurman and J.R. Fienup, "Analysis of the general image quality equation," Proc. of SPIE 6978 69780F-1.

[180] G.W. Kamerman, "On Image Information Density," *Optical Engineering* to be published, 2013.

Data File Size, Compression, Dissemination, and Communication Bandwidth Requirements

A given sensor produces data at some rate, and then the application domain dictates whether data is stored or communicated. Whether stored or communicated, it may be compressed to a more compact form, ideally with minimal loss of information content. One reason for compression of stored data might be to make better use of available storage—for example, to store mapping data for a greater area of coverage or to extend mission duration. As Table 4-5 suggests, navigation of completely autonomous vehicles faces very little in the way of PED challenges, as the data is both produced and consumed locally by the navigation subsystem of the vehicle.

On the other hand, as discussed in an earlier NRC report *Seeing Photons*,[181] unmanned aerial vehicles (UAVs) used for intelligence, surveillance and reconnaissance (ISR) may represent the most challenging ladar application for communications, as the data are meant to be consumed at a considerable distance from where they are gathered; an abstract view of this application is shown in Figure 4-62.

Since the Wireless Link W is typically highly constrained relative to the sensor capabilities, compression is often employed between the sensor and the ground station. There are many forms such compression can take, including sampling, lossy compression, image compression, video compression, and compressive sensing. There is an interesting tradeoff between placing more computational intelligence in the sensor to provide sensor-side analysis and the SWaP considerations of the airborne sensor.

Fusion or Synergy with Other Sensing Modalities

Ladar is extremely powerful as a complement to other sensing capabilities. For example, some combination of GPS, radar, video cameras, and ladar was used by almost all entrants in the DARPA Grand Challenges and the Urban Challenge to maximize the operating capabilities of the self-driving vehicles. The absolute frame of reference for navigating the vehicle could be determined to a reasonable degree of precision with differential GPS, but obstacle detection is best performed with one of the other sensor technologies.

Conclusion 4-22: Ladar may be most powerful when viewed as part of a system of complementary sensors rather than as a stand-alone multifunction sensor.

Conclusion 4-23: Use of ladars may be most cost-effective where they offer sensory capabilities (such as precision stand-off ranging and vibrometry) that are difficult for other sensors to achieve.

Computational Requirements, Processing Throughput, and Processor SWaP

Figure 4-63 shows the data processing steps associated with the ALIRT 3-D ladar system discussed in Chapter 2. One set of steps[182] in image formation from Jigsaw[183] data are illustrated in Figure 4-64. Each of these steps can consume substantial computing time (as shown in Table 4-6), which affects the total delay from the detection to the image or other exploitable data product.

---

[181] National Research Council, 2010, *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*, The National Academies Press, Washington, D.C.

[182] P. Cho, H. Anderson, R. Hatch, and P. Ramaswami, 2006, "Real-time 3D ladar imaging," *Lincoln Laboratory Journal* 16(1): 147.

[183] R.M. Marino and W.R. Davis, 2006, "Jigsaw: a foliage-penetrating 3D imaging laser radar system," *Lincoln Laboratory Journal* 15(1): 23.

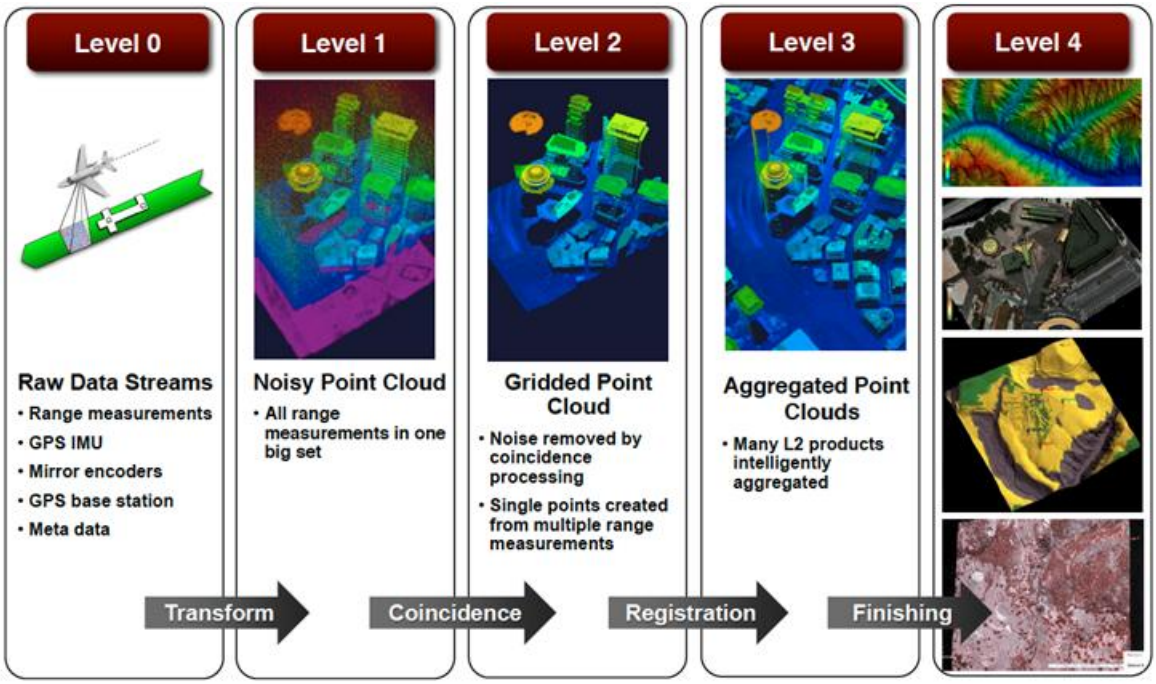

FIGURE 4-63 ALIRT data processing steps. SOURCE: Dale G. Fried, 2013, "Photon-counting laser radar development at MIT Lincoln Laboratory," April 24. Courtesy of MIT Lincoln Laboratory.

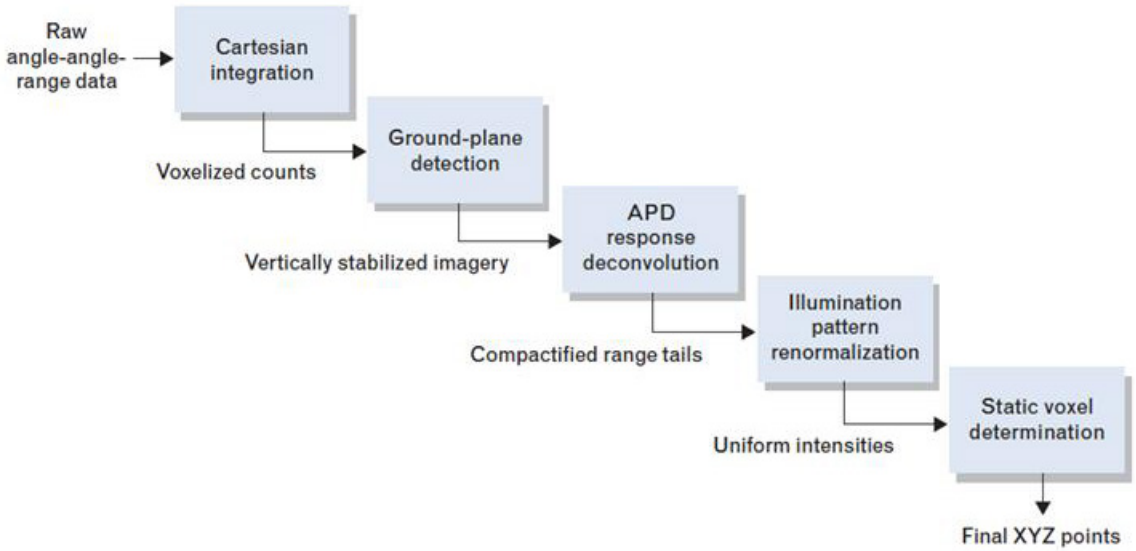

FIGURE 4-64 Steps in processing Jigsaw data. SOURCE: P. Cho, H. Anderson, R. Hatch, and P. Ramaswami, 2006, "Real-time 3D ladar imaging," *Lincoln Laboratory Journal*, 16(1): 147. Reprinted with permission of MIT Lincoln Laboratory.

Since the goal of this system is real-time 3-D imagery, 8.2 s of compute time per second of the data stream is unacceptable. Therefore, additional computing resources are brought to bear in a mode of computing called parallel processing, where multiple computations are carried out simultaneously. Fortunately, the algorithms allow parallel operation, and a set of 10 computers, as illustrated in Figure 4-65, can be employed to reduce the computation time to enable real-time operation.

TABLE 4-6 Computing Resources Consumed per Second of Imagery

Algorithm Task	Processing Time/Real Time
Raw data input	0.1
Cartesian integration	3.0
Ground detection	0.2
Response deconvolution	3.2
Static voxel determination	1.1
XYZ file output	0.7
Total	8.2

NOTE: Single processor timing results measured on a 3 GHz Pentium machine.
SOURCE: Peter Cho, Hyrum Anderson, Robert Hatch, and Prem Ramaswami, 2006, "Real-time 3D ladar imaging," *Lincoln Laboratory Journal*, 16(1): 147. Reprinted with permission of MIT Lincoln Laboratory.

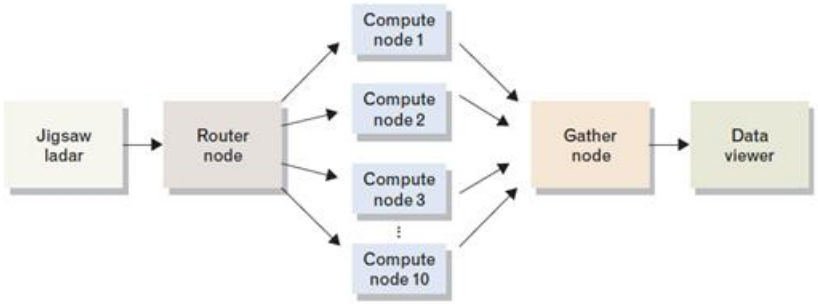

FIGURE 4-65 Parallelization of computation in Figure 4-64. SOURCE: Peter Cho et al., op. cit. Reprinted with permission of MIT Lincoln Laboratory.

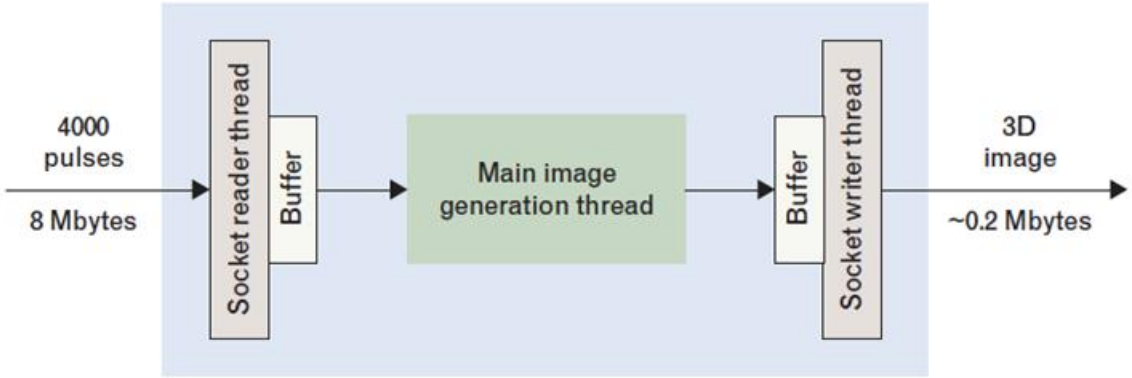

FIGURE 4-66 Data reduction in a compute node from Figure 4-65. SOURCE: Peter Cho et al., op. cit. Reprinted with permission of MIT Lincoln Laboratory.

An additional benefit of some steps is that they reduce the data rate through the system to a rate that is more sustainable by serial processing techniques, in preparation for delivery of imagery, as shown in Figure 4-66.

Currently, data from high-performance ladar sensors requires parallel processing. Processing capabilities with abundant parallelism, such as many core graphics processor units (GPUs), are commercially available, but SWaP requirements for UAVs are not a commercial consideration. DARPA's Power Efficiency Revolution for Embedded Computing Technologies (PERFECT) Program[184] has UAVs as a target application and seeks methods to move from the current energy performance level of 1 GFLOP[185]/W to greater than 75 GFLOPS/W. Interconnect architectures for some abundantly parallel machine architectures may be ill-suited to processing some sensor streams, but for others the performance gains may be substantial.

[184] See http://www.darpa.mil/Our_Work/MTO/Programs/Power_Efficiency_Revolution_for_Embedded_Computing_Technologies_(PERFECT).aspx.

[185] 1 GFLOP is 10^9 floating point operations per second.

5

Fundamental and Engineering Limits of Active Electro-Optical Sensing

ILLUMINATION SOURCES

There are a small number of what might be called truly "fundamental" limits to the performance of sources, and these involve the basic physics concept of energy conservation. While physics is involved in all other types of limits, the concepts here are more in the engineering category, involving, say, properties of the materials involved and thermal management of the devices. As such, these are not hard "limits" and may change with further development efforts. Some have resisted the effort that has gone into overcoming them, such as the lack of suitable semiconductor materials to provide laser operation over a good portion of the visible wavelength range or the lack of nonlinear materials with high transparency in the ultraviolet to allow generation of power in that region. Others, such as electrical conversion efficiencies for diode lasers and limits to the (average and peak) power outputs for solid-state lasers are undergoing continuing improvement. Appendix C contains several tables that are a snapshot in time for many of these "soft limits" (Tables C-5 to C-12), summarized below.

Solid-State Lasers

Bulk Format

Solid-state lasers require optical pumping; the most efficient method would be to use a narrow-line source at a suitable absorption frequency larger than the laser's, but as close as possible. The ratio of the energy of the laser photon to the pump photon is called the quantum defect. The theoretical efficiency for the 1,064 nm Nd:YAG laser line pumped by a laser diode at 808 nm is 76 percent. The theoretical best efficiency of the laser-pumped laser is then the product of the quantum defect and the pump laser efficiency. Table C-5 shows that pump diodes with 60 percent efficiency at 808 nm give a fundamental limit to the electrical efficiency for this Nd:YAG laser of 46 percent. In reality, such lasers demonstrate 20-25 percent efficiency, only half of the fundamental limit due to multiple engineering factors.

The maximum power out of solid-state lasers in the 1,000 nm wavelength region seems to have an engineering limit at the present time of ~ 2 kW in a single rod, 10 kW in a single disk, and 15 kW in a single slab. Higher powers are possible with multiple active media; the present record is >100 kW. Solid-state lasers at longer wavelengths are limited by multiphonon decay, but low-phonon hosts tend to be impractical for high-power operation.

The fundamental Schalow-Townes linewidth for lasers is only a few hertz and can be reached by active external stabilization. Otherwise, fluctuations in optical cavity length cause technical noise on the order of kilohertz owing to coupling between the pump power and gain medium refractive index, acoustic noise, and other cavity perturbations. If the laser is not stabilized, slow drifts of 10-50 MHz occur as a result of changing environmental conditions.

