

Fundamentals of Terrestrial Millimeter-Wave and THz Remote Sensing

E.R. Brown

Professor of Electrical Engineering
University of California, Los Angeles
Los Angeles, CA 90095

Email: erbrown@ee.ucla.edu ; drebrown@earthlink.net

Abstract

Having long been the realm of molecular chemistry, astronomy, and plasma diagnostics, the upper millimeter-wave band (~100 to 300 GHz) and the THz region above it have recently become the subject of heightened activity in the engineering community because of exciting new technology (e.g., sub-picosecond optoelectronics) and promising new “terrestrial” applications (e.g., counter-terrorism and medical imaging). The most challenging of these applications are arguably those that demand remote sensing at a stand-off of roughly 10 m or more between the target and the sensor system. As in any other spectral region, remote sensing in the THz region brings up the complex issues of sensor modality and architecture, free-space electromagnetic effects and components, transmit and receive electronics, signal processing, and atmospheric propagation. Unlike other spectral regions, there is not much literature that addresses these issues from a conceptual or system-engineering viewpoint. So a key theme of this chapter is to review or derive the essential engineering concepts in a comprehensive fashion, starting with fundamental principles of electromagnetics, quantum mechanics, and signal processing, and building up to trade-off formulations using system-level metrics such as noise-equivalent power and receiver operating characteristics. A secondary theme is to elucidate aspects of the THz region and its incumbent technology that are unique, whether advantageous or disadvantageous, relative to other spectral regions. The end goal is to provide a useful tutorial for graduate students or practicing engineers considering the upper mm-wave or THz regions for system research or development.

Keywords

Active and passive sensor, direct and coherent detection, heterodyne and homodyne receiver; blackbody and thermal radiation, Planck distribution, Rayleigh-Jeans limit, brightness temperature; atmospheric propagation, water vapor absorption, PCLnWin, HITRAN96 database; waveguide, feedhorn, planar antenna, parabolic dish, lenses, hyperhemisphere, antenna theorem, antenna directivity and gain, effective aperture, diffraction limit; Gaussian beams, beam waist, Rayleigh length, ABCD matrix; radiation noise, shot and thermal noise, quantum noise; classical and quantum mixer, classical and quantum square-law detection, bolometer, Golay cell, Schottky diode, low-noise amplifier; coupling and quantum efficiency, responsivity, conversion gain, noise figure, RF, IF, and post-detection bandwidth; Gaussian, Boltzmann, Poisson, Rayleigh, and Rician statistics; signal-to-noise ratio (SNR), noise-equivalent power (NEP), noise-equivalent delta temperature (NE Δ T), probability of detection, probability of false alarm, receiver operating characteristics (ROC); Poynting’s theorem, Friis’ transmission formula, Johnson-Nyquist theorem, Nyquist’s sampling theorem, North’s theorem, Whittaker-Shannon theorem.

I.	INTRODUCTION	5
A.	Background and Purpose.....	5
B.	Sensor systems	6
C.	THz Solid-State Devices: A Longstanding Challenge	7
II.	THZ RADIATION.....	10
A.	Poynting’s Theorem: Classical Picture	10
B.	Harmonic-Oscillator States: Quantum Picture	11
C.	Types of Radiation	12
C.1	Coherent Radiation	12
C.2	Thermal Radiation.....	13
C.2	The Special Case of Blackbody Radiation	15
D.	Fluctuations of Radiation	17
D.1	Quantum Derivation	17
D.2.	Coherent Radiation: Poissonian photon statistics	19
D.3	Incoherent or Thermal radiation: Boltzman Photon Statistics	20
D.4	Low-Frequency Limit: The Johnson-Nyquist Theorem.....	20
D.5.	High-Frequency Limit: Photon Shot Noise.....	21
E.	THz propagation.....	22
E.1	Atmospheric absorption	22
E.3.	Effect on THz signals: Radiative transfer.....	23
E.3.	Atmospheric scattering.....	25
F.	Measurement of THz power.....	25
III.	COUPLING OF THZ SENSORS TO FREE SPACE	26
A.	Routing between devices: THz Transmission Lines	26
A.1.	Metallic waveguides	26
B.	Coupling to free space: THz Antennas.....	27
B.1.	THz wire-like antennas	27
B.2.	THz aperture-like antennas: Dishes, Lenses, and Feedhorns	30
C.	The Number of Modes: Antenna Theorem and Interpretation of Planck’s Law	35
D.	Illustrative Examples of Modal Function.....	36
E.	Gaussian beams.....	38
E.1.	Gaussian-beam methodology	38
E.2.	Gaussian-beam formulation	39
E.3.	Transformation of Gaussian beams: An representative system example	42
F.	Friis’ Free-Space Transmission Equation	44
IV.	THZ RECEIVER TYPES AND PERFORMANCE METRICS.....	45
A.	Architectures	45
A.1.	Direct	46
A.2.	Preamplified Direct.....	46
A.3.	Heterodyne.....	47
A.4.	Pre-Amplified Heterodyne.....	47
B.	Signal-to-Noise Ratio and Noise Equivalent Power.....	47
C.	Noise Equivalent Power	48

D.	Noise Equivalent Delta Temperature	49
E.	Noise Figure and Friis Formula.....	50
V.	THZ SIGNAL AND NOISE PROCESSING.....	50
A.	Classical Square-Law Detection and Integration	51
B.	Quantum Square-Law Detection and Integration.....	53
C.	Classical Heterodyne and Homodyne Conversion and Detection	54
D.	Quantum Coherent Heterodyne and Homodyne Conversion and Detection	56
VI.	RADIATION-NOISE LIMITS ON SENSITIVITY	57
A.	Radiation-Noise-Limited NEP_{BD} of THz Direct-Detection Sensors	57
A.1.	Coherent Signal.....	57
A.2.	Thermal Signal.....	57
A.3.	Arbitrary Signal, Thermal Background.....	59
B.	Radiation-Noise-Limited NEP_{BD} of Classical and Quantum Coherent Receivers	62
VII.	PRACTICAL LIMITS ON RECEIVER SENSITIVITY: ELECTRONIC NOISE	62
A.	Physical Noise of Electronics.....	63
B.	Equivalent circuit representations	63
C.	Electrical Noise Limitations on THz Square-Law Detector.....	64
D.	Electrical Noise Limitation on Heterodyne Mixer	66
VIII.	RECEIVER PERFORMANCE LIMITATIONS AND STATISTICS.....	67
A.	Optimum signal-to-noise ratio before detection: North's Theorem	67
B.	Sensor Performance after Detection: Receiver Operating Characteristics	68
B.1.	Direct Detection Sensor	70
B.2.	Heterodyne (or Homodyne) Sensor.....	71
IX.	OVERALL PERFORMANCE OF FOUR TYPES OF PASSIVE SENSORS	72
A.	Sensor Scenario	73
B.	Direct Detection with Classical Square-Law Detector.....	74
B.1.	General Analysis.....	74
B.2.	Detector-Specific Analysis.....	77
B.3.	An Illustrative Example: Schottky-Diode Direct Detector.....	78
C.	Heterodyne	81
D.	Pre-Amplified Direct.....	82
E.	Pre-amplified Heterodyne	85
X.	ISSUES AND PERFORMANCE OF ACTIVE SENSORS.....	86
A.	Active Sensor Scenario	87
B.	Transmitter Types	89
B.1.	Continuous Wave.....	89
B.2.	Pulsed: T-Rays	89
C.	Receiver Types.....	90
C.1.	Direct Conversion with Classical Square-Law Detector.....	90
C.2.	Heterodyne and Homodyne Conversion with Classical Square-Law Detector	91
D.	An Illustrative Example: Design of an active sensor to measure absorption signatures	93
D.1.	Absorption Signature Characteristics.....	94

D.2. Sensor Design: Direct and Homodyne Differential Absorption Radars	95
D.3. Direct and Homodyne Sensor Performance at 425 GHz.....	97
XI. 2D IMAGING AND THE QUEST FOR POPULAR APPLICATIONS.....	99
A. Heuristics	99
B. Imaging by Spatial Sampling	101
C. Applications and Examples of MM-Wave Imagers	102
ACKNOWLEDGEMENT AND REFERENCES	103

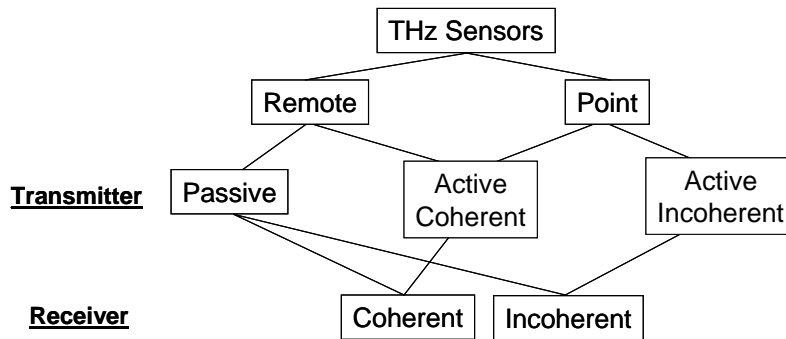


Fig. 1. Hierarchy of sensor technology in the mm-wave and THz regions

I. Introduction

A. Background and Purpose

The author has been in and around the field of remote sensing for over 20 years and over this period has observed a significant decline in the education and understanding of this topic, particularly since the end of the U.S./Soviet Cold War around 1990. There was little challenge finding interest and support for remote sensing when the biggest threat to U.S. National Security was intercontinental and theater ballistic missiles. This threat alone created entire industries to produce high-performance RF electronics, electromagnetic components, digital signal processing electronics, and signal-processing software, to name a few. Now, with the threats of terrorism, global warming, and bio-catastrophe, the field of remote sensing is experiencing a resurgence. And even though the threat ranges have shrunk from thousands of kilometers to thousands of meters and the threat agents have shrunk from missiles to molecules, many of the principles and methodologies developed during the Cold-War period should be applicable. A key motivation for this chapter is to provide a primer for students or young engineers or scientists considering remote sensing as a research topic or as a career. In the limited space of one book chapter, it is impossible to cover this topic in a comprehensive fashion. Therefore, it will focus on fundamental principles behind system architectures and remote sensing in the millimeter-wave and THz regions, which many consider to be at the upper end of the useful RF spectrum.