Mode-locked lasers have pulses that are given by the inverse bandwidth of the gain, with a repetition time equal to the round-trip time of the cavity. For Ti:sapphire, the theoretical pulse width is about 3.5 fs; experimentally the pulse width is about 25 percent larger because of dispersion in the optical

cavity, the finite spectral response of mirrors, and nonlinear effects. The calculated pulsewidth for the mode-locked, eye-safe Cr:ZnSe laser at 2,500 nm is 7 fs, while experiments report much longer pulsewidths (60 fs). Nd:YAG at 1,064 nm has a fundamental mode-locked pulsewidth of 2 ps.

Fiber Format

The quantum defect rate can be as low as 5 percent for a 1,030 nm Yb:fiber laser pumped at 976 nm. Practical limits to efficiency of diode-pumped fiber laser are due to poor spatial overlap of pumping and lasing regions in the core, incomplete absorption of the pump, losses in the laser material, excited state absorption of the pump or laser power, and upconversion from the upper laser level. In practice, 88 percent slope efficiency has been reported for Yb:fiber lasers. When the pump diodes have 65 percent efficiency, the Yb:fiber laser should have a fundamental limit of 62 percent electrical efficiency, but in reality it has an efficiency of 49 percent, mostly because pump light is lost in transport from diode facet to pump cladding.

Diode-pumped single fiber lasers have the ability to produce very high average powers in the 1000 nm wavelength region, limited mostly by stimulated Raman emission. A 1,030 nm Yb:fiber laser pumped by multiple 1,018 nm Yb:fiber lasers has produced 20 kW. In the eye-safe 2000 nm wavelength region, a single Tm-fiber laser can produce 1 kW when pumped by 790 nm diode lasers. A single-frequency, single fiber in the 1,000 nm region can put out 100 W with 5 kHz linewidth compared to 1 kW with 3 GHz linewidth. In both cases stimulated Brillouin scattering had to be overcome; the former used a thermal gradient along the length of the fiber; the latter used a frequency modulated source. In the eye-safe 2000 nm region, the average power in a single frequency from a single Tm-fiber laser is 600 W in a single frequency with < 5 MHz linewidth.

Q-switching a straight Yb:silica rod-type fiber can give 60 ns pulses of 0.45 MW peak power at 1030 nm (27 mJ) or a shorter pulse <1 ns with a peak power of 4.5 MW (4.3 mJ). Both regimes are limited by stimulated Raman scattering and optical breakdown. Flexible fibers typically generate less peak power: 25 kW in 400 ns pulses. The highest peak power comes from mode-locked pulses: Yb:silica laser pulses can be only 0.5 ps long, with a peak power from a single fiber of 3.8 GW (2.2 mJ). Such short pulses would be limited by self-phase modulation, as well as the mechanisms that affect Q-switched pulses, so to achieve this high power while avoiding nonlinear effects, the rod-type fiber system employs chirped-pulse amplification (CPA). Flexible fibers with CPA operate up to around 100 MW of peak power.

Nonlinear Optics

Harmonic generation can, in principle, be 100 percent effective for plane waves. In practice, many factors can reduce this efficiency: dephasing from beams of finite width; dephasing as a consequence of heating in the crystal from background absorption or multiphoton absorption and creation of color centers; losses at the entrance and exit faces of crystal; crystal or quasi-phase matching (QPM) material imperfections; and optical damage at material surfaces or in the bulk. All of these factors can limit second harmonic generation to 60-70 percent for Gaussian beams, though it is up to 90 percent for the high-energy, flat-top beams of the National Ignition Facility. Maximum third harmonic from Gaussian beams, theoretically, is 50 percent. The maximum amount of power generated is limited by the same phenomena that limit the efficiency. Lithium borate (LBO) has the lowest absorption of common nonlinear materials, and the powers are ultimately limited by damage in the coatings. Multiple hundreds of watts of second harmonic can be created from Nd- or Yb-doped lasers. The shortest wavelength achievable by a second harmonic is 176 nm, determined by phase-matching limits and vanishing nonlinear coefficients at short wavelength. This wavelength can be reached with KBBF crystals; the more readily available BBO crystals generate second harmonic out to about 205 nm.

Optical parametric oscillators OPOs), in principle, have a fundamental limit for overall conversion efficiency from pump to signal plus idler of 100 percent. The OPO is, in principle, a lossless process. It breaks up one pump photon of energy hn_p into two photons, hn_s and hn_i, where $hn_p = hn_s + hn_i$. The maximum efficiency of conversion from the pump to the signal is hn_s/hn_p. Similarly the maximum efficiency of conversion from the pump to the idler is hn_i/hn_p. These are called the Manly-Rowe relations and assume plane waves. They show that the sum of the power in the signal plus idler should equal the pump power if there are no extraneous losses. The actual efficiency can be 90 percent for continuous wave (CW) OPOs, while 50 percent is more typical for pulsed lasers. For pulsed OPOs, the finite buildup time of the parametric process in the cavity leads to further loss in efficiency. Often loss in conversion efficiency results from some light not being converted away from the pump wavelength.

While theoretically there is no limit to the power than can be handled by an OPO, the actual power is limited by the same processes that limit the efficiency of second harmonic generation. Typical OPO outputs are presently below 100 W but have been limited primarily by the power available in pump lasers. Hundreds of watts should be possible in the near-IR with materials like LBO.

DETECTORS

The smallest light packets, photons, carry information that detectors attempt to extract. How much information carried is an active research topic.

> The photon, the indivisible unit of electromagnetic energy, is a fundamental carrier of information. Numerous degrees of freedom are available for the conveyance of information on a photon including frequency, phase, arrival time, polarization, orbital angular momentum, linear momentum, superposition states, correlation, entanglement, etc.[1]

Active electromagnetic applications such as ladar extract more information than intensity from photons by including arrival time, wavelength (frequency) and sometimes polarization and phase in the classical regime.

Detector Fundamental Limits

Important parameters that characterize the fundamental properties of detectors are discussed next.

- Responsivity (Amperes/Watt). Responsivity is the ratio of the generated photocurrent to incident light power. The ideal would be for all the power to produce electron-hole pairs and for none to be wasted as heat or through other loss mechanisms. For detectors with gain, this parameter can be large, but gain also introduces excess noise (see discussion of gain and excess noise in Chapter 4).
- Signal to Noise (SNR). SNR is the ratio of the signal current to the root-mean-square average of the noise current. There is no ideal limit, but higher is better.
- Specific Detectivity (D*). Specific detectivity is used to compare detectors with different active areas and electronic bandwidths. It is defined as

$$D^* = \sqrt{A_d \Delta f} \times \text{SNR/Flux}$$

where A_d is the detector area, Flux is the incident photon power, and Δf is the bandwidth. This expression assumes the noise is constant white noise across Δf. Some detectors approach ideal performance when the

[1] See http://www.darpa.mil/Our_Work/DSO/Programs/Information_in_a_Photon.aspx.

wavelength of the incident photons nears the detector cutoff. The ideal is when the photons have just enough energy to create carriers by exciting them over the bandgap and there is no lost excess energy.

- Quantum Efficiency (percent). For detectors without gain, quantum efficiency (QE) is the ratio of the number of carriers to the number of incident photons, with 100 percent being ideal. For detectors with gain, the ratio can be large and is usually limited by voltage controls.
- Dark Current (electrons/second). Dark current is the current in the absence of light when the detector is operated in photoconductive mode. Dark current has unavoidable shot noise. For semiconductors, dark current decreases as bandgap increases. The ideal is zero dark current, but that only occurs at 0 K.
- Response time. Photon absorption in detectors produces carriers (electron-hole pairs) that generate a current under influence of electric fields. The sum of the electron and hole transit times determines the minimum response time. The detector resistance, $R_{d,}$ and capacitance, $C_{d,}$ add to the response time, $T = R_dC_d$. External circuitry adds additional RC delay in the response time. The response time lengthens the impulse response of the detector. The ideal response time for a detector with zero resistance is the detector width divided by the speed of the carrier's ballistic transport velocity, which is the speed of carrier transport without scattering and limited only by the carrier effective mass, a function of the band structure.

To highlight important detector fundamental parameters, a simplified analysis of passive detectors without detector gain is presented below. The model shows the limiting effects of several detector parameters but, as noted later, even with these constraints, systems with great utility are available. Better detectors will expand the EO system application options, but improved detectors are only one factor enabling improved systems.

To identify an object, a detector must distinguish the object's signal from the noise that comes from multiple sources. A key measure of performance is therefore the detector SNR. SNR is the ratio of object signal to the uncertainty of all the other contributions to the signal measurement. Higher SNR is better. The desire for better SNR drives much of the development of electro-active imaging systems, components, and signal processing. The goal of this analysis is to elucidate the pure characteristics of detectors without the many application- and system-specific factors such as atmospheric transmission, collection apertures, external readout noise, scene spatial noise and clutter, and like factors that are important for application success but not directly related to detector performance. At the end of this analysis is an example of a complex model that includes many application-specific parameters.[2] The more realistic complex model predicts performance (SNR) for the various combinations of detector focal plane arrays (FPAs), optics, and spectral target and scene characteristics. The model shows that detectors are only one of many aspects of overall system performance.

Signal

Detector signal is produced from the incident flux, *F*, (J/s) of photons multiplied by the photon to electron QE and integrated over time t.[3] Quantitatively, the signal *S* is

$$S = \frac{\mathrm{q}}{h\upsilon} FtQE = qtQE * (photon\ arrival\ rate)$$

where hυ is the energy of the photon (Planck's constant × frequency) and q is the electron charge. In addition to signal, there also is corrupting noise signal.

[2] See Figure 5-4.

[3] Not included are factors such as system aperture size and atmospheric transmission that affect system performance but are not exclusively related to the detector.

Noise Sources[4]

Incident Flux Shot Noise

Usually photons are random in nature,[5] described by Poisson statistics, and have arrival variations called shot noise. This noise in not related to any flaws in a detector and is unavoidable. Detector signal noise can never be less than flux shot noise, but some detectors operate near the shot noise limit[6] (see Figure 5-1), and there are ways to operate detectors to reduce other noise for the photon sources.[7] The performance for three commercial detectors compared to the ideal shot noise limit is seen in Figure 5-1, where it is apparent that the performance advantage produced by the high-gain Geiger-mode avalanche photodiode (GM-APD) detector[8] approaches the shot noise limit over much of the irradiance range, while the intrinsic noise of the detector degrades SNR for low light levels.

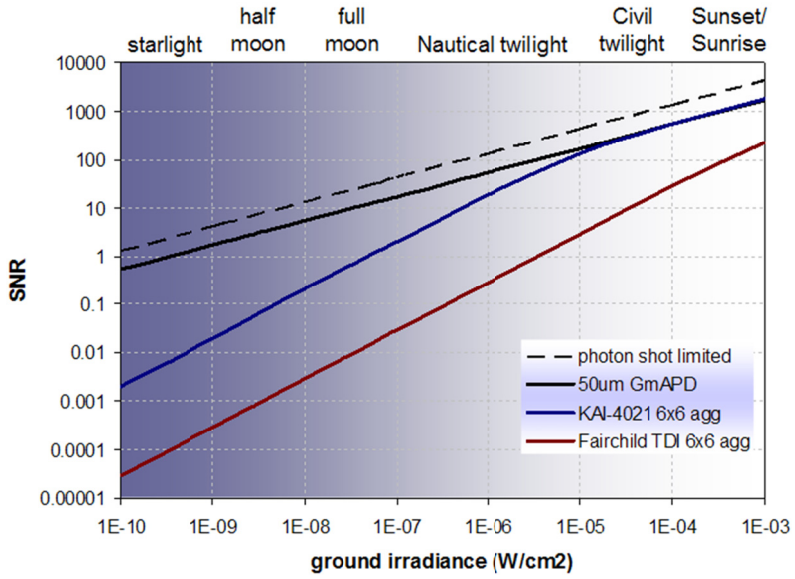

FIGURE 5-1 Curves demonstrating the theoretical sensitivity of an ideal shot-noise-limited 50 µm pitch, 100 percent QE sensor (dashed line); a 50 µm pitch, GM-APD sensor; a 6 × 6 aggregated 8 µm pitch Kodak KA4021 sensor; and a 6 × 6 aggregated 8 µm pitch Fairchild 128-stage time delay and integration (TDI) sensor. SOURCE: Michael Gartley. Detectors, RIT Course Number 1051-465 Lecture Noise. Available at http://ridl.cfd.rit.edu/products/training/Detectors%20465%2020083/Lecture/Lecture%2011-Noise/Lecture%20Noise.pdf.

[4] More detail on the following and additional noise sources is found in National Research Council, 2010, *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*, National Academies Press, pages 33-36.

[5] True for uncorrelated light photons from a source such as incandescence but with coherent light and proper selection, the light's amplitude noise can be reduced by transfer or "squeezing" fluctuations into the light phase resulting in subclassical shot noise light.

[6] Since for Poisson statistics the variance is equal to the mean, shot noise limited means the detector noise must be less than the shot noise = $\sqrt{N}$ where N is the number of incident photons.

[7] Examples include processing procedures such as balanced mode detection that reduces laser noise by splitting the received signal onto two detectors and rejecting any imbalance between the photocurrents generated by the reference and signal detectors (http://www.newport.com/New-Focus-Application-Note14-A-Survey-of-Methods-/919636/1033/content.aspx).

[8] See discussion of Geiger mode below and in Chapter 4.

Dark Current Noise

In semiconductor detectors, additional noise signal that comes from thermally generated electrons called dark current. It appears even when there are no impinging photons. The dark current is from electrons interacting with thermally excited crystal vibrations called phonons. The excited electrons that exceed the semiconductor bandgap energy produce unwanted dark current.

Dark current is like shot noise and cannot be discriminated from incident photon shot noise or other shot noise sources. The dark current shot noise is dependent on bandwidth (*BW*) and detector area (*A*):

$$n_{dark} = \sqrt{i_{dark}(BW * A)}$$

Another detrimental effect of dark current is that it can accumulate, reducing the detector charge storage capacity, and thereby limiting detector dynamic range and integration time. Most detector applications reduce detector dark current by cooling.

Other Noise Sources

The IR background light flux F_{back}, which depends on scene and target temperature, produces scattered photons that also contribute to detector shot noise.

$$n_{photon=} = \sqrt{i_{photon}(BW * A)}$$

The background noise contribution is also not directly detector related but can be significant for applications viewing dim targets. The readout of detectors adds several fundamental noise components such as Johnson, kTC (capacitor reset noise), persistence, and 1/f noise.[9] Some components of this noise are detector-related and others are system- or application-related.

Johnson noise arises from the random motion of electron flow that results from scattering, from electron phonon interactions, and motion induced by thermal interaction; this adds a shot noise component to the readout noise.

$$n_{Johnson} = \sqrt{4kT(BW/R)}$$

where *R* is the detector resistance. For low-resistance detectors like HgCdTe and low-impedance readout integrated circuits (ROICs), Johnson noise is not a significant detector limit except at the very lowest signals.

Readout circuits and detectors both have capacitance, and during detector readout there is generally an associated capacitor reset noise, called kTC. The noise is characterized by

$$n_C = \sqrt{kTC}/q$$

[9] There are other contributions to noise in real systems that are not "fundamental." Examples include pixel-to-pixel charge diffusion, electronic crosstalk, radiation interference from other components, and unstable power supplies. An operational, not fundamental, definition of readout noise, N_{read}, includes all noise sources. One way to measure this noise is to derive the standard deviation of multiple detector reads taken with minimum exposure time and under dark conditions at operating temperature.

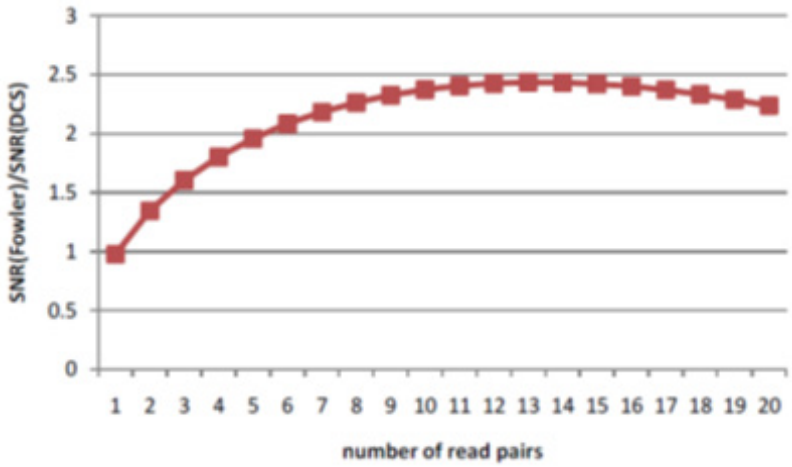

FIGURE 5-2 SNR gain is limited after about a dozen reads in MCDS. SOURCE: Massimo Robberto. Detectors: RIT Course Number 1051-465 Lecture Noise, available at http://ridl.cfd.rit.edu/products/training/Detectors%20465%2020083 /Lecture/Lecture%2011-Noise/Lecture%20Noise.pdf.

where C is the capacitance and q the electronic charge. Within limits (see Figure 5-2), this noise can be reduced by using multiple correlated double sampling (MCDS) data processing before the signal is digitized.[10]

Persistence is noise signal caused by the slow release of trapped charge after the detector is read out. Impurity or defects cause charge traps to decay exponentially with time, so unless traps are eliminated or frozen by cooling there is persistent long tail of thermally activated charge, adding noise and slowing response time. This noise source and excess noise are also important in detectors with gain, as discussed in Chapter 4.

> The principal sources of 1/*f* noise in electronic devices are almost invariably the slow fluctuations of properties of the condensed-matter materials of the devices. In many cases the specific sources of the fluctuations are known. These include fluctuating configurations of defects in metals, fluctuating occupancies of traps in semiconductors … [11]

In general, 1/f noise is technology-dependent and the theory of 1/f noise[12,13] is much disputed and remains a science and engineering mystery.[14] In certain materials such as HgCdTe, however, detailed models seem to well describe the contribution of 1/f noise from defects and surface states.[15] 1/f and

[10] For detectors that can be read without destroying the accumulated signal charge, MCDS and a related technique called Fowler sampling use several nondestructive reads from the detector while it is integrating signal. If N is the number of reads, by averaging the N reads the read noise is reduced by $1/\sqrt{N}$. MCDS processing decreases the system response time and adds to the system processing overhead and is therefore commonly used in astronomical imaging, where the image does not change and has a very low intensity, so that long integration times are unavoidable.

[11] http://en.wikipedia.org/wiki/Pink_noise#cite_note-Kogan-1996-2.

[12] Sh. Kogan, 1996, *Electronic Noise and Fluctuations in Solids*, Cambridge University Press, ISBN 0-521-46034-4.

[13] M. B. Weissman, 1988, "1/f Noise and other slow non-exponential kinetics in condensed matter," *Reviews of Modern Physics* 60(2): 537.

[14] http://www.scholarpedia.org/article/1/f_noise.

[15] M.A. Kinch, R.L. Strong, and C.A. Schade, 2013, "1/f Noise in HgCdTe Focal Plane Arrays," *J. Electronic Materials* 42(11): 3243.

Johnson noise are generally more important in thermal detectors than in photodetectors. 1/f noise is only important at lower frequencies.

Lumping all these readout noise sources into a single readout noise parameter, N_{read}, allows an estimate of the detector SNR that includes the signal and shot noise from the incident flux, background flux, F_{back}, dark current and total readout noise, and the effect of integration time, t:

$$SNR = \frac{S}{N} = \frac{\frac{q}{h\nu} FtQE}{\sqrt{\frac{q}{h\nu} FtQE + \frac{q}{h\nu} F_{back} tQE + i_{dark} t + N^2_{read}}}.$$

where the highlights show the object signal (blue) and the background signal (red) while the dark current noise is green and the total readout noise is black.

The SNR is improved by increasing the numerator compared with the denominator. Fundamental detector parameters that can be improved are the QE_v, i_{dark} and components of N^2_{read} such as $1/f$, kTC, persistence, and Johnson noise from the detector. Other parameters such as t and the component of kTC introduced by external readout circuitry are controlled by the system design and engineering practices.

As seen in the equation above, the signal increases with exposure time, t. Noise also increases with time, but generally not by as much. Within limits, SNR improves with integration time as the dominant noise term changes from readout to shot noise (see Figure 5-3).

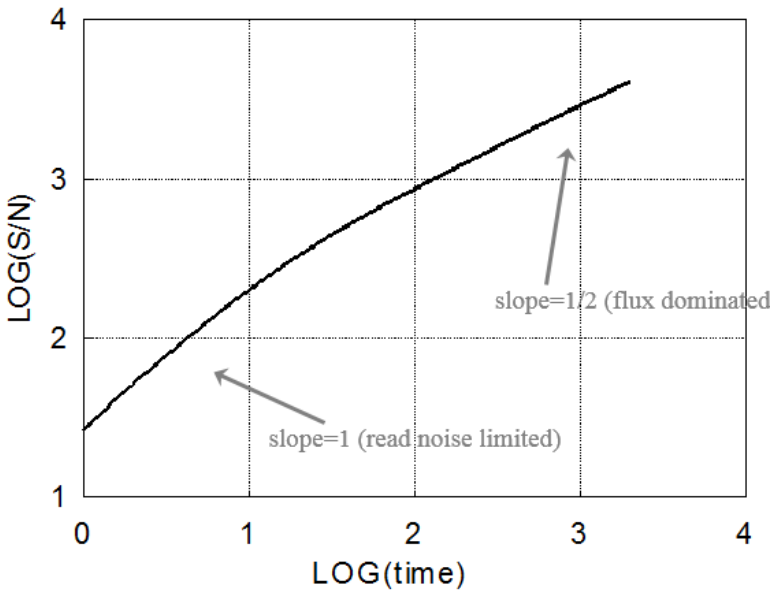

FIGURE 5-3 Signal with a minimum noise at $t = 0$ increases compared with noise as integration time increases, but the increase slows as the overall SNR becomes flux-, object-, and background-noise dominated. SOURCE: Don Figer, Detectors: RIT Course Number 1051-465 Lecture Noise, available at http://ridl.cfd.rit.edu/products/training/Detectors%20465%2020083/Lecture/Lecture%2011-Noise/Lecture%20Noise.pdf.

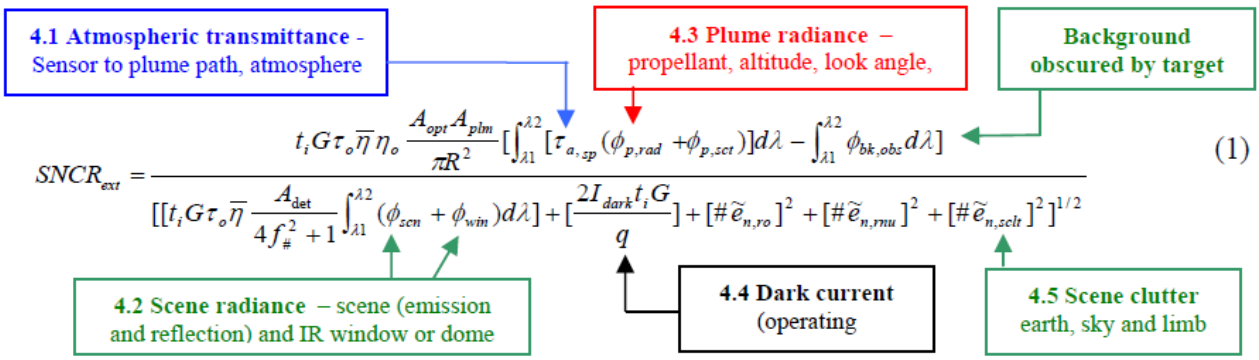

FIGURE 5-4 SNCR model for missile detection that includes many factors in addition to the detector performance. SOURCE: R. Richwine, A.K. Sood, R.S. Balcerak, and K. Freyvogel, 2007, "EO/IR sensor model for evaluating multispectral imaging system performance," *Proc. SPIE* 6543, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVIII, 65430W, April 30.

In practical applications of detectors there is an inherent interplay between the fundamental, operational, and system parameters. For example, obscurant can be often be better penetrated by using longer wavelengths, holographic processing[16] and range gated processing. Range gating may improve as the fundamental detector parameter response time improves. However, often there are other practical limits such as the RC time constant of the detector's external readout circuitry, which overwhelms the fundamental detector performance. As another example, the background flux is greatly reduced by being in space or cooling the aperture, but these options are not often operationally practical. For illustration of the complexity of noise estimation in real applications, a system performance model is provided (Figure 5-4) that has many of the important detector and nondetector factors that determine the real-world application signal to noise and clutter ratio (SNCR) for missile detection:[17]

Fundamental detector properties are important constraints on applications but for practical systems there are many other factors that require time to manufacture, engineering attention, operational constraints, and—finally—the funds necessary to build, train, and use the system.

Detector Gain

The SNR estimate assumes the detector signal responds to incident photon linearly and without gain, but detectors can be very nonlinear when they have intrinsic gain from operating in the linear avalanche or Geiger mode. Avalanche gain comes with the cost of excess signal noise but with the advantage that the gain can overcome readout and other system noise sources. See the subsection "Comparison of Linear- and Geiger-mode Systems" in Chapter 2 and the subsection "Comparison of Linear- vs. Geiger-Mode APDs" in Chapter 4.

[16] M. Locatelli, E. Pugliese, M. Paturzo, V. Bianco, A. Finizio, A. Pelagotti, P. Poggi, L. Miccio, R. Meucci, and P. Ferraro, 2013,"Imaging live humans through smoke and flames using far-infrared digital holography," *Optics Express* 21(5): 5379.

TABLE 5-1 Infrared Detector Parameter Limits with Comments on the Major Implications

Parameter	Limit	State of the art	Implications
Dark current at 140 K, HgCdTe arrays, 2.5 μm cutoff	10^{-3} electrons/s/pixel[a]	0.01 e/sec/pixel at 90 K	Improves SNR in low background
kTC noise	0 electrons	3 electrons/64 reads	Reduces SNR
Johnson noise	0 electrons	Few electrons/s	Important mostly for thermal detectors
1/f noise	Generally a theoretical mystery but well understood in particular cases[b]	Few electrons/s	Important for low frequencies
Quantum efficiency	100 percent	70-80 percent	Improves SNR

NOTE: See also Table 4-3, "Performance Parameters Comparison of Linear-, Geiger-, and Coherent-Mode Detectors" for more details on state-of-the-art APDs.
[a] J.W. Beletic, R. Blank, D. Gulbransen, D. Lee, M. Loose, E. C. Piquette, T. Sprafke, W. E. Tennant, M. Zandian, and J. Zino, 2008, "Teledyne imaging sensors: infrared imaging technologies for astronomy and civil space," Proceedings of the SPIE Conference on Astronomical Instrumentation (Marseille, France).
[b] M.A. Kinch et al., op. cit.

Fundamental Parameters

The key fundamental detector characteristics, their theoretical limits, and estimates of the current state of the art for selected detectors are summarized in Table 5-1, along with comments on implications of achieving those limits.

Key Detector Technologies and Trends

Materials and Processing Trends

Since detectors are not perfect crystals with uniform boundaries and contacts, material and fabrication improvements tend to reduce noise and minimize signal loss. For example, a serious source of semiconductor detector dark current noise is current generated from bulk and surface states that have lower energy than the bandgap.[18] There is some evidence that surface states may also contribute to 1/f noise. There are always efforts to fabricate detectors with lower impurities, fewer crystal defects, better contacts, and cleaner surfaces. Detectors are grown on lattice matched substrates that reduce the crystal strain physical imperfections. Lattice matched atomic layers with high purity materials and precise chemical compositions are formed with techniques such as HgCdTe grown by metal organic vapor phase epitaxy (MOVPE) on GaAs substrates. Special care is sometimes taken to anneal the detectors and add the treatment of hydrogen overpressure that is thought to create passivizing interstitial hydrogen. Many processing steps are proprietary and not published in detail but they reduce defects or deactivate them so the result is less thermally generated noise, better response time and lower signal losses. Better material perfection, uniformity and purity is a continuous quest.

[18] Semiconductor dark current from electrons excited over the bandgap is unavoidable, but materials with a larger bandgap have smaller dark current at the cost of requiring higher energy (smaller wavelength) photons for sensing. There are, however, some materials, such as graphene, that have unusually small coupling of the crystal vibrations (phonons) to electrons, and these materials promise much smaller dark currents at ambient temperature.

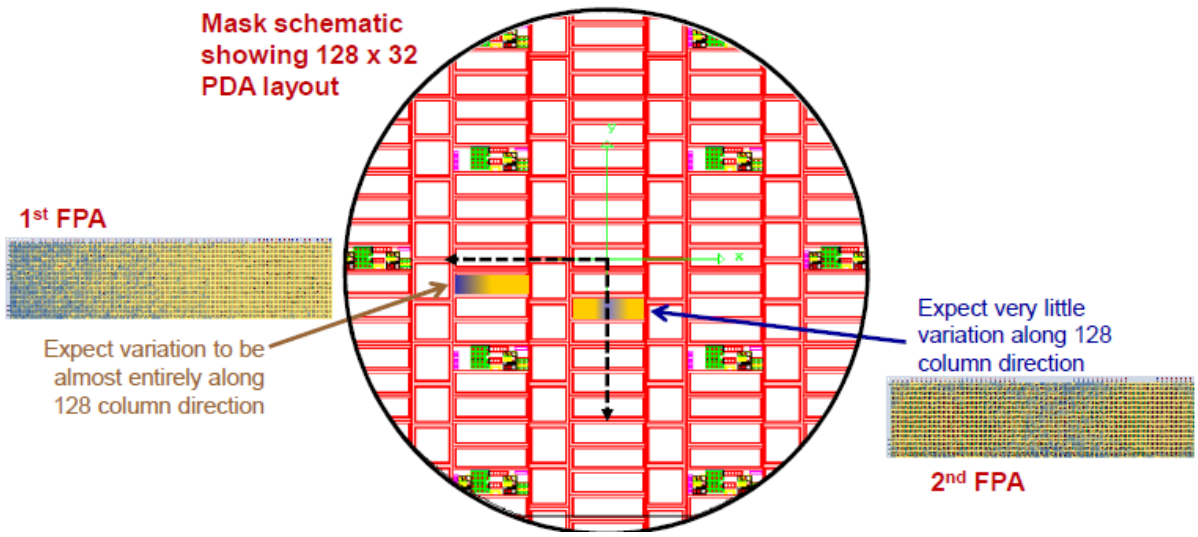

FIGURE 5-5 FPA variation within each chip and on the wafer. The pixel variations in dark count rate (DCR) are shown on first FPA and in photo detection efficiency (PDE) on the second FPA. Variations are shown by the color changes. The epitaxial growth process generally makes radial variations in DCR and PDE. SOURCE: Mark Itzler, Princeton Lightwave, presentation to the committee on Jan 30, 2013. Approved for public release, unlimited distribution.