A secondary purpose for this chapter is to provide some quantitative contrast of the various remote-sensor architectures. There have been several new approaches demonstrated during the past decade that fall outside the traditional types of mm-wave and THz remote sensors. For example, an active sensor has been developed to detect concealed weapons, and passive focal plane arrays have been developed for all-weather imaging. The present chapter will contrast these primarily within a sensor system context, not at the device or component performance. In spite of its limited popularity, the THz region offers several new sensor applications such as the remote detection of

airborne toxic bioparticles, the identification of concealed weapons and contraband, and all-weather aircraft landing.

By focusing on the sensor principles and architectures, a sacrifice will be made in terms of technological detail. There is simply not enough space to cover all the electromagnetics, solid-state physics, and signal processing associated with the myriad of components and devices being used today in THz sensors. There is also not enough space to give proper credit to some of the emerging THz technologies that have shown great results in the laboratory but have not yet been used for remote sensing. Principal among these are the optoelectronic technologies, such as THz generation by ultrafast photoconductive switches and mixers using solid-state and semiconductor lasers operating mode-locked or cw at near-infrared wavelengths. One of these techniques, called “T-ray imaging,” has advanced to the point of commercial production. Fortunately, several excellent review articles have already been published on both the traditional^{1,2,3} and optoelectronic technologies,^{4,5,6} and the reader is referred to one of these articles for more information.

B. Sensor systems

The THz region of the electromagnetic spectrum has long been the realm of basic sciences, such as molecular chemistry and astrophysics, but has not been broadly utilized for commercial or military systems because of the lack of transceiver technology and because of the strong attenuation of THz radiation through the terrestrial atmosphere. Historically, the most common THz remote sensors have been passive radiometers utilizing coherent (i.e., heterodyne) detection. The most common systems for point sensors have been Fourier transform spectrometers utilizing incoherent transmitters and receivers.

As in other regions of the electromagnetic spectrum, scientists and engineers have long sought to construct systems in the THz region to *sense* the propagation through or reflection from objects or regions of space having scientific or technical interest. Such *sensor* systems are generally classified according to three criteria which together define the sensor *modality*. They are: (1) the proximity of the sensor and the object of interest, (2) the source of THz radiation or, equivalently, the design of the *transmitter*, and (3) the sink of the THz radiation or, equivalently, the design of the *receiver*. These three criteria are shown in the sensor hierarchy of Fig. 1. In point sensors, the object or material of interest (henceforth referred to as the *target*) is located in close proximity to the sensor and the THz radiation is provided by a transmitter. The transmitter can be either coherent or incoherent. A good example of a THz point sensor with an incoherent transmitter is a Fourier transform spectrometer having a hot incandescent or discharge-tube source. A good example of a THz point sensor having a coherent source is a network analyzer, be it vector or scalar. Because of the proximity, practically all of the transmitted power can be directed through or off-of the target, so little transmitted power is required.

In remote, or stand-off sensors, there is a relatively large separation between the target and the sensor. The separation generally means that only a small fraction of the transmitted power can be delivered usefully to the target. If this small fraction can be detected back at the receiver, the modality is called *active* and the resulting sensor is called a *radar* (*radio detection and ranging*). It is fascinating that the radar concept was first proposed in the early part of the 20th century, promoted by none other than G.

Marconi, and was successfully developed into working systems even more quickly than wireless communications. But in the THz region the relatively low power of sources and the high opacity of the atmosphere has hindered radar development, and system developers have relied on a second modality called *passive*. This generally entails the illumination of the object or region by incoherent radiation from its surroundings, which if intense enough can be detected by the remote receiver. A good example of a passive THz remote sensor is the total power radiometer often used to measure trace gases in the atmosphere or emissions from interstellar media. The advantage of a passive system is clearly that the source radiation comes for free. The downside is that the atmospheric attenuation and fading effects can quickly degrade the sensor performance.

The third defining criterion for sensor modality is the receiver architecture. As in the transmitter there are two types, coherent and incoherent, both discussed later in Sec. IV. Within the coherent category there are two types: (1) heterodyne and homodyne. Within the incoherent category there are also two types: (1) power detection and photon detection. As shown in Fig. 1, a coherent transmitter is generally coupled to a coherent receiver, and an incoherent transmitter is generally coupled to an incoherent receiver. It is surprising to some people that a passive system can have either receiver type. A coherent receiver measuring an incoherent signal (e.g., white noise) essentially filters out that particular Fourier component of the signal that *coheres* to the local oscillator waveform. In other words, it cross-correlates with only one space-time mode of the otherwise random spectrum of incoming radiation.

One might on first glance think that a passive THz sensor operating at temperature T and relying on power received via thermal radiation also of temperature T is a violation of the first law of thermodynamics. After all, this law would require that in equilibrium the power transferred from the object to the sensor should be matched by a power transferred from the sensor to the object. But THz sensors, like RF systems in general, are categorically not in thermodynamic equilibrium! A good way to see this point is to think about the fact that the first active component (i.e., one capable of doing or receiving electromagnetic *work*) in most RF receivers is a low-noise amplifier (LNA). Modern solid-state LNAs operating up to several GHz have very low noise figures and correspondingly low noise temperatures, typically in the range of 100 K or below. This means that they produce noise equivalent to a device at much lower temperature, made possible by the fact that the LNA is an active device operating well out of thermodynamic equilibrium. This is why active devices are so pervasive in all electromagnetic sensors, both in the transmitter and the receiver.

C. THz Solid-State Devices: A Longstanding Challenge

No summary on the THz region would be complete without some discussion of its technology, ranging from passive quasi-optical components to quantum-effect solid-state detectors and unique coherent sources. Because many articles in recent years have reviewed this technology thoroughly, the discussion here will be kept to a minimum. And in keeping with the author's viewpoint on research, this summary occurs after the discussion of phenomenology and system-level issues. The development of new technology with a system perspective allows the engineer or scientist to design and characterize components and devices with far greater insight and practicality than otherwise.

Because of the relative lack of components in the THz region, some researchers have referred to it as the electromagnetic frontier. Although a bit misleading, this statement is not without justification. The THz region has presented pioneering investigators with one experience common to most “frontiers”, namely technical difficulty. The difficulty is based in two well-known but seldom discussed physical characteristics of operating any system in the THz region at room temperature. First, in the THz region the photon energy $h\nu$ is much less than $k_B T$ around room temperature, where k_B is Boltzman’s constant. This is in contrast to systems operating at near-infrared and shorter wavelengths. Then according to the Boltzmann *canonical* distribution, quantized states separated by $h\nu$ tend to have the same population. Hence, stimulated absorption is nearly as likely as stimulated emission. This affects the ability to measure or utilize atomic and molecular transitions at terrestrial temperatures and pressures.

Second, the THz region tends to be high enough in frequency that the classical behavior (i.e., drift and diffusion) of electrons and holes in semiconductors begins to disappear, and the coupling between electromagnetic waves and semiconductor devices weakens. In short, at frequencies for which $\omega\tau \geq 1$ an electromagnetic wave is oscillating too fast for free carriers to respond. A more precise way to understand this effect is to recall from classical electromagnetics that the energy transfer from an electromagnetic wave to free carriers is described by $\overline{\mathbf{J} \cdot \mathbf{E}}$ where $\mathbf{J} \cdot \mathbf{E}$ is the Joule term of Poyning’s theorem and denotes the time average over many cycles of the electromagnetic fields. When the $\omega\tau > 1$, the free carrier current density \mathbf{J} according to (1) approaches a quadrature relationship to the electric field so that $\overline{\mathbf{J} \cdot \mathbf{E}} \rightarrow \overline{\cos \omega t \cdot \sin \omega t} = 0$. The incoming electromagnetic wave that every device must respond to simply cannot transfer energy into the free carriers, and the device function disappears.

A good example of this is the channel conductance of any field effect transistor (FET). Be it a 3D, 2D, or 1D confinement, the transport of free carriers in the channel is governed by their concentration and their ac conductivity. In the simple model of ac conductivity put forth by Drude, this is given by

$$\sigma = \frac{\sigma_0}{1 + j\omega\tau}$$

where σ_0 is the dc conductivity and τ is the momentum relaxation time. As an example, we consider electrons in GaAs which at room temperature have a mobility of approximately $6000 \text{ cm}^2/\text{V}\cdot\text{s}$ and an effective mass m^* of $0.067m_0$. According to kinetic theory, the mobility is related to τ by $\mu = e\tau/m^*$, so that under these conditions $\tau \approx 0.23 \text{ ps}$. The corresponding 3-dB conductivity frequency is $f = (2\pi\tau)^{-1} = 0.69 \text{ THz}$. Of course, under device operating conditions the large internal electric fields make the above small-field approximations less accurate.

A second example is the bipolar junction transistor (BJT). Whether constructed with homo- or hetero- p-n junctions, a BJT is also ultimately limited in speed by an RC time constant. A leading contribution to R is the base majority carrier resistance, which depends on the majority carrier concentration and ac resistivity. Again, in the THz region, the resistivity begins to increase significantly. The second phenomenon that greatly affects electronics is that in solid-state material the THz radiation generally obeys $\omega\tau_c > 1$, where τ_c is the momentum relaxation time, and generates an electrical photocurrent at the difference frequency. The difference-frequency current, in turn, generates THz power by connecting it to a suitable load, such as a planar antenna.