Processing large focal plane arrays is more complicated than making single detectors. The material and processing improvements are especially difficult where precise atomic stoichiometry and layer thickness control is needed for a few tens of micron-size detector pixel over the relatively vast area of the multimillimeter size chips and even larger multicentimeter-sized production wafers. Achieving near-fundamental detector performance requires extremely precise and accurate process control over large areas. Typical state-of-the-art fabrication variations for InGaAsP/InP Geiger-mode avalanche photodiodes (GmAPDs) with high-efficiency single photon sensitivity at 1.06 μm are seen in Figure 5-5.

Pixel Size

Even if the specific detectivity D* does not improve, array pixel size is shrinking and will continue to do so because of technological advances. This trend is seen in Figure 5-6.

Smaller pixel size allows smaller system SWaP because smaller diffraction-limited optics and detector size reduction means the load on the cooling subsystem is reduced. There is some benefit from pixels even smaller than the incident wavelength from oversampling, but this benefit increases processing overhead.

Superlattice Detectors

The properties of superlattice detectors are being improved in several ways in attempts to achieve fundamental performance limits. Because the energy band structure in an n-type GaAs/AlGaAs quantum well infrared photodetector (QWIP) enforces directional photo absorption selection rules, normal incident radiation absorption is not possible.[19] Optical structures that enhance normal incident absorption are seen in Figure 5-7.

[19] Y. Fu1, M. Willander, W. Lu, and Wenlan Xu, 1998, "Optical coupling in quantum well infrared photodetector by diffraction grating," *J. Appl. Phys.* 84: 5750.

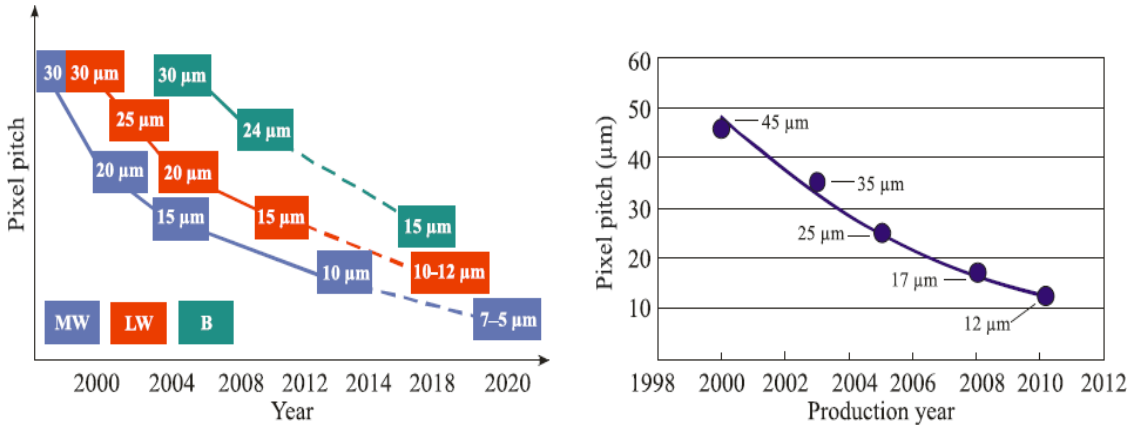

FIGURE 5-6 Pixel pitch for HgCdTe photodiodes (*left*) and amorphous silicon micro-bolometers (*right*) have continued to decrease due to technological advancements. SOURCE: *Left*, G. Destefanis, P. Tribolet, M. Vuillermet, and D.B. Lanfrey, 2011, "MCT IR detectors in France," *Proc. SPIE* 8012: 801235-1; *Right*, A. Hoffman, 2006, "Semiconductor processing technology improves resolution of infrared arrays," *Laser Focus World*, February, 81.

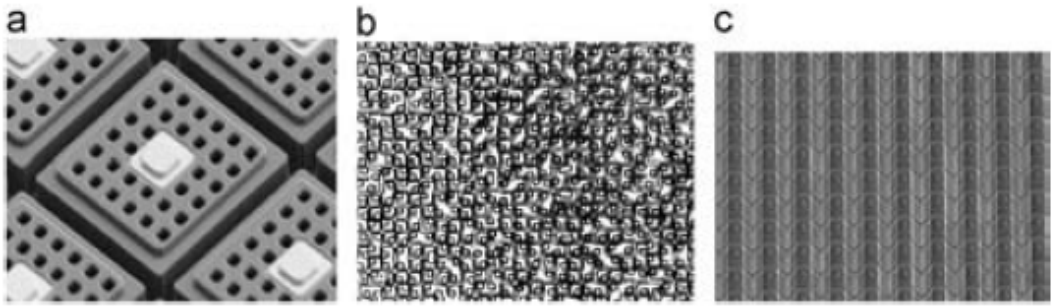

FIGURE 5-7 Grating light-coupling mechanisms used in QWIPs: (a) gratings with optical cavity, (b) random scattered reflector, and (c) corrugated quantum wells. SOURCE: A. Rogalski, 2012, "Progress in focal plane array technologies," *Progress in Quantum Electronics* 36: 342, with permission from Elsevier.

QWIPs (n-type) have a gain of <1. The lower gain on QWIP detectors must be compensated by higher gain ROICs, which add to the readout noise. Better ROICs and improved quantum efficiency are recognized as important developments for progress in QWIPs.[20]

Quantum Dots

Researchers are pursuing quantum dot infrared photodetectors (QDIPs) because in 2002 a theoretical analysis predicted that "quantum dots are expected to have the potential to outperform quantum wells by several orders of magnitude and compete with HgCdTe."[21] QDIPs use a distribution of dots of material for photon absorption and carrier production, while QWIPs use a thin heterogeneous semiconductor for the same purpose.

[20] Yetkin Arslan, Tahir Colakoglu, and Cengiz Besikci, 2013, "Diffraction-grating-coupled high quantum efficiency InP/InGaAs quantum well infrared photodetector focal plane array," *IEEE Journal Of Quantum Electronics*, 49(2): February.

[21] J. Phillips, 2002, "Evaluation of the fundamental properties of quantum dot infrared detectors," *Journal of Applied Physics* 91(7): April.

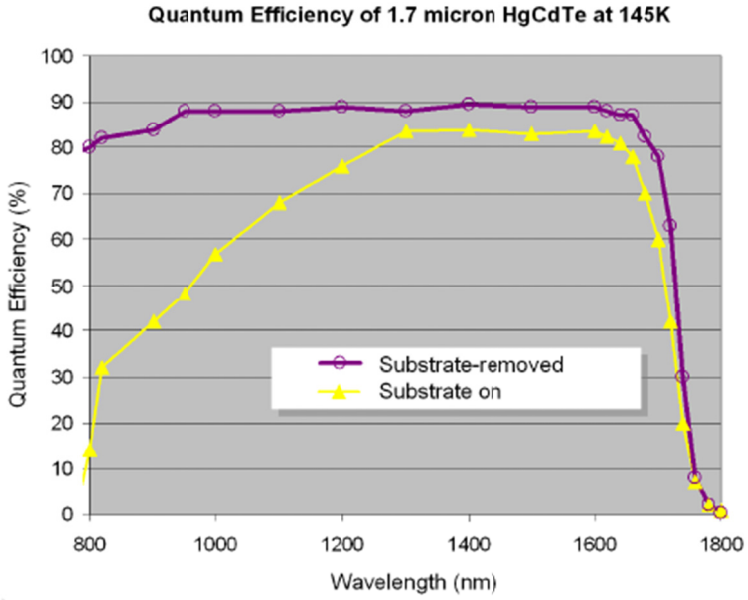

FIGURE 5-8 HgCdTe detector spectral range extended by removal of substrate. SOURCE: J.W. Beletec, R. Blank, et al., 2008, "Teledyne imaging sensors: infrared imaging technologies for astronomy and civil space," *Proceedings of the SPIE,* Conference on Astronomical Instrumentation (Marseille, France), Courtesy of Teledyne Imaging Sensors.

Currently, quantum dot infrared photodetectors have problems related to dot size and uniformity.[22] QDIPs are promising, but there are as yet no commercial or military applications.[23]

HgCdTe Versus Other Materials

For the next decade, HgCdTe technology will likely continue to be the IR detector material of choice for MWIR and LWIR applications because of its excellent properties and its increasing manufacturability. Despite "competition from alternative technologies, HgCdTe is unlikely to be seriously challenged for high-performance applications—those requiring multispectral capability and fast response"[24]. HgCdTe arrays are sensitive over a broad spectral range and can incorporate low-noise gain, reducing the demands on ROICs. As seen in Figure 5-8, HgCdTe detectors without a substrate show good quantum efficiency into the near IR, and there are plans to extend this spectral range into the visible and, possibly, the ultra-violet.

However, the chemical stability of HgCdTe is not perfect, and nonuniformity over large areas is a challenge (see Figure 5-5). There is a steep learning curve for those seeking to fabricate high-uniformity HgCdTe focal plane arrays.

[22] P. Bhattacharya, S. Ghosh, and A. D. Stiff-Roberts, 2004, "Quantum dot optoelectronic devices," *Ann. Rev. Mater. Res.*, 34: 1, August.

[23] See additional discussion of QDs in Chapter 4.

[24] A. Rogalski, 2012, "Progress in focal plane array technologies," *Progress in Quantum Electronics* 36(2): 342-473.

"Type II InAs/GaInSb superlattice structure is a relatively new alternative IR material system and has application potential for LWIR/VLWIR spectral ranges with performance comparable to HgCdTe, but it does not compete at the shorter wavelengths"[25]. Sb-based superlattices are fabricated using standard III-V technology that may be more competitive due to lower costs of series production.[26]

HgCdTe technology will continue to dominate military passive and advanced MWIR and LWIR active applications of IR detectors because of its excellent detector properties such as broad spectral response and low noise gain. Type II InAs/GaInSb and related superlattice structures are an "alternative IR material system that has great potential for dual LWIR/VLWIR spectral ranges with theoretical performance comparable to HgCdTe with the similar cutoff wavelength"[27]. Based on what appeared to be more mature material and fabrication processes, hopes were high, but actual progress has been slow. More research is needed to understand avalanche gain limitations, gain noise, quantum efficiency, and choice of optimal semiconductor alloys.

SIGNAL PROCESSING

Limits of Signal Processing

Signal processing consists of a series of computational steps in the transformation of data into information. In the case of a ladar, the steps might begin with a ROIC attached (typically bump-bonded) to a detector array. The ROIC puts out multiple concurrent data signals that must be "processed" by a computational method ("algorithm"). Figure 5-9 illustrates a set of steps mapped onto some hardware components.

The computational method is typically decomposed into steps. The decomposition is based on either or both of the following: (1) orderings of operations required by the mathematics of the method (e.g., a series of transforms performed in order) and (2) current technology limits on storage, communication and computation. These technology limits can be component limitations (e.g., the clock rate of a processor or the cycle times of storage devices) or architectural limitations (e.g., mismatches among components, such as wireless communications links and airborne sensor collection capabilities or limitations on parallelism inherent in the algorithm).

Each step will have an associated delay (measured in seconds) and throughput (measured in bits per second). The delay through the signal processing system as a whole is the sum of the delays for each step, plus the time needed to communicate data between each step. Several schemes can be used to improve performance, some of which are enabled by improvements in algorithms and their implementations, and others of which are enabled by improvements in hardware technologies such as semiconductor manufacturing.

An example of an algorithmic improvement is one that allows multiple data items to be processed at once, in a mode of operation called parallel processing. When n operations can be performed at the same time, the delay per operation is reduced by a factor of n and the throughput is increased by a factor of n. So, if a signal processing step allows for 10 operations to be processed at once and 10 hardware units are available, it takes 1 unit of delay to process 10 operations, so each operation accumulates 1/10 of the delay required for nonparallel operation. Since 10 operations on b bits of data are performed at once, the

[25] A. Rogalski, 2009 "Infrared Detectors for the Future" *ActA PhysicA Polonica A* 116: 389-406.

[26] A. Rogalski, Material considerations for third generation photon detectors, http://www.phy-astr.gsu.edu/qwip2006/Presentations/Material%20considerations%20for%20third%20generation%20detectors%20-%20A%20R.pdf. Accessed on March 14, 2014.

[27] A. Rogalski, 2010, Infrared Detectors, Second Edition, CRC Press, pp. 846–847.

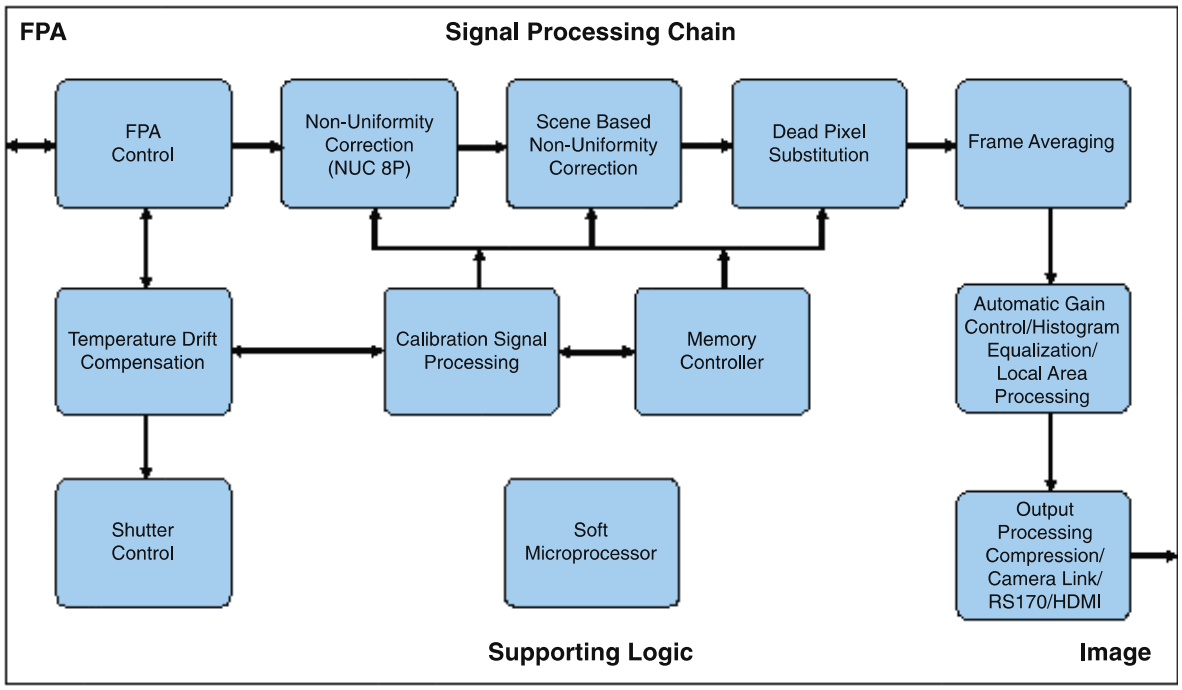

FIGURE 5-9 Components of a signal processing algorithm for an FPA. SOURCE: Figure 4-10 from National Research Council, 2010, *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*, The National Academies Press, Washington, D.C.

throughput is 10 × *b*, or 10 times as much. Unfortunately, not all computation is parallelizable, and the gains for parallelism are limited to the portions of the computation that can be parallelized. Amdahl's Law[28] states this succinctly, as:

$$\text{Speedup for } n \text{ processors} = 1/(serial + ([1-serial]/n))$$

where *serial*, a number between 0 and 1, is the fraction of the computation that must be executed one operation at a time. Fortunately, much of the signal processing can be parallelized and the major limits are technological in nature (e.g., the number *n* of operations that can be carried out in parallel). These limits are being addressed aggressively—for example, by companies such as NVidia that develop graphics processor units (GPUs) that carry out thousands of operations simultaneously. For example, an NVidia Tesla K20X offers[29] 2688 cores and peak performance figures of 1.31 double-precision TFLOPS[30] and 3.95 single-precision TFLOPS.