Nevertheless, several technological breakthroughs have occurred during the past decade that make THz sensors more practical than ever before. First there has been a rapid development of THz detector and mixer technology including Schottky varistive mixers, superconducting-insulating-superconductor tunnel junction mixers, and superconducting hot electron bolometers. Second, there has been a parallel rapid growth of THz solid-state source technology including Schottky varactor multipliers. Third, there has been a rapid development of ancillary technologies in the millimeter-wave and near-infrared regions to support the THz sources. In particular, there has been a steady development of monolithic microwave- and millimeter-wave integrated circuits (MMICs) up to around 100 GHz. Specifically, MMICs solid-state power amplifiers are now used to drive Schottky multipliers much more efficiently than ever before. Second: (1) growth and fabrication of semiconducting material having photocarrier lifetime less than 1 ps, (2) modern microfabrication techniques that allow sub-micron electrode features to be patterned on the photoconductor surface, leading to subpicosecond electrical time constants, and (3) integration of photoconductive elements with compact planar antennas, leading to efficient coupling of the THz radiation to free space.

An ancillary breakthrough that strongly supports the optoelectronic THz generation approach has occurred in the field of solid-state and semiconductor lasers. Solid-state materials such as $\text{Ti:Al}_2\text{O}_3$, have been developed that provide unprecedented values of gain-bandwidth and can provide high levels of power tunable over 10s of nm. Various techniques such as distributed Bragg reflectors, distributed feedback structures, and external cavities have all been integrated with semiconductor laser diodes to produce sources with useful output power (>1 mW) and high spectral purity. And in the popular fiber-optic telecommunication band around 1550 nm, the erbium-doped fiber amplifier (EDFA) has been developed that can boost the power of spectrally-pure laser-diode sources up to ~ 1 -W level.

A key advantage of THz sensors over infrared and visible ones is ultimate noise limits. As will be shown later in Sec. VI, both direct and coherent receivers operate against fundamental noise limits that depend on the background radiation and the photon frequency. In the coherent case, the limit is simply the photon shot noise that has noise equivalent power equal to $h\nu/\eta$, where η is the receiver coupling efficiency. Fig. 2 shows this limit plotted for $\eta = 1$ in terms of the photon energy $h\nu/e$ and equivalent temperature, $h\nu/k_B$. Of course, this same advantage is shared by RF receivers operating at lower frequencies, which is one of the reasons RF communications, be it wired or wireless, is generally superior to photonic communications in terms of sensitivity.

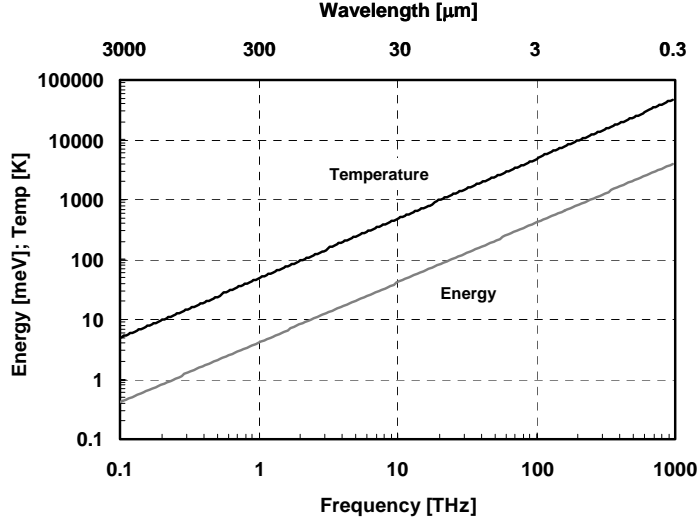


Fig. 2. Quantum-limit defined by the minimum energy – one photon – per spatial mode from the mm-wave region through the visible region of the electromagnetic spectrum.

II. THz Radiation

The THz portion of the electromagnetic spectrum is loosely defined as that region between approximately 300 GHz and 3 THz. As such it overlaps two older regions – the submillimeter-wave region between 300 GHz and 3 THz and the far-infrared region between 3 and 30 THz. Although, lagging behind the RF, infrared, and visible regions in terms of technological maturity, but is still subject to the same laws of electromagnetics (i.e., Maxwell’s equation), quantum mechanics (i.e., photon effects), and radiative transport as any other region.

A. Poynting’s Theorem: Classical Picture

One of the profound results of electromagnetic theory (i.e., Maxwell’s equations) is the existence of traveling or propagating waves, which must always be associated with two quantities, E and H – the electric and magnetic field vectors. An important theorem pertaining to propagating waves is the energy, or Poynting’s theorem, which states that the propagating instantaneous power associated with the wave is given by $\vec{S} \equiv \vec{E} \times \vec{H}$. S represents the instantaneous power in the wave, and is applicable no matter what its time or space dependence. Since E and H are always orthogonal in free-space, $H = E/z_0$, and

$$\vec{S} \equiv \frac{E(t)^2}{\eta_0} \hat{s}$$

where z_0 is the intrinsic impedance of free space (377Ω) and \hat{s} is the unit vector along the direction of S.

Poynting's vector and radiation propagation become particularly simple in the special case of sinusoidal time dependence, sometimes called harmonic waves. The electric and magnetic field vectors, E and H, then have solutions of the form $\vec{E} = \text{Re}\{\tilde{E}e^{j\omega t}\}$ and $\vec{H} = \text{Re}\{\tilde{H}e^{j\omega t}\}$, where the wig over denotes a phasor (complex, time-independent), quantity. And the time-averaged power is given by $\vec{S} \cdot \vec{A} = 1/2 \text{Re}\{\vec{E} \times \vec{H}^*\} \cdot \vec{A}$.

An important quantity for sensors is the time-averaged power flowing into the sensor aperture, $P_{inc} = \vec{S} \cdot \vec{A} = (1/2) \text{Re}\{\vec{E} \times \vec{H}\} \cdot \vec{A}$, where \vec{A} is the sensor areal vector (pointed perpendicular to the sensor surface). An even more important quantity is the power *usefully* absorbed,

$$P_{abs} \equiv \eta \cdot P_{inc} = \frac{1}{2} \eta \cdot \text{Re}\{\vec{E} \times \vec{H}\} \cdot \vec{A} \quad (1)$$

and η is the power coupling efficiency. Since η is the fraction of incident power *absorbed*, it must account for the effects of reflection at the environment-sensor interface, unabsorbed radiation that passes through the sensor, etc. The majority of sensors couple radiation in from free space propagating perpendicular to the surface. In this case \vec{H} is perpendicular to \vec{E} , $|\vec{H}| = |\vec{E}| / z_0$, and

$$P_{inc} = 1/2(E^2 A)/z_0 = 1/2(\epsilon_0 c E^2 A) = c U_E A, \quad (2)$$

so that

$$P_{abs} = \eta c U_E A \quad (3)$$

where U_E is the energy density for the electric field and c is the speed of light in vacuum.

B. Harmonic-Oscillator States: Quantum Picture

When the traveling wave phenomenon is addressed in quantum mechanics, the oscillatory nature in the classical picture is captured by the harmonic oscillator model in the quantum picture. The reader is assumed to understand the important derivation, so here we only state the result for the energy eigenvalues for a harmonic oscillator,

$$U_n = (n + 1/2) \cdot h \nu \quad (4)$$

In this expression, n is a positive integer starting at zero, h is Planck's constant, ν is the oscillator natural frequency, and the $1/2$ is the "zero-point" energy, i.e., the residual oscillator energy eigenvalue when $n = 0$. As taught in basic quantum mechanics courses, the zero-point energy can not do useful work. For example, it can not be used to transmit information in a radar or communications system.

C. Types of Radiation

C.1 Coherent Radiation

The classical sinusoidal waveforms are both replaced by a wave function called the *coherent state* of frequency $\nu = \omega/2\pi$. The amplitude, or occupancy, of this state corresponds to the instantaneous power associated with the classical field amplitudes E or H, and the occupancy number represents the number of *photons* in this state. Each photon has energy $h\nu$, where h is Planck's constant. The coherent state is emitted by oscillators, be they electronic resonator-based sources or atomic lasers.

A necessary aspect of the quantum-mechanical picture is fluctuations associated with the sensor measurement process. If one had a detector fast and sensitive enough to measure the individual photons, one finds that the photon number in the coherent state itself is random and obeys Poisson statistics. In other words, the probability of measuring n photons in the mode in an arbitrary time interval is given by

$$p(n) = \frac{\langle n \rangle^n}{n!} e^{-\langle n \rangle}$$

where $\langle n \rangle$ is the mean number of photons measured in this same interval over many different measurements. It is simple to verify that this distribution is normalized (i.e., $\sum_0^{\infty} p(n) = 1$) and that $\sum_0^{\infty} np(n) = \langle n \rangle$.

This may first appear to be contradictory to what we know about coherent radiation – namely, a sinusoidal dependence of the field quantities (E and H) at frequency ω and a $\sin^2(\omega t)$ dependence of the instantaneous power. This applies to a coherent state as well, but in terms of the mean photon number in the mode ν . In other words, the Poisson distribution gets replaced by

$$p(n) = \frac{[\langle n \rangle(t)]^n}{n!} e^{-\langle n \rangle(t)}$$

where $\langle n \rangle(t) \propto \sin^2(\omega t)$, the same time dependence as the instantaneous intensity.

A lucid discussion of the coherent state and its statistical properties is found in Ref. ⁷

Given the photon picture, an equivalent way to represent a coherent wave and a sensor is through the average measured photon rate J_p , defined as the average number of photons usefully absorbed by the sensor per unit time. From Eqn 3 This is given simply by

$$J_p = \frac{P_{abs}}{h\nu} = \frac{c \cdot \eta \cdot U_E \cdot A}{h\nu}$$

where h is Planck's constant and ν is the frequency. Even for the relatively weak THz coherent sources, this flux is astronomically high. For example, a source putting out 1 μ W at 600 GHz ($h\nu = 2.48$ meV) is emitting a photon rate of 2.5×10^{15} photon/s !

A practical issue associated with any real sensor in the THz region is that the limited detector bandwidth does not allow for resolution of individual photons at such a high rate. In practice, the average number of photons is given by the sensor resolution time δt , and the mean number of photons measured during this interval is

$$\langle n \rangle = J_p \cdot \delta t = \frac{P_{abs} \cdot \delta t}{h\nu} \approx \frac{P_{abs}}{\delta f \cdot h\nu}$$

where δf is the measurement bandwidth.

C.2 Thermal Radiation

Thermal radiation results from a volume of electromagnetic modes coming into thermal equilibrium with a heat bath at temperature T . In the terrestrial environment of the Earth, thermal radiation is omnipresent and tends to be very strong in the THz region because of absorption by water vapor. To get a quantitative expression for the average power received by a sensor exposed to thermal radiation, it is best to start with the quantum picture in which the radiation is described not by the electric field and intensity (Poynting) vector, but rather by the occupation of each quantized electromagnetic mode. The occupation probability is then given by the Boltzman distribution $p(n) = C \exp[-$

$U_n/kT]$ where C is a normalization factor, $C = \sum_{n=0}^{\infty} e^{-U_n/kT}$, and k_B is Boltzman's constant [1.38×10^{-23} W/K, MKS]. The mean occupation is then given by the Planck function

$$\langle n(\nu) \rangle = \frac{1}{e^{h\nu/k_B T} - 1} \equiv f_P$$

where k_B is the Boltzman constant and T is the temperature of the bath. And the mean energy is given by

$$\langle U \rangle = (\langle n \rangle + 1/2) \cdot h\nu = (f_P + 1/2) \cdot h\nu$$

So if a sensor is pointed at a thermal source, such as a highly attenuating sky, how much of the radiation is incident on the sensor? The answer lies in electromagnetic modal theory and in thermodynamics. The plane wave modes are the most convenient for decomposing free-space radiation for analytic purposes. However, they are not physically realizable in any receiver because of their infinite lateral spatial extent. More convenient modes are the spatially-orthogonal set defined by the sensor antenna. These so called lateral or "spatial" modes can be used to decompose any radiation transmitted or received by the antenna provided they are defined properly in the antenna coordinate system. A remarkable concept from statistical mechanics is that the Planck function is valid for any orthogonal set of modes, no matter what antenna they apply to. So we arrive at the result that the mean thermal energy *incident* from free space at frequency ν is just the energy quantum, $h\nu$, times the mean number of photons in that mode, summed over all spatial modes

$$\langle U \rangle = \sum_{m=1}^M f_P(\nu) \cdot h\nu$$

where m is the spatial mode index, and the zero-point has been excluded because it can not couple energy on the average. Note that the maximum number of spatial modes M is shown explicitly, representative of the fact that many THz sensors accept well more than one spatial mode but rarely enough that the summation can be approximated by a spatial integral as is usually done in the infrared and visible regions.

What happens if we fix the spatial mode and change the frequency? Each frequency in the Planck function corresponds to a unique harmonic oscillator and, therefore, to a unique mode. In the language of lasers, the different frequencies correspond to different "longitudinal" modes. Therefore, the total energy in a frequency

range $\Delta\nu$ is to be thought of as a sum over all the possible longitudinal modes for each lateral mode

$$\langle U \rangle = \sum_{m=1}^M \sum_{n=1}^N f_p(\nu) \cdot h\nu$$

where n is the longitudinal-mode index. A simple way to estimate the number of longitudinal modes is to assume that the thermal radiation is separated from the sensor antenna by a distance L and that boundary conditions require that the electromagnetic intensity be a maximum at both the radiator and sensor. Then the lowest frequency longitudinal mode corresponds to a half-wavelength between the two, $\nu_{\min} = c/2L$. In open-cavity lasers and Fabry-Perot resonators, this quantity is called the free spectral range. We also assume that the sensor is filtered so that it responds only to radiation lying within a “passband” ν_0 to $\nu_0 + \Delta\nu$, and that the sensor responds only to the half of the longitudinal modes propagating in the direction from the source. The number of longitudinal modes $N(\nu)$ that exist at each ν is then given by

$$N(\nu) = \frac{1}{2} \cdot \frac{\nu - \nu_0}{c/2L}$$

And the mean electromagnetic energy is given by

$$\langle U \rangle = \sum_{\nu_0}^{\nu_0 + \Delta\nu} \sum_{m=1}^M N(\nu) \cdot h\nu \cdot f_p(\nu)$$

Remote sensors are usually configured so that targets and radiators are both in the far field of the transmit and receive antennas, and hence $c/2L \ll \nu_0$, where ν_0 is the bottom of signal passband. In this case the sum over n is quasi-continuous and we can approximate the sum by an integral

$$\langle U \rangle = \sum_{m=1}^M \int_{\nu_0}^{\nu_0 + \Delta\nu} f_p(\nu) \cdot g(\nu) \cdot h\nu \cdot d\nu = \sum_{m=1}^M \frac{L}{c} \int_{\nu_0}^{\nu_0 + \Delta\nu} f_p(\nu) \cdot h\nu \cdot d\nu$$

where $g(\nu) = dN/d\nu = L/c$ is the density of forward-traveling states.

We can now estimate the average energy density $\langle U' \rangle$ just in front of the antenna aperture of area A ,

$$\langle U' \rangle = \frac{1}{AL} \langle U \rangle = \sum_{m=1}^M \frac{1}{cA} \int_{\nu_0}^{\nu_0 + \Delta\nu} f_p(\nu) \cdot h\nu \cdot d\nu ,$$

and from Eqn 2 the ensemble-averaged power received by the sensor antenna becomes

$$\langle P_{inc} \rangle = \sum_{m=1}^M \int_{\nu_0}^{\nu_0 + \Delta\nu} f_p(\nu) \cdot h\nu \cdot d\nu .$$

Using the coupling (Eqn 1) coefficient from the classical analysis, we find the power absorbed by the sensor is

$$\langle P_{abs} \rangle = \sum_{m=1}^M \int_{\nu_0}^{\nu_0 + \Delta\nu} \eta_m(\nu) f_p(\nu) \cdot h\nu \cdot d\nu \quad (5)$$

Remarkably, this expression has no dependence on c and no direct dependence on the antenna area, although this is included implicitly through the number of spatial

modes contained in the sum as we shall see in Sec. III.C. In addition, it is a straightforward exercise in statistical mechanics to show that it *can be generalized to any form of radiation statistics between the source and the sensor*, so that

$$\langle P_{inc} \rangle = \sum_m^M \int_{\nu_0}^{\nu_0+\Delta\nu} h\nu \langle n_m(\nu) \rangle \cdot d\nu \quad (6)$$

$$\langle P_{abs} \rangle = \sum_m^M \int_{\nu_0}^{\nu_0+\Delta\nu} \eta_m(\nu) \cdot h\nu \langle n_m(\nu) \rangle \cdot d\nu$$

And we show explicitly the dependence of n on m since other possible radiation statistics, besides the Boltzmann distribution behind the Planck function, may show a dependence of occupancy on mode type.

In the THz region Eqn 5 for thermal radiation commonly gets applied under two limiting cases – high and low background thermal radiation. In the high background case, $T > h\nu/k_B$ over the entire band $\Delta\nu$, so that $e^{h\nu/k_B T} - 1 \approx h\nu/k_B T$. Let's suppose that a sensor is receiving thermal noise power in this limit in M spatial modes over an arbitrary spectral bandwidth $\Delta\nu$. Let's also assume that the electromagnetic waves associated with these modes are perfectly coupled to the sensor (i.e., all incident waves are absorbed without reflection). In this case, the average received power is the product of the average occupancy per mode $\langle n(\nu) \rangle$ times the energy per mode, $h\nu$, integrated over the longitudinal modes (i.e., the spectral band). So neglecting the zero-point term

$$\langle P_{abs} \rangle = M \int_{\nu_0}^{\nu_0+\Delta\nu} \langle n(\nu) \rangle h\nu \cdot d\nu = M \int_{\nu_0}^{\nu_0+\Delta\nu} k_B T \cdot d\nu = M \cdot k_B T \cdot \Delta\nu \quad (7)$$

This case, called the Rayleigh-Jeans limit, is very common at THz frequencies and below. For example, if $\nu = 1$ THz and $T = 290$ K (room temperature), we have $h\nu = 4.1$ meV, and $k_B T = 25.0$ meV, so that $h\nu/k_B T = 0.164$ and $e^{h\nu/k_B T} = 1.178$, and then the Rayleigh-Jeans approximation is accurate to about 9%. One interesting aspect of the Rayleigh-Jeans limit is that $\Delta\nu$ can be arbitrarily large if the highest frequency in the spectral band satisfies $h\nu/k_B T \ll 1$. So it is not limited to narrow-band sensors as is sometimes stated in the literature.

C.2 The Special Case of Blackbody Radiation

In the special case of uniform, isotropic thermal radiation in an enclosed cavity, the spatial modes can be counted exactly. This is a classic problem in elementary physics and results in the following expression commonly known as blackbody radiation law.

$$\frac{d|S|}{d\nu} \equiv I_\nu = \frac{2\pi h \nu^3 \cdot \epsilon}{c^2 (e^{h\nu/k_B T_s} - 1)}, \quad (8)$$

where T_S is the source physical temperature and ε is the source emissivity. This expression, which defines the commonly used optical quantity called the frequency specific irradiance I_ν , gives the randomly polarized power per unit area that crosses a flat surface of area A when illuminated from *one side* by isotropic blackbody radiation. It is very useful when a mm-wave or THz detector is exposed to thermal “background” radiation over a much greater solid angle than the desired incident signal. The Rayleigh-Jeans approximation then applies with the simple result (written in terms of the free-space wavelength, c/ν),

$$I_\nu = \frac{1}{A} \frac{dP_{inc}}{d\nu} = \frac{2\pi \cdot \varepsilon \cdot k_B T_S}{\lambda^2}.$$

A second form of the blackbody law is more useful when one is concerned with calculating the thermal radiation into a mm-wave or THz receiver over a narrow solid angle as defined, for example, by a receiver antenna. In this case, one quantifies the power per unit area emanating from the reference surface of area A in the direction perpendicular to the surface and over a solid angle $d\Omega$. The resulting *brightness* function is given by

$$B_\nu = \frac{2h\nu^3 \cdot \varepsilon}{c^2 (e^{h\nu/k_B T_S} - 1)}$$

Note that the irradiance derives from the brightness by integrating over a hemisphere and weighting by the Lambert factor. And in the Rayleigh-Jeans limit one obtains the following form

$$B_\nu = \frac{2 \cdot \varepsilon \cdot k_B T_S}{\lambda^2} \equiv \frac{2 \cdot k_B T_B}{\lambda^2} \quad (9)$$

Note the definition given here for the brightness temperature T_B includes the emissivity, which is not known for most sources. This is the temperature quantity most often used in remote sensing, particularly mm-wave and THz astronomy.