A technology improvement such as an improvement in semiconductor fabrication can affect the ability to integrate multiple steps of an algorithm on a single chip. This reduces the delay for communicating data between steps of an algorithm and reduces the potential for errors and noise in a communications path. Further, integration may permit further processing optimizations when data are

[28] Gene Amdahl, 1967, "Validity of the single processor approach to achieving large-scale computing capabilities," AFIPS Spring Joint Computer Conference Proceedings, 483.

[29] http://www.nvidia.com/object/tesla-servers.html.

[30] 10^{12} floating point operations per second.

TABLE 5-2 Limitations on Various Signal Processing Parameters

Parameter	Limit	State-of-the-Art	Implications
Digital process energy	kTLn(2) = 0.0178 eV	1×10^4 eV	Battery life, system size, and Moore's Law seems to prevent improvement beyond 100 ×.
Parallelized speedup	1/(*serial*+([1-*serial*]/*n*)) (See discussion in text.)	64 (problem dependent, technology dependent)	Faster processing for certain problems
Power efficiency in multiply accumulates per second	Assuming digital limit kTln(2) ~1 nw	10 /mW	Analog processing might improve the state of affairs by 1000×.

accessible to multiple steps of a computation—for example, processing it in place or exploiting the principle of locality of reference.[31]

The ultimate limits of parallelism and integration are achieved when all of the digital functions are combined with the ROIC. While seemingly far-fetched, technology and algorithmic directions pointing towards this outcome are being explored by, among others, the Defense Advanced Research Projects Agency (DARPA) in the RAPID UPSIDE[32] Program. These limits are summarized in Table 5-2.

Another important system consideration is energy. In the shorter term, DARPA has a program called Power Efficiency Revolution for Embedded Computing Technologies (PERFECT)[33] focused on gaining higher performance processing on a lower energy budget, with a program target of at least 75 GFLOP/W. While it is not clear how to achieve it, there is a lower bound on the energy required for computing derived from thermodynamics, called Landauer's principle,[34] which states that there is a minimum amount of energy, kT ln(2), required to change a bit. This prediction has recently been verified experimentally.[35] Landauer's principle applies to irreversible computations (those that increase entropy) but arguments have been made[36] that reversible computing[37] might be able to overcome this lower limit, at least for portions of the computation. The current state-of-the-art processor's power consumption is several orders of magnitude higher than that predicted by Landauer's principle, so this fundamental limit is not a practical barrier.

Finally, whenever practical, adding more signal processing functionality to the ROIC is a preferred option. This contributes to making the systems smaller and lighter and may even result in lower overall power consumption. Adding functionality on the ROIC however, often comes at the expense of added thermal management due to the added thermal dissipation, and more significantly as the bandwidth

[31] Locality of reference is observed in multiple forms in computation, but the important idea is that data nearby to each other in time (referenced recently) and space (nearby in the data set) are more likely to be referenced.

[32] RAPID UPSIDE, Solicitation Number: DARPA-BAA-12-53.

[33] http://www.darpa.mil/Our_Work/MTO/Programs/Power_Efficiency_Revolution_for_Embedded_Computing_Technologies_(PERFECT).aspx.

[34] Rolf Landauer, 1961, "Irreversibility and heat generation in the computing process," *IBM Journal of Research and Development*, 5: 183.

[35] Antoine Berut, et al., 2012, "Experimental verification of Landauer's principle linking information and thermodynamics," *Nature*, 483: 187.

[36] Graham P. Boechler, Jean M. Whitney, Craig S. Lent, Alexei O. Orlov and Gregory L. Snider, 2010, "Fundamental limits of energy dissipation in charge-based computing'" *Applied Physics Letters*, 97(10).

[37] Tommaso Toffoli, 1980, "Reversible computing," Springer: Berlin, Heidelberg.

of the circuitry goes up. This aspect of adding functionality on the ROIC is common to passive detectors as well and has been extensively addressed in a recent NRC report.[38]

In summary, there are a few unique efforts to develop more efficient image processing algorithms and specialized IR image processing hardware. More importantly, EO IR systems will continue to benefit from spinoffs from the imaging processing hardware developed for games as well as the low-power processors developed for mobile phones and cameras. These will make smaller, lighter, and lower-power IR systems available worldwide.

There are multiple axes along which processing performance is measured, including throughput, latency, and SWaP requirements. Technological improvements in parallel processing and integration are promising but may come at the expense of SWaP. The technology goal can be expressed as maximizing the ratio of performance/SWaP. Evidence from biological systems (e.g., the human brain) and physics (e.g., Landauer's principle) indicates that substantial opportunities exist.

Recommendation 5-1: Novel approaches to signal processing based on biological systems and targeting physical limits should be explored.

PROPAGATION EFFECTS

Atmospheric Propagation

The atmosphere affects active EO sensors in three main ways: absorption, scattering, and refractive index variations (the latter causing beam spreading or fluctuations). For active EO sensors, these effects occur both for the illumination beam and for backscatter from the object to the sensor. Obviously absorption and scattering are negative influences. Normally, refractive index variations are also negative, but there are occasions where something called lucky imaging can turn these variations into a positive influence.[39] In lucky imaging, portions of the image that are improved by the index of refraction variations are kept and combined over time with other improved portions of the image. Angular resolution beyond the diffraction limit can occur using lucky imaging. Index of refraction variations can cause beam spreading, beam wander, image dancing, and beam scintillation.

Absorption is a process whereby atmospheric molecules absorb energy from photons. This is a quantum process, so photons with a certain amount of energy, at certain wavelengths, are preferentially absorbed. Figure 5-10 shows atmospheric transmission vs. wavelength for the visible and infrared regions.

Scattering is another significant effect of the atmosphere. Rayleigh scattering occurs when the photons scatter off of particles that are small compared to the wavelength of light. Rayleigh scattering varies inversely with the fourth power of the wavelength and is isotropic.[40] When the wavelength of light is very much larger than the particles, very little scattering occurs. This is why microwave radar usually has excellent transmission through the atmosphere, even through clouds, fog, and often rain. Infrared radiation is not scattered as much as visible radiation. Mie scattering occurs when particles are about the same size as the wavelength of the light and is biased toward forward scattering. Mie scattering tends to scatter more than Rayleigh scattering. Particles much larger than the wavelength of electromagnetic radiation exhibit reflection and refraction, similar to hitting a large object. Total atmospheric transmission, combining losses due to absorption and due to scattering, is characterized at a given wavelength by Beer's law:

---

[38] National Research Council, 2010, *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*, National Academies Press.

[39] David L. Fried, 1978, "Probability of getting a lucky short-exposure image through turbulence," *Optical Society of America* 68(12): 1651.

[40] C.E. Barnett, 1942, "Some application of wavelength turbidimetry in the infrared," *J. Phys. Chem* 46(1): 69.

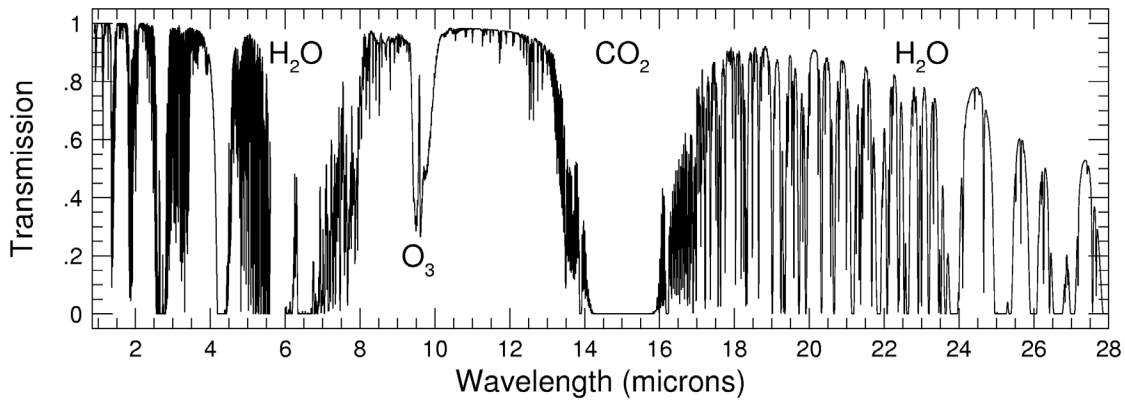

FIGURE 5-10 Atmospheric absorption. Figure created by R. Pogge from ATRAN model data of Lord, S.D. (1992, NASA Technical Memor. 103957) from the Gemini Observatory website (http://www.gemini.edu).

$$\tau = \exp[-\alpha(\lambda)L]$$

where τ is the absorption over some distance and α is the absorption per unit distance for a given wavelength.[41] Beer's law provides an exponential decay of the signal level vs. distance. For large values of α, Beer's law can dramatically reduce signal level after some distances. There is a phenomenon called a precursor wave, which exhibits only linear decay vs. wavelength, but the explanation is thought to be that the decay is the sum of exponential decays with different decay rates at different wavelengths. The sum of the wavelengths exhibits linear decay, but the mix of wavelengths transmitted shifts to those wavelengths with lower exponential decay.

Turbulence can be described in two limiting cases: thin turbulence or distributed turbulence, as shown in Figure 5-11.

Thin turbulence is what is usually considered, because it is easier to compensate. With thin turbulence it is assumed that all the phase distortions are at, or near, the pupil plane. As a result, phase change that occurs in the pupil plane will completely compensate for the thin turbulence optical distortions. Adaptive optics systems can provide phase compensation for thin turbulence.

Aero-optical effects occur near an air vehicle, and therefore fit into the region of pupil plane disturbances. The difficulty with compensating for aero-optical effects is that they tend to be much more rapid than other sources of turbulence, requiring perhaps an order of magnitude faster compensation (on the order of ten kHz instead of 1 kHz).

Due to the diffraction limit, when an illuminator beam is emitted from the active EO system and sent to the target, the beam can be made narrower by using larger optics. Large optics mean high antenna gain and a narrow beam. This puts high flux (watts per square meter) on a target. The atmosphere, however, can limit how narrow the illumination beam is even with large optics. David Fried defined the Fried parameter, r_0, which is the largest effective diameter that can be used under given atmospheric conditions.[42] When calculating the beam divergence of the illuminator beam, the estimated full width, half max, beam divergence of the illuminator beam in a vacuum is:

$$\vartheta \approx \frac{\lambda}{D}$$

[41] Larry Andrews, 2004, *Field Guide to Atmospheric Optics*, SPIE FG04: 4.

[42] D. L. Fried, 1966, "optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *Journal of the Optical Society of America* 56(10): 1372.

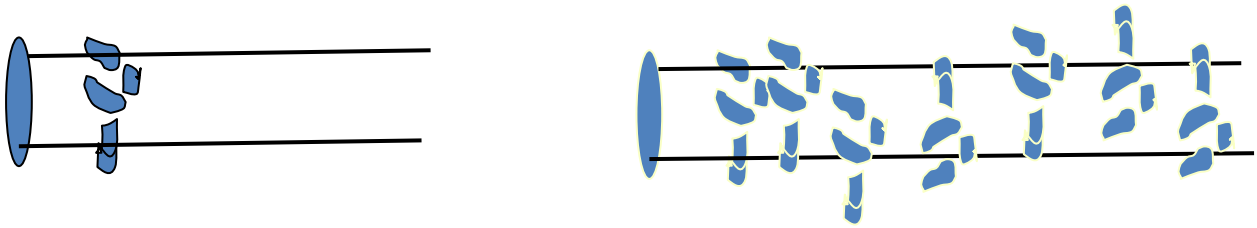

FIGURE 5-11 Thin turbulence is shown on the left, and distributed turbulence—sometimes called deep turbulence—is shown on the right.

where D is the diameter of the illuminator aperture. In the atmosphere, however, the beam divergence calculation will substitute the Fried parameter, $r_{0,}$ for D, because the atmosphere will not allow an area larger in diameter than r_0 to have constant phase across it (to act like a single aperture). The Fried parameter depends on the structure parameter, or strength of the atmospheric turbulence, C_n^2, of the atmosphere for that day, time, and location.[43] It also depends on the path length. The Fried parameter[44] may be calculated as

$$r_0 = \left[0.423k^2 \int_{Path} C_n^2(z)dz \right]^{-\frac{3}{5}}$$

where $k = 2\pi/\lambda$.

One major assumption is usually the frozen atmosphere assumption—that is, the atmosphere stays frozen but just moves past at some velocity. The faster it moves, the more turbulent the atmosphere. The velocity of the wind will influence the Greenwood time constant,[45] the time interval over which turbulence remains essentially unchanged.

Scintillation refers to the twinkle of a star, the spatial or temporal fluctuations in irradiance caused by small random index of refraction fluctuations. The Rytov variance characterizes scintillation.

The illuminator beam also suffers from Beer's law transmission. Light scatters off the object of interest in either a specular or Lambertian fashion. Lambertian scatter provides scattering over a wide angle of return, again encountering Beer's law as light returns to the active EO sensor receiver. The index of refraction variations in the atmosphere make the return light come back in a varying, crooked, path. Detail in the object being viewed moves around as the atmosphere distorts its image. When this moving detail is averaged, it reduces angular resolution.

One limit is how much loss might be acceptable. Beer's law can be used to calculate how many absorption lengths would yield an acceptable amount of atmospheric loss, or an acceptable allowance in the link budget. Table 5-3 shows the loss as a function of the number of absorption lengths.

Active EO sensors require two-way transmission, so a 15 db one-way loss means a 30 db loss both ways, or a loss in link budget by a factor of 1,000. From this table, it is obvious that no more than 4 or 5 absorption lengths loss can be tolerated for any reasonable-power ladar. A loss of 1,000 or 10,000 is a significant link budget loss.

Another limit is the amount of beam spreading that can be tolerated for the illuminator beam. If flash illumination is used, then this is not much of a limit. Assuming a 32 × 32 pixel FPA on receive, the illuminator beam can be 32 times larger than a pixel, and the illuminator aperture can be 32 times smaller than the receive aperture. For a 20-cm receive aperture the transmit aperture can be less than 1 cm. Another way to think of this is that the Fried parameter can be less than 1 cm without interfering with the illuminator beam.