Fig. 3 shows brightness curves for three temperatures, 290 K, 120 K, and 40 K. The values $T = 120$ K and $T = 40$ K correspond to the typical sky equivalent blackbody temperature looking up from sea level along the zenith at 100 GHz and 70 GHz, respectively.⁸ Also shown are the linear curves for the Rayleigh-Jeans approximation. Clearly the latter is a good approximation at 290 K, deviating by at most a factor of 2.6 at the highest plotted frequency of 10 THz. But as the temperature decreases, the Rayleigh-Jeans becomes progressively worse, overshooting the peak that occurs in the THz region for temperatures roughly less than 150 K. The displacement of this peak to longer wavelength with decreasing temperature is the famous Wien law,⁹

$$\lambda_{max} T = \text{constant} = 5.098 \text{ [mm-K]}$$

For example, the peak brightness of a source having a frequency-independent ε and a physical temperature of 290 K is $17.6 \mu\text{m}$ (17.1 THz). This peak shifts to $\lambda_{max} = 127 \mu\text{m}$ (2.35 THz) for a 40-K source, and 1.89 mm (159 GHz) for a 2.7-K source. The latter temperature corresponds to the cosmic background brightness temperature – a remarkable discovery by radio astronomers and strong evidence for the “big bang” theory of the universe. The 159-GHz peak also suggests that the THz region is very important for cosmic radiative transfer – a key reason why much of the technology development in the THz region during the past 25 years or so has been driven by the astronomy application.

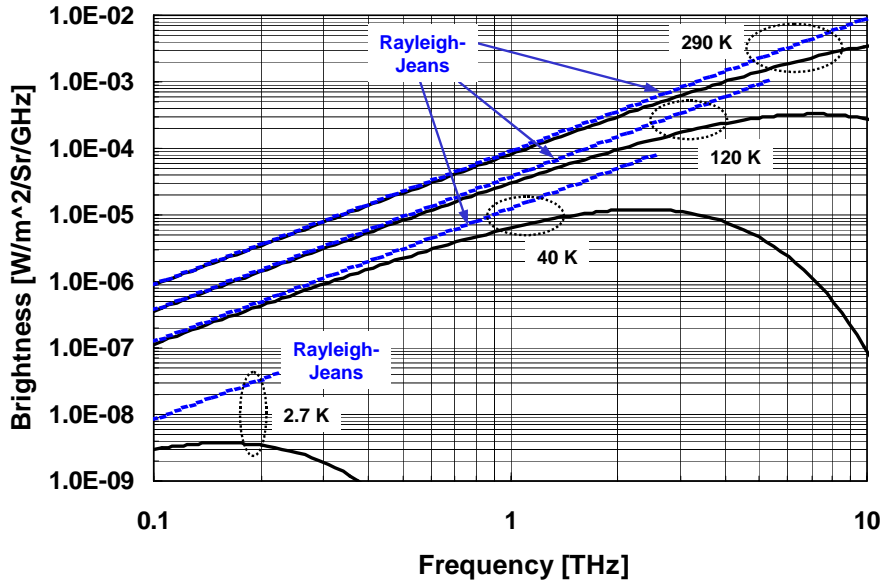


Fig. 3. Blackbody brightness spectra in THz region for different source temperatures.

D. Fluctuations of Radiation

Although most of the principles of electromagnetics necessary to understand THz sensors were understood prior to 1900, a few came later with the advent of quantum mechanics. Perhaps the most fundamental principle is the occurrence of fluctuations in the radiation. Radiation fluctuations associated with the sensing process does depend on the spectral region and has two limiting forms: classical (electric-field thermal-noise) fluctuations in the long-wavelength limit and corpuscular (photon shot noise) fluctuations in the short-wave limit. In the THz region, both types of radiation noise may be exhibited depending on the type of sensor, the type of radiation detected, and other factors. Since a quantitative understanding of radiation noise cannot be made without the tools of statistical mechanics and probability theory, this section starts out with a short review of the basic concepts in these two fields.

D.1 Quantum Derivation

Quantum mechanics teaches us that the act of measuring radiation, like the measurement of any other physical observable, is inherently probabilistic with a metric of uncertainty or deviation given by Planck's constant h . Even when measuring the power from an extremely coherent source, such as a laser or high-Q cavity oscillator, one expects a fluctuation in the measurement results depending on the type of radiation that one is measuring. In general, this is a complicated issue in quantum statistical mechanics or quantum electrodynamics, but the analysis becomes relatively simple in two extremes: (1) coherent radiation, or (2) thermal, or blackbody, radiation. In these cases, we learn that the most common description for radiation fluctuations is the Poisson statistics.

A convenient launch point for the discussion of all radiation noise is Eqn 6, which is generally applicable to any incident radiation of arbitrary photon statistics provided that the fluctuations are perfectly random. The fluctuation in the incident power around the average is then given by

$$\Delta P_{inc} \equiv P_{inc} - \langle P_{inc} \rangle = \sum_m^M \int_{\nu_0}^{\nu_0 + \Delta\nu} h\nu \cdot \{n_m(\nu) - \langle n_m(\nu) \rangle\} d\nu \equiv \sum_m^M \int_{\nu_0}^{\nu_0 + \Delta\nu} h\nu \cdot \Delta n_m(\nu) \cdot d\nu$$

where n is now to be regarded as a discrete random variable. Probability theory teaches us that an important measure of the fluctuations is the variance, or “mean-square” fluctuation

$$\langle (\Delta P_{inc})^2 \rangle = \langle \left(\sum_m^M \int_{\nu_0}^{\nu_0 + \Delta\nu} h\nu \cdot \Delta n_m(\nu) \cdot d\nu \right) \cdot \left(\sum_{m'}^{M'} \int_{\nu_0}^{\nu_0 + \Delta\nu} h\nu \cdot \Delta n_{m'}(\nu) \cdot d\nu \right) \rangle$$

where m' is a dummy summation index. In evaluating this expression we utilize the fact that perfectly *random* fluctuations in different orthogonal modes will be uncorrelated, so that only cross-products having the same mode index will survive the ensemble averaging

$$\langle (\Delta P_{inc})^2 \rangle = \langle \sum_m^M \left(\int_{\nu_0}^{\nu_0 + \Delta\nu} h\nu \cdot \Delta n_m(\nu) \cdot d\nu \right)^2 \rangle$$

In general, this expression is quite difficult to evaluate because it requires the knowledge of the radiation statistics in all possible modes at all possible frequencies. It is greatly simplified by the “band-limited” assumption that the spectral width $\Delta\nu$ is narrow enough that both $h\nu$ and $\Delta\nu$ can be considered constant over the range of the integration. One then gets

$$\langle (\Delta P_{inc})^2 \rangle \approx \langle \sum_m^M (h\nu_0 \Delta n_m(\nu_0) \cdot \Delta\nu)^2 \rangle = \sum_m^M (h\nu_0 \Delta\nu)^2 \langle [\Delta n_m(\nu_0)]^2 \rangle \quad (10)$$

where the order of mode summation and ensemble averaging have been interchanged since Δn is the only random variable.

To get the mean-square fluctuation of the absorbed power, we weight each term in Eqn 10 by the power coupling coefficient

$$\langle (\Delta P_{abs})^2 \rangle \approx \langle \sum_m^M \eta_m \cdot (h\nu_0 \Delta n_m(\nu_0) \cdot \Delta\nu)^2 \rangle = \sum_m^M \eta_m \cdot (h\nu_0 \Delta\nu)^2 \langle [\Delta n_m(\nu_0)]^2 \rangle$$

Note that even after band limiting, the sum must still be carried out over η_m and n_m since the coupling efficiency of different spatial modes in the THz region is highly variant, and the radiation statistics may be different than Boltzmann. In the special case of only one spatial mode, we find

$$\langle (\Delta P_{abs})^2 \rangle = \eta \cdot (h\nu_0 \Delta\nu)^2 \langle \Delta n(\nu_0) \rangle^2 \quad (11)$$

In the literature on remote sensing and RF signal processing this narrow $\Delta\nu$ approximation is often called “band-limited”. In the author’s experience, this is too often treated as an assumption rather than an approximation. Fortunately, it almost always leads to best-case predictions, often called fundamental noise limits. Experimental reality can only be worse, so system engineers need not be concerned about violating the “laws of physics.” These fundamental limits will be examined in the subsequent sections in the two special cases of purely coherent and purely incoherent radiation.