---

[43] Larry Andrews, 2004, *Field Guide to Atmospheric Optics*, SPIE FG04: 11.
[44] John W. Hardy, 1998, *Adaptive Optics for Astronomical Telescopes*, Oxford University Press, 92.
[45] See http://en.wikipedia.org/wiki/Greenwood_frequency.

TABLE 5-3 Atmospheric Loss as a Function of Absorption Lengths

Number of Absorption Lengths	Transmission τ	Loss (db)
1	0.367879	4.3
2	0.135335	8.7
3	0.049787	13
4	0.018316	17.4
5	0.006738	21.7
6	0.002479	26.1
7	0.000912	30.4
8	0.000335	34.7

Another limit is the turbulence on receive, which will limit angular resolution. The diffraction limit on receive will be limited to the diffraction associated with an aperture diameter equal to the Fried parameter unless some form of compensation is employed. Methods of compensation were discussed in Chapter 4.

Atmospheric compensation techniques are key for correcting aberrations caused by the atmosphere. Traditional adaptive optics only correct for phase variations at the pupil plane. There is no effective approach to reduce exponential loss except going to longer wavelengths. Longer wavelengths will significantly reduce exponential loss from scattering for small particles.

Conclusion 5-1: Any long-range active EO sensor operating in the atmosphere will be limited by the atmosphere. The amount of limitation will vary from hour to hour and day to day.

Conclusion 5-2: The ability to compensate for extended turbulence will greatly aid long-range active EO sensors. If some method could be found to reduce the exponential decay, such as a soliton wave through highly scattering or absorptive media, this would be immensely beneficial for active EO sensors.

Propagation Underwater

"Active EO sensors transmit energy that is elastically or inelastically returned from the probed region of interest and analyzed for target signatures"[46]. Whether or not active EO sensing is practical underwater depends on whether the target is visible. Two cases can be considered: (1) active illumination from below the surface of the water—that is, the source and detector are immersed, and (2), active illumination and detection from above the surface of the water.

The ultimate limit to observability is the transparency of the water. Besides the natural processes of absorption and scattering in pure water, there may be further degradation of visibility due to suspended particles, turbidity and turbulence, and water color. The final limits will depend on the reflectance and contrast of the target, the character of the optical image, and the detection sensitivity.

[46]Arete Associates, Active EO (electro-optical), http://www.arete.com/operational_customers/active_eo.aspx. Accessed on March 14, 2014

The degree to which seawater is transparent is a function of the combined effects of scattering and absorption of light by the water. Both absorption and scattering cause the amount of light to decrease exponentially with distance z: $I = I_o \exp(\alpha z)$, where α is the fraction of light absorbed per unit length, called the loss coefficient. The loss coefficient will be the sum of the coefficients for all loss mechanisms. For example, the loss coefficient $\alpha = \alpha_{water} + \alpha_{dissolved} + \alpha_{particle}$, assumes that the ocean or lake contains losses inherent in pure water plus those due to dissolved substances and those due to suspended particles. Each of these many coefficients have absorption and scattering components.[47]

Transparency of Pure Water

The absorption of pure water has a strong wavelength dependence and is found to be smallest in the blue-green wavelengths (see Figure 2-23). The minimum absorption is near the blue wavelength 418 nm, where the absorption coefficient $\alpha_{abs} = 0.005$ m^{-1}. The water absorption increases rapidly at red wavelengths because of overtones of infrared vibrational resonances.

In addition to absorption, even pure water exhibits scattering losses. Rayleigh scattering is scattering from molecules, particles, and density fluctuations that are small with respect to the wavelength of the light. This scattering is much larger for shorter wavelengths in the blue because the scattered intensity is proportional to $1/\lambda^4$, where λ is the wavelength of light. This increased scattering in the blue pushes the wavelength of minimum total attenuation more to the green. The optimum operating wavelength would be where the sum of the two absorption coefficients is a minimum, close to the argon laser wavelength of 488 nm. This encouraged early researchers to investigate active underwater sensing using argon lasers. More recently frequency-doubled Nd:YAG lasers, operating at 532 nm, have become more practical. According to Figure 5-12, at this wavelength $\alpha_{abs} = 0.046$ m^{-1} and $\alpha_{scat} = 0.026$, for a total attenuation coefficient of 0.072 m^{-1}. This means that in the purest water, the laser signal at 532 nm will drop to 1/e in a distance of 14 meters. If used for active sensing, the signal must make a round trip and still be detectable. This loss must be figured in the link budget. In addition, any scattering must be gated out or minimized in some fashion.

A range-gated system was built in China using time resolution to minimize the signal from scattering.[48] It was tested in a swimming pool and apparently was able to detect objects at 26 meters. Given the data from Figure 5-12, if the water were ideal, with a loss coefficient of $\alpha = 0.072$ m^{-1}, at 26 m there should still be 15 percent of the incident light. In fact, the signal was very weak at this distance. Images were reported only at 15 m distance. Clearly the loss in real water is much larger than the ideal, even in a quiet pool.

Real water is rarely pure, and additional factors are much more important than intrinsic water absorption and Rayleigh scattering. The other major concern for water transparency is Mie scattering, which is due to particles that are larger than, or on the same order of the size as, the wavelength of the incident light. Mie scattering has no strong wavelength dependence. The direction of scattered light peaks forward. Particles in water such as algae and mud usually introduce both scattering and absorption. In water, Mie scattering from particles larger than a wavelength of light is usually the most predominant. It has these general characteristics: (1) all wavelengths are scattered by roughly similar amounts, (2) most of the light is scattered in a forward direction, much less is scattered backwards or sideways, and (3) Mie scattering is very much stronger than Rayleigh scattering and increases as the concentration of particles increases.

[47] http://www.deepocean.net/deepocean/index.php?science07.php.

[48] Jin Weiqi, Cao Fengmei, Wang Xia, Liu Guangrong, Huang Youwei, Qi Huaichuan, Shen Fei, 2008, "Range-gated underwater laser imaging system based on intensified gate imaging technology," *Proc. of SPIE* 6621: 66210L.

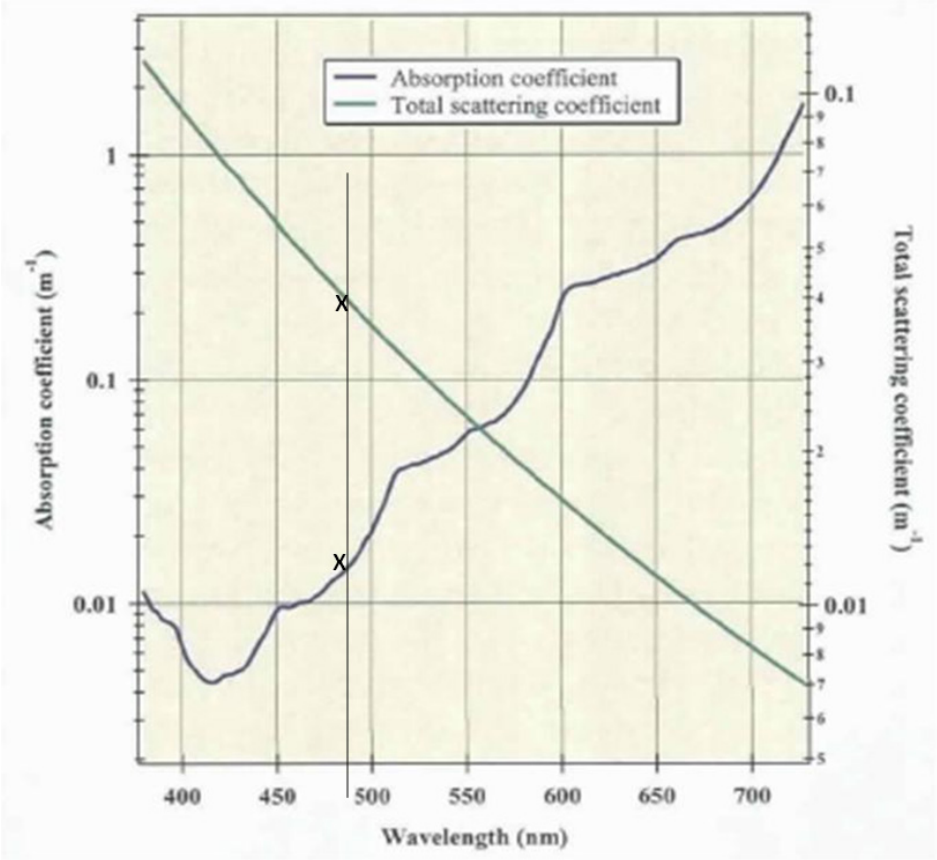

FIGURE 5-12 Spectral absorption coefficient α_{abs} and total scattering coefficient α_{scat} for pure water. The two losses at wavelength 488 nm sum to a total minimum attenuation. SOURCE: Adapted from Figure 5 in R.A. Maffione, 2001, "Evolution and revolution in measuring ocean optical properties," *Oceanography* 14(3): 9.

CONCLUDING THOUGHTS AND OVERARCHING CONCLUSION AND RECOMMENDATION

Since the game-changing invention of microwave radar technology in the Second World War, radar has developed a mature technology and exploitation base and has proliferated to a wide variety of military and commercial applications. Similarly, passive EO sensing systems have developed widespread commercial applications and have been exploited for military use in space, in the air on intelligence, surveillance, and reconnaissance (ISR) missions, and on the ground in night vision systems. Active EO sensing systems enjoy the best of both worlds in that they have resolution advantages over microwave radar and illumination advantages over passive EO systems. Recent advances in laser illuminators, sensitive broadband detectors, and rapid data processing are combining to enable active EO sensing to bring revolutionary advances in areas such as mapping, targeting, autonomous robotics, environmental monitoring, weather prediction, and intelligence gathering. In the context of this committee's task to focus on those areas of active EO sensing that could produce technological surprise adverse to U.S. national security, the committee arrived at the following overarching conclusion and recommendation:

Overarching Conclusion: Active electro-optical sensing is a rapidly emerging technology with many applications across intelligence, military, scientific, and commercial domains. It has the potential to alter the balance in many of these areas. Interest in developing and applying active electro-optical

sensing has consequently risen in other nations, putting U.S. leadership at risk, to the degree that in some instances the United States no longer leads.

Overarching Recommendation: The uses and development worldwide of active electro-optical (EO) sensing should be tracked aggressively by the U.S. intelligence community, and active EO sensing should be aggressively developed by the United States.

Appendixes

A

Committee Biographies

Paul McManamon, *Chair,* is technical director of the Ladar and Optical Communication Institute (LOCI) at the University of Dayton. He is owner of Exciting Technology LLC. Until May 2008 he was chief scientist for the Sensors Directorate, Air Force Research Laboratory (AFRL), Air Force Materiel Command, Wright-Patterson Air Force Base, Ohio. The Sensors Directorate consists of about 1,250 people responsible for developing new sensor technology for the Air Force. Dr McManamon was responsible for the technical portfolio of the Sensors Directorate, including radio-frequency (RF) sensors and countermeasures, electro-optical (EO) sensors and countermeasures, and automatic object recognition. He has developed multidiscriminate EO sensors, including multifunction laser radar, novel EO countermeasure systems, and optical phased-array beam steering. Dr. McManamon participated in three Air Force Scientific Advisory Board (AFSAB) summer studies: *New World Vistas* (1995); *A Roadmap for a 21st Century Aerospace Force* (1998); and *Sensors for Difficult Targets* (2001). He was instrumental in the development of laser flash imaging, initiating the ERASER program as a method to enhance EO target recognition range by a factor of 4 or 5. Dr. McManamon is widely recognized in the electro-optical community. He was the president of the International Society of Optics and Photonics (SPIE) in 2006 and is involved in numerous other scientific organizations.

Walter F. Buell, *Vice Chair,* is principal director of the Electronics and Photonics Laboratory at the Aerospace Corporation. Dr. Buell has been a part of the technical staff at Aerospace for many years. Previously, he has served as systems director for NGA InnoVision Programs, Systems Director for Advanced Studies, manager of the Lidar and Atomic Clocks Section of the Photonics Technology Department, as well as the principal investigator for Synthetic Aperture Ladar programs at Aerospace. Dr. Buell's research interests also include laser cooling and trapping of atoms, atomic clocks, laser remote sensing, and quantum information physics. He has published more than 25 papers in atomic, molecular, and optical physics and holds 3 patents. Dr. Buell has a Ph.D. in physics from the University of Texas at Austin and is a member of the American Physical Society (APS), the Institute of Electrical and Electronics Engineers (IEEE), and SPIE.

Melissa Goodhart Choi is the group leader of the Active Optical Systems Group at the Massachusetts Institute of Technology's (MIT's) Lincoln Laboratory. This group focuses on development of advanced laser radar concepts and is currently involved in the development and operation of experimental and deployed airborne and ground-based three-dimensional laser radars; data exploitation; and simulation and modeling efforts for various applications, with significant efforts in the area of coherent laser radar. She earned a Ph.D. in applied mathematics from North Carolina State University in 1999.

John Devitt is the director of engineering at Raytheon Vision Systems. He was the former division chief of the Electro-Optical Sciences Laboratory's remote sensing group at the Georgia Tech Research Institute (GTRI). Previously he was manager of the Systems Analysis and Test group in the Infrared Products Engineering organization at L-3 Cincinnati Electronics (CE). He has more than 20 years' experience in advanced technology projects in EO-IR, optics, sensors, and related areas, including direct experience in

leading both large-format focal plane array (FPA) development and novel detector programs. He has led major research and development, manufacturing technology, and other advanced technology developmental projects with broad technical and economic scope at both L-3 CE and, previously, the GE Global Research Center. This has involved coordinating multiple cross-technical and organizational groups within businesses, managing university and national laboratory subcontracts, and being a key interface with the major government agencies. In 1993, Mr. Devitt was awarded the GE Sanford Moss Award for Most Outstanding Test and Measurement Program. He has 12 U.S. patents and numerous publications. Mr. Devitt is a certified Six Sigma Green Belt and has an M.S. in physics from Ohio State University.

Elsa Garmire (NAE) is Sydney E Junkins Professor of Engineering Sciences at Dartmouth College, Hanover, New Hampshire, where she served as dean of the Thayer School of Engineering from 1995 to 1997. Prior to that she was William Hogue Professor of Electrical Engineering, professor of physics, and director of the Center for Laser Studies at the University of Southern California, where she had been since 1975. She served on the research staff at Caltech for 9 years under the direction of Amnon Yariv. Dr. Garmire's research has focused on lasers and optics, including opto-electronics, nonlinear optics, optical devices, fiber optics, quantum electronics, device fabrication, and semiconductors. She and her students have authored over 200 journal papers and received 10 patents. She has supervised 30 Ph.D. and 15 M.S. theses. The focus of this research has been to contribute both experimentally and theoretically to the understanding of nonlinear optics, integrated optics, and semiconductor lasers. Dr. Garmire has served the research community as a member of the boards of five professional societies and as president of the Optical Society of America. She has organized seven conferences related to her research interests and has been associate editor for four technical journals. In honor of her contributions to quantum electronics and nonlinear optics, Dr. Garmire has been elected a member of the National Academy of Engineering (NAE) and the American Academy of Arts and Sciences. She was promoted to fellow of the IEEE, the Optical Society of America, the American Physical Society, and the Society of Women Engineers SWE). She received the SWE Achievement Award (its highest award), has been a Fulbright scholar, and was elected an honorary member of Phi Beta Kappa. Dr. Garmire has advised government policy makers by serving on numerous technical advisory boards, panels, and committees for the Army, the Department of Energy (DOE), the Air Force, the National Science Foundation (NSF), the National Research Council (NRC), and the NAE. She has also been involved in public policy through the public affairs committees of the APS and the American Institute of Physics, as well as on the NRC Committee on Science, Engineering, and Public Policy. In 2007 and 2008, Professor Garmire spent a year as a Jefferson Science Fellow at the U.S. State Department, advising in the area of telecommunications. She received an A.B. in physics from Harvard and a Ph.D. in physics from MIT under the direction of Nobel prize winner Charles H. Townes.