D.2. Coherent Radiation: Poissonian photon statistics

We start with the case of coherent radiation – so important to the wide array of active sensors (e.g., radars) used in the RF region, and passive coherent sensors that use a local oscillator. When quantized, this form of radiation becomes the coherent state, whose quantum statistics are given by the Poisson distribution.

$$p(n) = \frac{[\langle n \rangle (t)]^n}{n!} e^{-\langle n \rangle (t)}$$

where $\langle n \rangle (t) \propto \sin^2(\omega_c t)$. This state corresponds to only one electromagnetic mode at frequency ν_c which generally has a very narrow bandwidth $\delta\nu$ that is assumed to lie within the sensor passband, $\Delta\nu$. To apply Eqn 11 above, we need to calculate the variance of the Poisson distribution – a standard exercise in probability theory. It is found that

$$\langle (\Delta n)^2 \rangle = \sum_0^{\infty} (\Delta n)^2 p(n) = \langle n \rangle$$

In other words, the variance is equal to the mean. Substitution then yields the power variance

$$\langle (\Delta P_{abs})^2 \rangle \approx \sum_m^M \eta_m \cdot (h\nu_c \Delta\nu)^2 \langle n_m(\nu_c) \rangle$$

To simplify this further, we note that the use of Eqn 11 assumes that the mean population per longitudinal mode is flat with n , which is clearly not the case with a coherent source. But we can replace the coherent source with an effective source, flat across $\Delta\nu$, through the relation $\langle n_m \rangle_{\text{eff}} = P_{\text{inc},m} / (h\nu_c \Delta\nu)$, where $P_{\text{inc},m}$ is the time-averaged incident power in the m th spatial mode. When integrated over the sensor spectral passband, this always gives the correct photon flux, and leads to the power variance

$$\langle (\Delta P_{abs})^2 \rangle \approx \sum_m^M \eta_m \cdot h\nu_c \cdot \Delta\nu \cdot P_{\text{inc},m} \quad (12)$$

In the special but practical case that the coherent power is contained in only one spatial mode, we can immediately derive the standard deviation and RF power spectral density

$$S_p(\nu) \equiv \frac{\sqrt{\langle (\Delta P_{abs})^2 \rangle}}{\Delta\nu} = \frac{\sqrt{\eta \cdot h\nu_c \cdot P_{\text{inc}}}}{\sqrt{\Delta\nu}} \quad (13)$$

Note that Eqn 12 represents another interesting influence of quantum mechanics on sensor theory. Namely, fluctuations occur in the incident power of a coherent oscillator that, classically, one would expect to be noise-free. These fluctuations occur because of the quantized nature of the photons in the oscillation and the fact that such quantization always brings an uncertainty when “measurements” are made. In this case, measurement is simply the process of absorption.

D.3 Incoherent or Thermal radiation: Boltzman Photon Statistics

For any remote sensor calculations, particularly in the THz region, one needs to consider the fluctuations of thermal radiation. The terrestrial thermal radiation level is quite high, and even in outer space there is a (cosmic) background radiation corresponding to a 2.7-K thermal source. It is a basic exercise of statistical mechanics to show using the harmonic oscillator energy function and Boltzman statistics that

$$\langle (\Delta n)^2 \rangle \approx \langle (n - \langle n \rangle)^2 \rangle = \langle n^2 \rangle - \langle n \rangle^2 = \langle n \rangle (\langle n \rangle + 1) = f_p (1 + f_p) \quad (14)$$

where f_p is the Planck function. This expression has two interesting limits. In the low-frequency, or Rayleigh-Jeans limit when $h\nu \ll kT$, $f_p \gg 1$ and $\langle (\Delta n)^2 \rangle$ goes to f_p^2 , which is approximately $(kT/h\nu)^2$. In the high-frequency limit when $h\nu > kT$, $f_p \ll 1$ and $\langle (\Delta n)^2 \rangle$ goes to $\langle n \rangle$ or f_p , which goes to $e^{-h\nu/kT}$. The first limit is generally valid in the THz region at typical terrestrial temperatures. The second limit is generally anywhere in the near-infrared or shorter wavelengths under the same terrestrial conditions. For example, for $T = 300$ K, $\langle (\Delta n)^2 \rangle$ approaches $\langle n \rangle$ asymptotically just above 10 THz.

From the viewpoint of probability theory, note that the tendency for $\langle (\Delta n)^2 \rangle$ to go to $\langle n \rangle$ for low photonic state occupancy is exactly the behavior predicted by Poisson statistics, and the tendency to go to $(\langle n \rangle)^2$ at high occupancy is the behavior expected for binomial statistics, which become Gaussian in the limit of high state density. It is then appropriate to think of thermal radiation of all types as inherently Poissonian. This is fully consistent with the central-limit theorem of probability theory.¹⁰

Substitution of Eqn 14 into the band-limited expression for absorbed power fluctuation of M spatial modes then yields

$$\langle (\Delta P_{abs})^2 \rangle \approx \sum_m^M \eta_m(\nu_0) \cdot (h\nu_0 \Delta\nu)^2 \langle n_m \rangle (\langle n_m \rangle + 1) = \left[\sum_m^M \eta_m(\nu_0) \right] \cdot (h\nu_0 \Delta\nu)^2 f_p (f_p + 1)$$

where the sum over modes is carried out only over the coupling efficiency. This expression represents a remarkable effect of thermal photons called “bunching.” As the occupancy of photon states grows past unity, there is a tendency for the fluctuations to grow in the amplitude much more quickly. This is a ramification of the fact that photons are really massless bosons and, as such, are correlated quantum mechanically such that they tend to condense in a single state at a rate that grows with the population of that state. Remarkably, such “bunching” occurs routinely in THz sensors and RF sensors, in general, as shown in the next section. It has a classical explanation in terms of mixing between the electric fields in the same spatial but different longitudinal modes.¹¹

D.4 Low-Frequency Limit: The Johnson-Nyquist Theorem

A very important result for power fluctuations in the low-frequency limit is the Johnson-Nyquist theorem. Since in this limit $h\nu/kT$ is very small, $e^{h\nu/kT} \approx 1 + h\nu/kT$ (by Taylor expansion) so that $f_p \approx kT/h\nu \gg 1$, and thus $\langle (\Delta n)^2 \rangle \approx (kT/h\nu)^2$, which is a big number at THz frequencies for T near room temperature. As done above, let's suppose that a sensor is receiving thermal noise power in this limit contained in M spatial modes

over an arbitrary spectral bandwidth $\Delta\nu$. In this case the mean-square absorbed power from the statistical fluctuations Δn is given by

$$\langle (\Delta P_{abs})^2 \rangle = \left[\sum_m^M \eta_m \right] (h\nu_0 \Delta\nu)^2 f_p (f_p + 1) \approx (h\nu_0 \Delta\nu)^2 (kT / h\nu_0)^2 = \left[\sum_m^M \eta_m \right] (kT \Delta\nu)^2$$

which can be stated in a more familiar form as

$$P_{RMS} \equiv \sqrt{\langle (\Delta P_{abs})^2 \rangle} = \sqrt{\sum_m^M \eta_m} \cdot k_B T \cdot \Delta\nu = \langle P_{abs} \rangle / \sqrt{\sum_m^M \eta_m}$$

So by comparing this to the average power of Eqn 7, we observe the remarkable result that for one spatial mode ($M=1$) and perfect coupling, the rms absorbed power fluctuation is exactly equal to the average power – a sign of Gaussian statistics.

An important consequence of the Johnson-Nyquist theorem follows from the relationship established above between the fluctuations and the power spectral density, which allows us to write

$$S_p(\nu) \approx \frac{P_{RMS}}{\Delta\nu} = \sqrt{\sum_m^M \eta_m} \cdot k_B T$$

a result of such fundamental importance to THz sensor theory, and RF systems in general, that it is found in nearly every calculation of system-level performance.

D.5. High-Frequency Limit: Photon Shot Noise

Although RF systems are traditionally analyzed assuming validity of the Johnson-Nyquist theorem, the THz region is exceptional because for some systems $h\nu$ can approach kT so that $\langle n \rangle$ drops to order unity. As done above, let's suppose that a sensor is receiving thermal noise power in M orthogonal spatial modes over a “band-limited” $\Delta\nu$ that is narrow enough that $\langle n \rangle$ can be treated as a constant. Let's also assume that the electromagnetic waves associated with these spatial modes are all equally-well matched to the sensor (i.e., all incident waves experience the same coupling efficiency). With these assumptions, the power from the fluctuations Δn is given by

$$\langle (\Delta P_{abs})^2 \rangle = \sum_m^M \eta_m \cdot (h\nu_0 \Delta\nu)^2 f_p (f_p + 1)$$

which can be written immediately as an effective power spectral density

$$S_p(\nu) \approx \frac{P_{RMS}}{\Delta\nu} = h\nu_0 \sqrt{\sum_m^M \eta_m} \sqrt{f_p (f_p + 1)}$$

This can be clarified somewhat by noting that the mean power incident from all spatial modes is simply $\sum_m^M f_p \cdot h\nu \cdot \Delta\nu$, so we can re-write this as

$$S_p(\nu) \approx \frac{P_{RMS}}{\Delta\nu} = \frac{\sqrt{h\nu_0 \cdot \left[\sum_m^M \eta_m \right] \cdot P_{inc} \cdot (f_p + 1)}}{\sqrt{\Delta\nu}}$$

In the high-frequency limit, $h\nu \gg k_B T$, we have $f_p \ll 1$ so that the fluctuations in absorbed power become

$$\langle (\Delta P_{abs})^2 \rangle = \sum_m^M \eta_m \cdot h\nu_0 \cdot \Delta\nu \cdot P_{inc} ,$$

and the spectral density is given by

$$S_p(\nu) \approx \frac{P_{RMS}}{\Delta\nu} = \frac{\sqrt{h\nu_0 \cdot P_{inc} \cdot \sum_m^M \eta_m}}{\sqrt{\Delta\nu}}$$

Note that in the unimodal case, this is identical in form to the spectral density for a coherent signal, Eqn 13. It is satisfying that such a fundamental result can be derived in two different ways. As will be explained later in Sec. V.B, this result is called “photon shot noise”.

E. THz propagation

E.1 Atmospheric absorption

More than any other region of the electromagnetic spectrum, the atmospheric transmission in the THz region is dominated by one factor – water vapor. Being a polar molecule with a nonlinear molecular orientation, water displays a strong absorption line for nearly all of its rotational modes. Most of these modes and their optical properties have been conveniently catalogued over the past few decades into a database called HITRAN96. Started by the U.S. Air Force Geophysics Laboratory, the database contains the kinetic coefficients for over 1 million atomic and molecular lines of importance to atmospheric radiative transport between the microwave and visible regions. The set of water vapor lines occurring between 100 and 1000 GHz is shown as a “stick” diagram in Fig. 4. The ordinate is a line-strength coefficient for the Voight model of atomic absorption – a model that combines the essential characteristics of the Lorentzian model of collisional broadening with the Gaussian model of Doppler broadening.

Note that the spectral region in Fig. 4 contains 228 spectral lines. Perhaps the best known of these is the 183 GHz line, which has been used by NASA and other government agencies as the basis for space-based atmospheric-sounding radiometers. Remarkably, many of the water lines that occur at higher frequencies approaching 1 THz and beyond are orders-of-magnitude stronger than the 183-GHz line.

Given the water lines and a model of atmospheric pressure and temperature, one can compute the transmission over chosen paths using a radiative transport code. A popular tool over the past 20 years has been FASCODE, also developed by the U.S. Air Force. In the late 1990s this was converted to a Windows environment.¹² An example transmission spectrum up to 2.0 THz is plotted in Fig. 5. This was computed over a 1 km path length at 1 atm of pressure, 296 K temperature, 60% relative humidity, and a slant angle of 5°. The strongest lines of Fig. 4 are now pressure broadened, and even stronger lines occur above 1.0 THz. In places where neighboring lines are in close proximity, the absorption remains very strong between the features, creating what some have called the “THz pea soup.”

An important comment about the HITRAN96 database is that all of its absorption lines apply to monomeric water. At 1 atm pressure and high humidity, it is known that water behaves in a more complicated fashion whereby individual molecules interact to form dimers, trimers, etc. This is, of course, the physical basis for precipitation so it is not surprising that it is occurring at high humidity. The effect of dimerization is an increase in the number of lines in Fig. 4 and a further thickening of the absorption spectrum in Fig. 5.

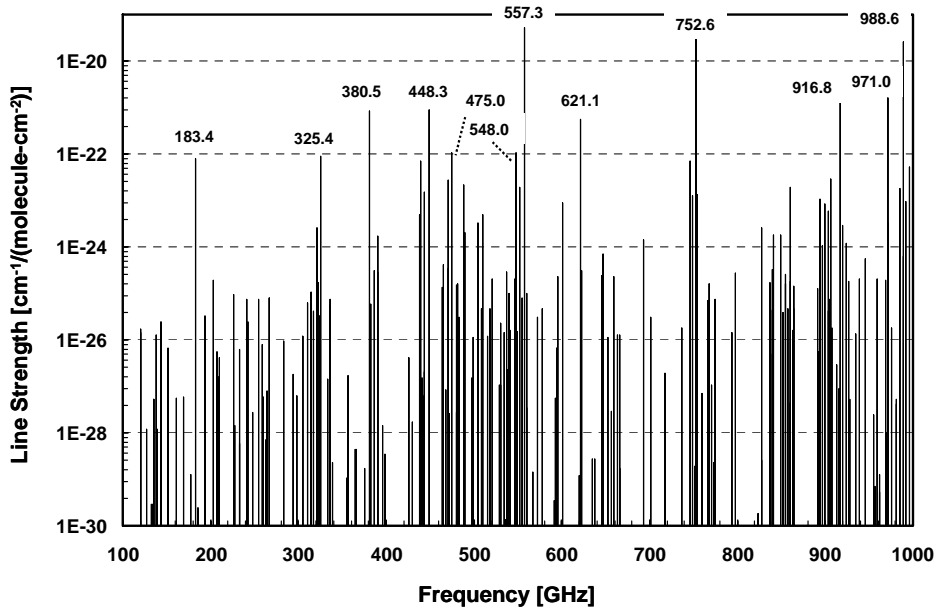


Fig. 4. Stick diagram of water vapor absorption lines between 0.1 and 1.0 THz (from HITRAN96 database). Labels denote the center frequency of 12 strongest lines.

E.3. Effect on THz signals: Radiative transfer

Given the existence of a molecular absorption and a pressure at or near 1 Atm, the density of molecules is generally large enough that the interparticle separate is less than a wavelength in the THz region. Hence, the attenuation from absorption can be approximated through the exponential fall in intensity predicted by the Lambert-Beer law

$$\tau \equiv \frac{I_t}{I_i} = \exp(-\rho \cdot \sigma_A \cdot z)$$

where I_i is the incident intensity, I_t is the transmitted intensity, ρ is the particle density, σ_A is the absorption cross section, and z is the path length.

Of course, absorption is concomitant with energy transfer, so we expect an absorbing medium to also contribute radiation according to the principle of detailed balance. The overall radiative transfer has a well-known solution when the source body of thermodynamic temperature T_S subtends a solid angle Ω with respect to the sensor and the intervening medium is much broader than Ω . Because brightness is the quantity conserved over a fixed solid angle, it is the convenient radiative function and is given at the sensor receiver by¹³

$$B_R = \tau(\nu) B_S + (1-\tau) B_M$$

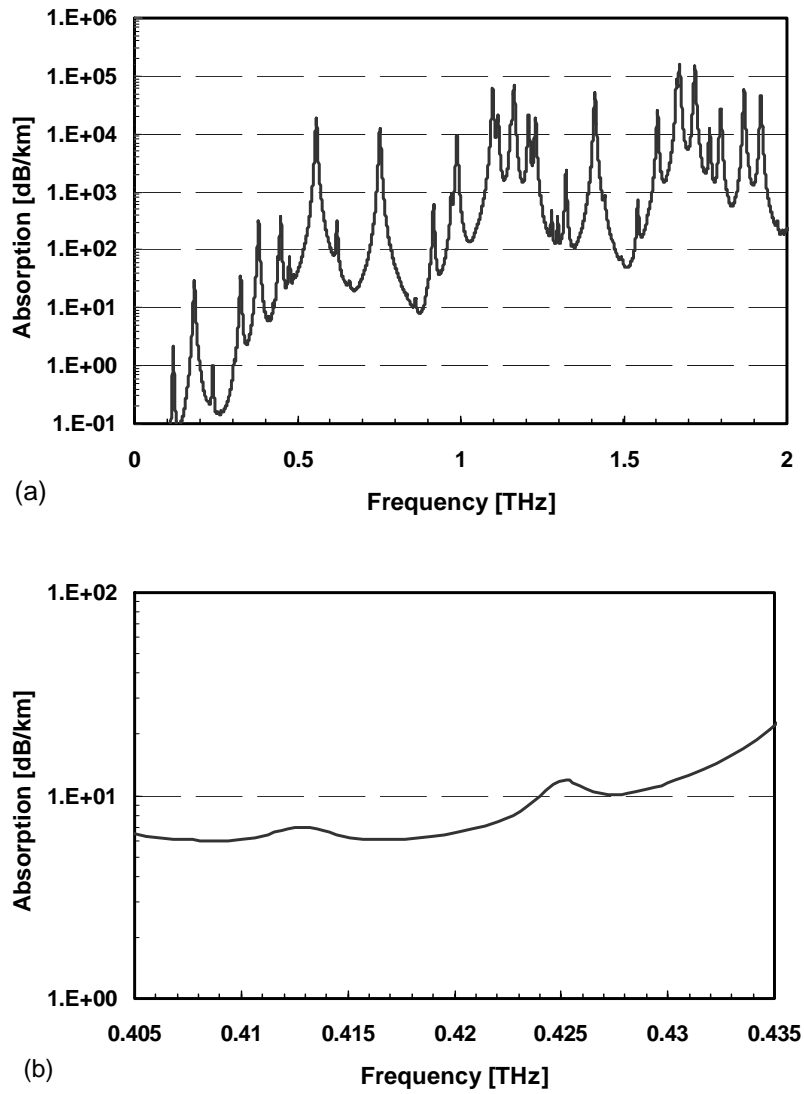


Fig. 5. (a) Transmission of atmosphere through 1 KM of path as computed by PcLnWin. (b) Magnified view of the region around 420 GHz, which will be used later in the chapter for an active sensor simulation.

where B_S and B_M are the brightnesses of the source and medium, respectively. Again, for terrestrial remote sensing the Rayleigh-Jeans limit is quite accurate in the mm-wave and THz regions, so that from the definition of brightness Eqn 9, we can write

$$T_R = \tau(\nu) T_S + (1-\tau) T_M$$

Note that if the background and medium temperatures are in thermodynamic equilibrium, $T_S = T_M$ and we find $T_R = T_S$.

E.3. Atmospheric scattering

In the millimeter-wave and THz regions, the absorption mechanisms in the atmosphere are known to be strong, particularly from water vapor. In principle, one also expects attenuation to occur also through scattering. Because of the much longer wavelength compared to the visible or infrared, one would expect the THz scattering to be much weaker from the common, micron-scale airborne particle types, such as fog, ice grains, dust, pollens, etc. However, there are conditions, such as precipitation, in which scattering should play an important role. Unfortunately, little or no experimental data exists on scattering under these conditions owing to the lack of calibrated THz instrumentation and the general difficulty in distinguishing absorption from scattering effects.

A general statement can be made that if the scattering is strong enough to compete with or dominate the absorption. In this case, the simple “rectilinear” expression given above should probably be abandoned and a more sophisticated radiative transfer carried out. To see why, suppose that a ground-based sensor and airborne source are separated by a rain cloud thick enough that the attenuation by scattering is comparable to that by absorption. Under this condition, the raindrops will almost certainly reflect significant radiation into the sensor that originates from the ground below. The reader is referred to two excellent texts that address the scattering problem in great detail.^{14,15}

F. Measurement of THz power

An essential task for the calibration of detectors and qualification of sources is the accurate measurement of power and frequency. Measurement of power in the THz region is a surprisingly difficult task, particularly at microwatt levels and below. As in the lower millimeter-wave region, thermal detectors are often the most accurate detector, although not the most sensitive. This is because of their relative simplicity and ability to efficiently absorb the radiation. The practical requirement on any thermal detector is the sensitivity be good enough to detect the low ($\ll 1 \mu\text{W}$) THz output that occurs when a new source is first being developed.