Gary W. Kamerman is the president and chief scientist for FastMetrix, Inc., where he has lead the science team that oversaw the design, development, and deployment of several advanced remote-sensing laser systems for the U.S. government. He supervised the design and demonstration of deployable high-performance computer systems, and designed nonintrusive optical instruments for high-speed velocimetry, interior ballistics analysis, remote chemical detection, and combustion diagnostics. Mr. Kamerman invented and patented the optical computing system to process laser radar signals before photodetection. He is a Ph.D. candidate at Kiev Polytechnic Institute, focusing his studies on applied physics.

Kenneth A. Kress is a senior scientist for KBK Consulting, Inc., an affiliate of Montana State University's Department of Physics, and a consultant for Booz Allen Hamilton, where he specializes in quantum information science and other technical evaluations and strategic planning for intelligence and defense applications. Some of his past clients include DARPA's Microsystems Technology Office, Noblis, (formerly known as Mitretek Systems), GTRI, and Lockheed Martin's Special Programs Division. From 1971 to 1999 he worked in a series of positions at the Central Intelligence Agency's

Directorate of Operations, Office of Development and Engineering, and, finally, its Office of Research and Development (ORD). He was first a research and development manager, later a program manager, and finally an ORD senior scientist responsible for management support, the development of technical and strategic plans, and DOD inter-agency coordination for advanced technology. He is the inventor of the solid-state neutron detector, for which he won an award in 1981. He holds a Ph.D. in physics from Montana State University.

Jeanette Lurier is an engineering fellow at Raytheon Company. Ms. Lurier has over 30 years' experience in aerospace engineering and system architecture and has been with Raytheon since 1985. She has spent most of her career working on active and passive EO/IR systems for space applications and developing new technologies in lasers and IR sensors. Ms. Lurier is a Raytheon Certified Architect and is currently managing the ladar product line of Raytheon Space Systems. In this role she focuses on developing strong customer relationships and leveraging her work as mission systems engineer to guide architecture studies, mission analysis, technology roadmap development, and technology development, paving the way for new business opportunities in cutting-edge capabilities. Ms. Lurier originally joined the company as an optical engineer and has worked on all aspects of program execution, including concept generation, hardware design, development, fabrication and test, system design integration and architecture, and technology development. Ms. Lurier holds a B.A. in physics from Cornell University and an M.S. in optical engineering from the University of Rochester.

Pradip Mitra is senior director of advanced development programs at DRS Technologies in Dallas, Texas, where he has worked for the past 10 years. He oversees multiple IR technology development efforts that include active imaging detectors and sensors. Before joining DRS he worked at Lockheed Martin for 13 years, where he was involved in IR materials and detector research. He holds 18 issued patents in the IR technology area and has coauthored over 70 papers. He received a Ph.D. in physical chemistry from Boston University and an M.Sc. from the Indian Institute of Technology in Kanpur, India.

Peter F. Moulton (NAE) is a principal research scientist for Q-Peak, Inc. After finishing graduate school he was employed in the Quantum Electronics Group at MIT's Lincoln Laboratory. In 1985 he joined a start-up company, Schwartz Electro-Optics (SEO), as vice president and managed the founding of the company's Research Division in Concord, Massachusetts. Dr. Moulton became senior vice president of SEO in 1997 and was involved in spinning off the Research Division as a separate company, Q-Peak, in 1998, and in the sale of Q-Peak to its current parent company, Physical Sciences Inc., in 2001. He served as the vice president and chief technology officer of Q-Peak until 2012, when he moved to a part-time position as a principal research scientist. Dr. Moulton's technical work began in the field of bulk solid state lasers and in recent years has extended to include nonlinear optics and fiber lasers. Dr. Moulton is a fellow of the Optical Society of America (OSA) and of the IEEE. In 1997 he was awarded the R.W. Wood Prize from the OSA and the William Streifer Scientific Achievement Award from IEEE/LEOS, and in 2013 he won the IEEE Photonics Award. He was elected to the NAE in 2000. He received an A.B. in physics from Harvard College in 1968 and M.S. and Ph.D. degrees in electrical engineering from MIT in 1972 and 1975.

Jonathan Smith is the Olga and Alberico Pompa Professor of Engineering and Applied Science at the University of Pennsylvania. He has served as a program manager at the Defense Advanced Research Projects Agency (DARPA. Before that, Dr. Smith was at Bell Telephone Laboratories and Bell Communications Research, which he joined at the AT&T divestiture. His current research interests include programmable network infrastructures, network defense, and cognitive radios. Dr. Smith served on the Network and Information Technology Technical Advisory Group of the President's Council of Advisors on Science and Technology, is an IEEE fellow, and is a member of the NRC Board on Army Science and Technology (BAST). He received his A.B. in mathematics magna cum laude from Boston College and M.S. and Ph.D. degrees in computer science from Columbia University.

Abbie (Tippie) Watnik is a Karle Fellow and Research Physicist in the Applied Optics Branch within the Optical Sciences Division at the Naval Research Laboratory in Washington, D.C. Dr. Watnik's current work focuses on applying holographic and computational imaging techniques and principles to intelligence, surveillance and reconnaissance (ISR) systems and countermeasures (CM) to support projects of interest for our national defense. Her research interests include active imaging, diffractive imaging, aberration correction, phase reconstruction, mathematical algorithm development, and nonconventional approaches to image correction. Dr. Watnik earned a Ph.D. and an M.S. from the University of Rochester. As an NSF Graduate Research Fellow and a Harvey Fellow, Dr. Watnik studied aberration correction for multiplane aberrations using a nonlinear optimization approach that employs sharpness metrics as well as constructed and processed gigapixel synthetic-aperture high-resolution imagery. She is a recipient of the Rochester Precision Optics Award and winner of the Mark Ain Technical Business Model Competition and holds a Ph.D. in optics from the University of Rochester.

Eli Yablonovitch (NAS/NAE) is currently a professor of electrical engineering and computer sciences at the University of California (UC), Berkeley. He worked for 2 years at Bell Telephone Laboratories and then became a professor of applied physics at Harvard. In 1979 he joined Exxon to do research on photovoltaic solar energy. Then, in 1984, he joined Bell Communications Research, where he was a Distinguished Member of Staff, and also director of solid state physics research. In 1992 he joined UC, Los Angeles, where he was the Northrop Grumman Professor of Electrical Engineering. In 2007 Dr. Yablonovitch assumed his current position, where he is the James and Katherine Lau Chair in Engineering. Professor Yablonovitch's work has covered a broad variety of topics: nonlinear optics, laser-plasma interaction, infrared laser chemistry, photovoltaic energy conversion, strained-quantum-well lasers, and chemical modification of semiconductor surfaces. Currently his main interests are in opto-electronics, high- speed optical communications, high-efficiency light-emitting diodes and nanocavity lasers, photonic crystals at optical and microwave frequencies, and quantum computing and quantum communication. He received his Ph.D. degree in applied physics from Harvard University in 1972.

B

Meetings and Participating Orginizations

January 29-31, 2013
The Keck Center of the National Academies
Washington, D.C.

Trends in Active EO Technology
Ralph Burnham, Fibertek, Inc.

Compact Femtosecond Laser System with 2 mJ Output Femtosecond Laser Technology and Applications
Peter Moulton, Q-Peak, Inc.

Photon Counting 3-D LADAR Cameras Based on Geiger-mode Avalanche Diodes (GmAPDs)
Mark Itzler, Princeton Lightware, Inc.

March 5, 2013
Raytheon Company
Los Angeles, California

3D Holographic Laser Radar
Joseph Marron, Lockheed Martin Space Systems Company

Active EO Detectors Linear Mode HgCdTe at Raytheon Vision System
Justin Wehner and Michael Jack, Raytheon Company

March 6-7, 2013
The Arnold and Mable Beckman Center
Irvine, California

Application of LIDAR Against Soft Targets
Steven B. Alejandro, Reperi LLC

High Resolution 3D Video
Paul Banks, Tetravue
Quantum Metrology
Claudio G. Parazzoli, The Boeing Company

LIDAR Focal Plane Technologies
George Williams, Voxtel, Inc.

April 17-19, 2013
The Keck Center of the National Academies
Washington, D.C.

Secure Quantum Technologies
Robert W. Boyd, University of Rochester

Lidar Sensors for Navigations and Landing
Farzin Amzajerdian, NASA Langley Research Center

UPSIDE: Unconventional Processing of Signals for Intelligent Data Exploitation
Dan Hammerstrom, Defense Advanced Research Projects Agency

Active EO Imaging Technology
Keith Lewis, Sciovis

Emerging and Enabling Lidar Technologies and Techniques for NASA's Space-based Active Optical Remote Sensing Missions
Upendra Singh, NASA Langley Research Center

Boundary Conditions: Closed vs. Open Systems

May 10, 2013
Sandia National Laboratories
Albuquerque, New Mexico

UV-Laser Induced Fluorescence Remote Sensing Technology and Applications
Randy Schmitt and Bill Seng, Sandia National Laboratories

Fiber Laser Science and Technology
Daniel Soh, Scott Bisson, and Sean Moore
Presented by Wen Hsu, Sandia National Laboratories

THz and Quantum Cascade Lasers
Michael C. Wanke, Sandia National Laboratories

C

Laser Sources and Their Fundamental and Engineering Limits

This appendix summarizes in table form the characteristics of lasers and other light sources systems important to active EO sensing and their fundamental and practical engineering limitations. Tables C-1 to C-4 are called out in the laser discussion in Chapter 4. Tables C-5 through C-12 are summarized in Chapter 5.

TABLE C-1 Key Characteristics of Edge-Emitting Interband Diode Lasers

Material	Wavelength Range (nm)	Single-emitter Power	Single-Emitter Efficiency (%)	Bar Power (W)	Bar Efficiency (%)
GaInN	~380	0.2	10-15		
	400, 450	1.2-1.6	25-30	--	--
	~520	0.05	<<10	--	--
AlGaInP	639-690	0.75-1.5	20-25	--	--
	632, 635, 638	--	--	2.5-8 W	>25
	675	--	--	20 W	>35
GaAlAs	793, 808, 852	5-7 W	60	60-200	50-60
InGaAs	915-976 1,064	10-15 5 W	60-70 50-55	120-200 60	60-70 50-55
InGaAsP/ AlInGaAs	1,470-1,532 1,600-1,700	5-7 W 3 W	30-45 20-25	60-100 40	30-40 20-25
AlGaInAsSb	1,900-2,100 2,300-2,500	1-2 W 1 W	10-15 5-10	15-20 W	10-20

SOURCE: Data provided by Steven Patterson, DILAS, Tucson, Ariz.

TABLE C-2 Key Characteristics of State-of-the-Art Cascade Diode Lasers

QCL: InP/InGaAs/InAlAs	Wavelength of best operation: 4-5 μm; output power: 5 W, CW with 21% efficiency (obtained with AlAs inserts to increase the effective barrier heights). Wavelengths between 2.9 and 4 μm and from 5 to 150 μm operate with lower performance.
QCL: GaAs/GaAs/AlGaAs	Wavelength of best operation: 10 μm; output power: 80 mW at 77 K; CW up to 150 K. Far-IR wavelength: 100 μm; output: 8 mW, CW at 45 K with 0.2 % efficiency. No CW above 117 K. Full wavelength range = 9-300 μm; non-competitive in mid-IR.
QCL: GaSb/(InAs/AlSb)	Wavelengths; 2.6 – 5 μm. CW only with TE cooler
QCL: InP/InGaAs/AlAsSb	To date, non-competitive with InGaAs/InAlAs at any wavelength. No CW room temperature operation.
ICL on GaSb	Wavelength: 3-4.2 μm; output power: 360 mW; efficiency: 15%. Wavelengths from 4.2-6 mm with lower performance.
ICL on InAs	Wavelength: 5.3 μm; Output Power: 40 mW at 180 K; CW up to 248 K. Wavelength range: 5.3-10.4 mm.

NOTE: QCL, quantum cascade laser; ICL, interband cascade laser.
SOURCE: Data provided by Jerry Meyer and Igor Vurgaftman, Naval Research Laboratory, Washington, D.C.

TABLE C-3 Characteristics of Several Common Bulk Laser Materials

Material	Wavelength (nm)	Storage Time (msec)	Cross section (cm^2)	Gain Linewidth (nm)	Saturation Fluence (J/cm^2)
Nd:YAG	1,064	0.24	2.8×10^{-19}	0.6	0.66
Nd:vanadate	1,064	0.09	1.1×10^{-18}	1.0	0.17
Nd:YLF	1,047	0.485	1.8×10^{-19}	1.0	1.0
Nd:glass	1,050-1,060	0.3-0.4	$3\text{-}4 \times 10^{-20}$	20-30	4.7-6.3
Yb:YAG	1,030	0.95	2.1×10^{-20}	9	9.2
Yb:YAG(77K)	1,030	0.85	1.1×10^{-19}	1.5	1.8
Er:YAG	1,645	7.6	5.0×10^{-21}	5	24
Er:glass	1,550	7.9	8.0×10^{-21}	55	16
Ho:YAG	2,090	8.5	1.3×10^{-20}	25	7.3
Ho:YLF	2,050	15	1.8×10^{-20}	25	5.3
Ti:sapphire	800	0.0032	3.0×10^{-19}	225 (100 THz)	0.83
Cr:ZnSe	2,450	0.006	1.3×10^{-18}	1,000 (50 THz)	0.06

TABLE C-4 Properties of Hybrid Lasers

Material	Wavelength (nm)	Pulse Energy (mJ)	Pulsewidth (ns)	Pulse Rate (Hz)
Er:YAG[a]	1,617	30	42	30
Er:YAG[b]	1,645	4.2	100	1,000
Er:YAG[c]	1,645	1.6	1.1	10,000
Er:YAG[d]	1,645	60 W		CW
Ho:YLF[e]	2,050	170	20	100
Ho:YLF[f]	2,050	100	20	1,000
Ho:YLF[g]	2,050	115 W		CW
Ho:YAG[h]	2,090	125	20	100
Ho:YAG[i]	2,090	22	70	1,000
Ho:YAG[j]	2,090	1.7	50	35,000

[a] J.W. Kim, J.I. Mackenzie, J.K. Sahu, and W.A. Clarkson, "Hybrid fibre-bulk erbium lasers—Recent progress and future prospects," 7th EMRS DTC Technical Conference, Edinburgh, 2010.
[b] D.Y. Shen, J.K. Sahu, and W.A. Clarkson, "Highly efficient in-band pumped Er:YAG laser with 60 W of output at 1645 nm," *Opt. Lett.* 31:754, 2006.
[c] R.C. Stoneman, R. Hartman, E.A. Schneider, A.I.R. Malm, S.R. Vetorino, C.G. Garvin, J.V. Pelk, S.M. Hannon, and S.W. Henderson, "Eye-safe 1.6-μm Er:YAG transmitters for coherent laser radar," Proceedings 14th Coherent Laser Radar Conference, July 8-13, 2007, Snowmass, Colo.
[d] D.Y. Shen, J.K. Sahu, and W.A. Clarkson, "Highly efficient in-band pumped Er:YAG laser with 60 W of output at 1645 nm," *Opt. Lett.* 31:754, 2006.
[e] A. Dergachev, "45-dB, Compact, Single-Frequency, 2-μm Amplifier," paper FTh4A.2 in Lasers, Sources, and Related Photonic Devices, OSA Technical Digest (CD), Optical Society of America, 2012.
[f] Ibid.
[g] Ibid.
[h] K. Schmidt, C. Reiter, H. Voss, F. Maßmann, and M. Ostermeyer, "High Energy 125mJ Ho:YAG (2.09 um) MOPA Double Pass Laser System Pumped by CW Thulium Fiber Laser (1.9 um)," paper CA3_4 in CLEO/Europe and EQEC 2011 Conference Digest, OSA Technical Digest (CD), OSA, 2011.
[i] Ibid.
[j] A. Hemming, J. Richards, A. Davidson, N. Carmody, S. Bennetts, N. Simakov, and J. Haub, "99 W mid-IR operation of a ZGP OPO at 25 percent duty cycle," *Opt. Express* 21:10062, 2013.

TABLE C-5 Fundamental Limits of Diode Lasers: Interband, Edge-Emitting

Property	Value	Limit Reason	Comments
Electrical efficiency (theory)	100%	Fundamental energy conservation	One photon per one injected carrier
Electrical efficiency (actual)	40-70%	Multiple device issues	Ohmic losses, injected carrier spreading away from lasing region, active region absorption loss, Auger losses (at long wavelengths)
Wavelength	>380 nm	Materials	Limit of GaAlN material, no other large-bandgap II-VI has allowed PN junction fabrication
	<520 nm	Materials	Limits of GaAlN material
	>630 nm	Materials	Lack of semiconductors with bandgaps in green-yellow-red region that can form PN junctions
	<2,500 nm	Materials	Limits on active materials and DBR structures
Power out (1 emitting facet, 1 TM)	0.1-0.5 W	pn junction physics, facet damage	Height of emitting region limited to 0.5 μm by junction height, width limited by multi-mode operation, intensity limited by facet damage levels
Power out (1 emitting facet, Multi-TM)	15 W (at 9xx μm, less at others)	pn junction physics, transverse lasing, facet damage	Same height limit as above, width of emitting region limited by lasing in transverse direction, intensity limited by facet damage
Power out (1-cm-long, linear bar, multiple emitters)	200 W (at 9xx μm, less at others)	Device temperature, limited by heat removal	Efforts to improve heat removal are underway, to allow limit to become facet damage, but cost and reliability are challenges. Efficiency improvements will allow bars to generate higher powers
Spectral linewidth (single TM)	Several MHz	Coupling of diode current fluctuations to cavity refractive index	Can be reduced to the kHz range through use of an external cavity

NOTE: TM = transverse mode.
Refer to Table C-1 for device properties as a function of wavelength.

TABLE C-6 Diode Lasers Vertical Cavity (VCSEL)

Property	Value	Limit Reason	Comments
Electrical efficiency (theory)	100%	Fundamental energy conservation	One photon per one injected carrier
Electrical efficiency (actual)	40-60% (8xx-9xx nm)	Multiple device issues	Ohmic losses, injected carrier spreading away from lasing region, absorption loss in active region, Auger losses (at long wavelengths)
Wavelength	>410 nm	Materials	Limit of GaAlN material, no other large-bandgap II-VI has allowed PN junction fabrication
	<503 nm	Materials	Limits of GaAlN material
	>630 nm	Materials	Lack of semiconductors with bandgaps in green-yellow-red region that can form PN junctions
	<2,000 nm	Thermal activation	The smaller bandgaps can be bridged with thermally activated carriers
Power out (1 emitting facet, 1 TM)	5-15 mW	Active region heating	Diameter of 1 TM limited to ~4 μm on chip by optical cavity, limits power for a given current
Power out (1 emitting facet, MTM)	5 W	Active region heating region	Done with 300-μm emitting region diameter
Power out (VCSEL arrays)	230 W (maximum at 9xx nm)	Temperature limited by heat removal	Power from 0.22 cm^2 area, can scale further by increasing the area. Low-duty cycles increase peak power to nearly 1 kW from 5 × 5 mm area.
Spectral linewidth (1 TM)	Several MHz	Coupling of diode current fluctuations to cavity refractive index	Can be reduced to the kHz range through use of an external cavity

TABLE C-7 Diode Lasers: Quantum Cascade (QCL)

Property	Value	Limit Reason	Comments
Electrical efficiency (theory)	Variable %	Fundamental energy conservation	Each laser transition efficiency limited by energy of photon divided by bandgap of base material, but multiple transitions in series (typ. 25-75) increase the efficiency by this factor.
Electrical efficiency (actual)	21 %, cw, RT, (40-50% pulsed, 160 K)	Multiple device issues	Ohmic losses, injected carrier spreading away from lasing region, absorption loss in active region, losses through the injector regions. Wavelength = 4.9 μm.
Wavelength	>3.8 μm (cw)	Materials	Limits to depth of quantum well in InP structures, but efficiency is low on the short-wavelength end. Pulsed operation to ~3 μm
	<13 μm (cw)	Materials	Typical limit for room-temperature, cw operation. Pulsed/cryogenic operation to 30 μm
	>60 μm cryogenics for cw THz	Materials	Phonon absorption in InP-based devices prevents coverage of 30-60-μm wavelengths,
Power Out Single device	5 W (4.9 μm)	Thermal heating	Achieved in high-efficiency (21%) devices
Spectral linewidth (1 TM)	Sub-MHz	Current and 1/f noise	Can be reduced to the kHz range with current feedback. Intrinsic noise is several hundred Hz.

NOTE: Refer to Table C-2 for device properties as a function of semiconductor material system.

TABLE C-8 Diode Lasers: Interband Cascade (ICL)

Property	Value	Limit Reason	Comments
Electrical efficiency (theory)	Variable %	Fundamental: energy conservation	Eacc laser transition efficiency limited by the energy of the photon divided by the bandgap energy of the base semiconductor material, but ICLs employ multiple transitions in series so efficiency is increased
Electrical efficiency (actual)	15% (3.7 μm, cw, RT)	Multiple device issues	Ohmic losses, injected carrier spreading away from lasing region, absorption loss in active region, losses through the injector regions
Wavelength	>3 μm, cw	Materials	Limited depth of interband quantum well
	<6 μm cw	Materials	Too high current densities at longer wavelengths
Power out single device	0.36 W (3.7 μm)	Thermal heating	Achieved in high-efficiency (15%) devices

NOTE: Refer to Table C-2 for device properties as a function of semiconductor material system.

TABLE C-9 Solid State Lasers: Bulk Format

Property	Value	Limit Reason	Comments
Optical efficiency (theory)	Quantum defect ≡ ratio of laser wavelength to average pump wavelength	Fundamental: energy conservation	Example: 76% for 1,064-nm Nd:YAG laser pumped at 808 nm. Violated when interaction between active ions allows >1 excited state per pump photon
Optical efficiency (actual)	< quantum defect	Multiple	Poor spatial overlap of pump and lasing regions in material, incomplete absorption of pump, reflection of pump from material surface, losses in laser material, excited state absorption of pump or laser power, upconversion from upper laser level
Electrical efficiency (theory)	Pump electrical efficiency times quantum defect	Fundamental: Energy conservation	Example: Pump diodes with 60% efficiency at 808 nm with Nd:YAG laser at 1,064 nm → 46% electrical efficiency
Electrical efficiency (actual)	< pump electrical efficiency X quantum defect	Multiple	Optical efficiency < quantum defect; pump light loss in transport to laser material. Example: typical diode-pumped Nd:YAG lasers are 20-25% electrically efficient.
Wavelength	>286 nm	Materials	Transparency of host crystal
Wavelength	<7,150 nm	Materials	Long-wavelengths have multi-phonon decay, requires low-phonon hosts; may be impractical for high-power. Better hosts (e.g. YLF) can be used for <4,300-nm lasers
Average power (single device)	2 kW (rods) 10 kW (disks) 15 kW (slabs)	Material fracture, thermal effects	1,060-nm, diode-pumped systems. Higher powers with multiple active media: >100 kW is record (Nd:YAG). Yb:YAG used for thin disks
Average power (1 device, diffraction-limited)	50 W (rods) 1 kW (disks) 15 kW (slabs)	Thermal distortion of laser material	1,060-nm region, diode-pumped systems. Higher powers possible with multiple active media, >100 kW is present record, using Nd:YAG slab. Yb:YAG used for thin disks
Spectral linewidth (theory)	Several Hz	Fundamental: Schawlow-Townes limit	Set by spontaneous emission of gain medium into the laser mode. True for all lasers.
Spectral linewidth (actual, ms time scale)	Several kHz	Technical noise	Fluctuations in optical cavity length from coupling between pump power and gain medium refractive index, acoustic noise, other cavity perturbations. External stabilization can reduce technical noise to the Hz level.
Spectral linewidth (long term)	10-50 MHz	Environmental drift	Slow change in laser cavity temperature
Mode-locked pulsewidth (theory)	~3.5 fs	Fundamental: laser material gain-bandwidth.	Value is for Ti:sapphire at 800 nm, Cr:ZnSe is 7 fs at 2500 nm, Nd:YAG is 2 ps at 1,064 nm
Mode-locked pulsewidth (actual)	~4.5 fs Ti-sapphire at 800 nm	Dispersion in optical cavity, mirrors' spectral response, nonlinearities.	Cr:ZnSe is around 50 fs

TABLE C-10 Solid State Lasers: Fiber Format

Property	Value	Limit Reason	Comments
Optical efficiency (theory)	Quantum defect = ratio of laser wavelength to average pump wavelength	Fundamental: energy conservation	Example: 95% for 1,030-nm Yb:fiber laser pumped at 976 nm. Violated for a few systems when interaction between active ions allows more than one excited state per pump photon
Optical efficiency (actual)	< quantum defect: 88% slope efficiency for Yb:fiber	Multiple	Poor spatial overlap of pump and lasing regions in material, incomplete absorption of pump, losses in laser material, excited state absorption of pump or laser power, upconversion from upper laser level.
Electrical efficiency (theory)	Pump electrical efficiency times quantum defect	Fundamental: energy conservation	Example: Pump diodes with 65% efficiency at 976 nm with Yb:fiber laser at 1,030 nm ->62% electrical efficiency
Electrical efficiency (actual)	< above	Multiple	Actual optical efficiency lower than quantum defect, loss of pump light in transport from diode facet to pump cladding. Example: Yb:fiber at 1,030 nm -> 40% electrical efficiency
Wavelength	>248 nm	Materials	Transparency limit of fiber (up-conversion laser operation in ZBLAN fibers
Wavelength	<3,900 nm	Materials	Multi-phonon relaxation limits operation at longer wavelengths
Average power (single fiber, 1,000-nm)	20 kW	Stimulated Raman scattering	1,030-nm Yb:fiber laser pumped by multiple 1,018-nm Yb:fiber lasers
Average power (single fiber, 2,000-nm)	1 kW	Available pump power	Tm:fiber laser pumped by 790-nm diode lasers
Average power (single fiber, 1 frequency, 1,000-nm)	100 W (5 kHz linewidth) 1 kW (3 GHz linewidth)	Stimulated Brillouin scattering	Yb:fiber lasers. Removed SBS by … 100-W result: thermal gradient along length of fiber. 1 kW result: frequency modulated source
Average power (single fiber, 1 frequency, 2,000-nm)	600 W (<5 MHz linewidth)	Stimulated Brillouin scattering	Tm:fiber laser
Peak power (single fiber, ns pulses)	0.45 MW (27 mJ in 60 ns) 4.5 MW (4.3 mJ; <1 ns)	Simulated Raman scattering, optical breakdown	Yb:silica, rod-type fibers. Flexible fibers typically generate 25 kW peak powers in 400 ns
Peak power (single fiber, ps range)	3.8 GW (2.2 mJ in 0.5 ps)	Simulated Raman scattering, optical breakdown, self-phase modulation	Yb:silica, employs chirped-pulse amplification (CPA) to avoid nonlinear effects, rod-type fibers. Flexible fibers with CPA operate around 100 MW of peak power

TABLE C-11 Nonlinear-Optics-Based Sources: Harmonic Generation

Property	Value	Limit Reason	Comments
Conversion efficiency (theory)	100%	Fundamental: Energy conservation	For plane waves
Second-harmonic conversion efficiency (actual)	90% for flat-profile, high-energy beams (e.g., NIF) 60-70% for Gaussian-profile beam	Multiple	De-phasing from finite beam width and heating in crystal due to background and multi-photon absorption and creation of color centers, losses at entrance/exit faces, crystal, material imperfections, optical damage.
Third-harmonic conversion efficiency (actual)	90% for flat-profile, high-energy beams (e.g., NIF) 50% for Gaussian-profile beam	Multiple	De-phasing from finite beam width and heating in crystal due to background and multi-photon absorption and creation of color centers, losses at entrance/exit faces, crystal, material imperfections, optical damage.
Average power output (theory)	Unlimited	Process does not dissipate heat	Nonlinear process has limited power density due to optical damage.
Average power output (actual)	Widely variable	Background, multi-photon and coating absorption, color-center formation.	LBO material has the lowest absorption of common nonlinear materials, and is limited by absorption from coatings on the surface. Multiple hundreds of W for second-harmonic of Nd- or Yb-doped lasers.
Shortest wavelength, second harmonic	176 nm	Phase-match, vanishing nonlinear coefficients	176 nm for KBBF crystals. For more readily available BBO crystals, about 205 nm

TABLE C-12 Nonlinear-Optics-Based Sources: Optical Parametric Generation

Property	Value	Limit Reason	Comments
Conversion efficiency from pump to signal + idler (theory)	100%	Fundamental: energy conservation	For plane waves
Conversion efficiency from pump to signal power	Pump ÷ Signal wavelengths	Fundamental: photon conservation (Manley-Rowe)	For plane waves
Conversion efficiency from pump to idler power	Pump ÷ Idler wavelengths	Fundamental: photon conservation (Manley-Rowe)	For plane waves
Conversion efficiency from pump to signal + idler (actual)	90% for cw 50% typical for pulsed Buildup time reduces efficiency	Multiple	De-phasing from finite beam size, de-phasing due to heating in crystal from background absorption or multi-photon absorption and creation of color centers, losses at entrance/exit faces, material imperfections, optical damage.
Average-power output (theory)	Unlimited	Process does not dissipate heat	Ultimate limit due to optical damage.
Average-power output (actual)	Widely variable	Absorption: from coatings, multi-photons, parasitics, color-center formation.	See effects for conversion efficiency Levels currently below 100 W, but have been limited more by the pump laser power. Multiples of 100 W should be possible in near-IR with materials like LBO