Perhaps the best candidate thermal detector for room temperature operation is the Golay cell – a venerable device that combines absorption of radiation by a thin metallic film and transfer of the heat into a gas.¹⁶ The expansion of the gas moves a membrane on which is mounted an optical grating or similar component. The motion of the grating diffracts light between a source and a photodetector, causing a large change in output voltage of the detector. In the author’s experience, the Golay cell is very uniform in spectral response, being limited only on the long-wavelength end by the entrance aperture. For the typical 3-mm-aperture Golay cell, this long-wavelength cutoff is roughly 100 GHz. For frequencies around 1 THz, the NEP is roughly $2 \times 10^{-10} \text{ W}\cdot\text{Hz}^{-1/2}$. For the typical detector bandwidth of 10 Hz, the minimum detectable power is thus $\sim 1 \text{ nW}$.

The best candidate for cryogenic operation is the composite bolometer operating at 4.2 K (liquid helium).¹⁷ This device consists of a thin absorbing film in an integrating cavity on which is mounted a small Si (or perhaps some other semiconductor) thermistor having a high coefficient of thermal resistance, $(1/R) dR/dT$. The key advantage of

cryogenic over room-temperature operation is reduction in the thermal noise while maintaining high responsivity. Around 1 THz the 4.2-K Si composite bolometer can provide an optical NEP of roughly 1×10^{-12} W-Hz^{-1/2}. So in the typical bandwidth of 10 Hz, the minimum detectable power is ~ 10 pW. The disadvantage of the composite bolometer compared to the Golay cell is uniformity of response. Because of standing waves that can occur in an integrating cavity, the spectral variation in responsivity can be a factor of ten or more – a fact not appreciated by many people in the field.

III. Coupling of THz Sensors to Free Space

A critical aspect of any remote sensor is “connecting” THz radiation between the various signal processing components and devices, and “coupling” the THz radiation to and/or from the external medium in which the target is embedded. The connecting task, generic to all electromagnetic sensors, is carried out with THz transmission lines. The coupling is carried out with various types of antennas.

A. Routing between devices: THz Transmission Lines

THz transmission lines are usually scaled-down versions of microwave and millimeter-wave transmission lines and, as such, are relatively simple to analyze and design. The two most common types are metallic waveguides and coplanar-strip transmission lines. The former type is popular because if fabricated with high conductivity metal (e.g., oxygen-free copper) and smooth inside walls, it can display high bandwidth and low insertion loss. The latter type is popular because if fabricated on a dielectric or a high-resistivity semiconductor (e.g., quartz or semi-insulating GaAs), it can be integrated with THz devices such as superconducting tunnel junctions or Schottky diodes. Various types of scaled-down optical structures have also been investigated as THz transmission lines, such as dielectric waveguides. But because of high insertion loss, inability to integrate with devices, or both, these structures have not gained popularity and will not be addressed in this article.

A.1. Metallic waveguides

Perhaps the most common THz transmission line of all is the rectangular or circular metallic waveguide. The former is generally made up to about 300 GHz according to the “WR-N (Waveguide Rectangular) standard whereby the width-to-height ratio is approximately 2.0 and N is the width in units of 1/100th of an inch. For example, WR-3 has a width $W = 0.03$ inch and a height $H = 0.015$ inch. According to waveguide propagation theory, the cut-off frequency of the fundamental (TE₁₀ mode) is then $f_c = c/(2W) = 197$ GHz. WR-3 is about the smallest rectangular waveguide that is available commercially. Smaller sizes are made and follow the WR standard but are generally only made in a custom fashion.

Circular waveguide is also popular, primarily because it is easier to fabricate by simple machine-shop (e.g., milling) techniques. The fundamental mode is the TE₁₁ whose cut-off properties are given by $0 = J_0(kr)$ where $k = \omega/c$ is the wavenumber, r is the radius, and J_0 is the Bessel function of the first kind and order zero. The first “zero” of

this Bessel function is $J_0(1.841) = 0$, so that the cut-off wavelength in terms of diameter is simply $(\pi/\lambda_c)d = 1.841$. This yields $\lambda_c = \pi d/1.841$, or $f_c \equiv c/\lambda_c = 1.841c/\pi d$. For example, a circular waveguide having a diameter of 1 mm (a convenient size to make by conventional machine-shop techniques) displays a cut-off frequency of 175.8 GHz.

At frequencies well above the cut-off of the fundamental mode in either rectangular or circular waveguide, higher-order modes begin to propagate. For any uniform waveguide, these modes obey the fundamental relation for the cutoff frequency in terms of the width (a) and height (b) of the waveguide:

$$v_c^{m,n} = c\sqrt{(m/2a)^2 + (n/2b)^2} \quad (15)$$

An interesting use of this expression is to plot the number of modes $N(v)$ that can propagate at any frequency as a function of frequency above cut-off. Such a curve is easily generated by a spreadsheet with sorting capability, as exemplified later in Sec. III.D.

B. Coupling to free space: THz Antennas

A critical aspect of any remote sensor is the coupling to the external medium in which the target is embedded. Since the external medium is usually free space, the component that carries out this coupling is traditionally called the “antenna”. As in other regions of the electromagnetic spectrum, THz antennas generally fall into one of two categories: (1) wire antennas, such as dipoles, and (2) aperture antennas, such as feedhorns. The distinction between these two is a key topic of many textbooks and will not be addressed here.¹⁸ What distinguishes the THz region from microwave bands and below is that wire-like antennas are generally fabricated on substrates and feedhorns are often operated overmoded, as described further below.

B.1. THz wire-like antennas

Perhaps the first wire-like antenna to be used successfully at THz frequencies is the traveling-wave, unbalanced, linear (“cat-whisker”) antenna. It consists of a long wire (relative to a free-space wavelength) connected at one end to a ground plane (or a device over a ground plane) and at the other end to a feed port. To force the antenna pattern into one direction of space and thereby improve the directivity, the wire is mounted in a corner-cube reflector.¹⁹ Historically, because of its broadband nature and small reactance, the cat whisker antenna played a key role in the development of THz technology as the demonstration vehicle for Schottky diodes, resonant-tunneling diodes, and other THz semiconductor devices.

The drawback of cat-whisker corner-cube antennas was and is the difficulty in their packaging and incompatibility with integrated circuit techniques. Therefore, during the 1970s and early 1980s, significant research and development was put into printed-circuit antennas, also called planar antennas.²⁰ A key driver for this development was the success of one type of PC antenna – the microstrip patch – at microwave frequencies. By this time the utility of the patch was well known, particularly its amenability to two-dimensional arrays. But researchers found that the patch was prone to inefficiency at THz frequencies because of its propensity to launch substrate modes no matter how thin the substrate.²¹ This led to the exploration of other antenna types,

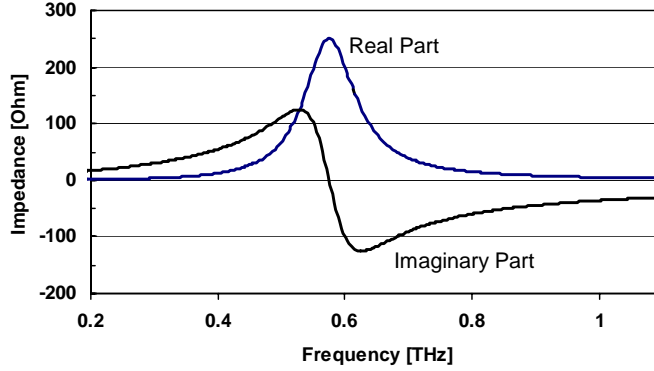


Fig. 6. Real and imaginary parts of the driving point impedance of a planar full-wave dipole antenna

including broad-side dipoles, slots, and various self-complementary antennas. End-fire planar antennas were also investigated, particularly linearly and exponentially-tapered (i.e., Vivaldi) antennas.²²

Amongst the first planar antennas to be used experimentally were the class of self-complementary antennas, such as the bow-tie and the logarithmic spiral. These are antennas that after some rotation operation less than 2π become their own Babinet complement. For both types the required rotation angle is just $\pi/2$. To achieve this in the log spiral, the edge of each arm emanates out on a locus given by $r_0 e^{c\theta}$ and each edge is rotated from its nearest neighbor by 90° . All self-complementary antennas display special electromagnetic characteristics, namely a driving-point impedance that is real and independent of frequency over a wide operational bandwidth (determined by the minimum and maximum lateral extent of the self-complementary structure). For a three-turn spiral, this bandwidth can easily be one decade. The value of the driving-point resistance is then given by

$$R_A = 60\pi/(\epsilon_{\text{eff}})^{1/2}$$

where ϵ_{eff} is the effective dielectric constant given by $(1 + \epsilon_r)/2$. For GaAs, $\epsilon_r \approx 13.0$ so that $R_A = 72 \Omega$. Another remarkable feature of self-complementary spirals is their insignificant load susceptance B_L over the bandwidth.

For some devices a higher antenna driving-point resistance is desirable. A good choice is a resonant planar antenna, such as a planar dipole or slot. The driving-point impedance of a planar dipole on a semi-infinite GaAs half-space is shown in Fig. 6.²³ Note that its resonant resistance is 250Ω - over three times greater than that of the self-complementary spiral.

Coupling wire-like antennas to free space

Although planar antennas offer a large variety of pattern and impedance characteristics, they do not necessarily solve the primary problem behind patch antennas – substrate modes. For example, a planar dipole on a GaAs substrate exposed to air on the back side will also support at least one substrate mode and thus will suffer from

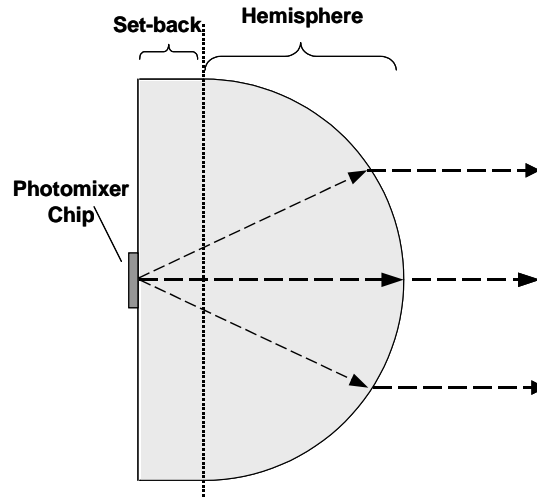


Fig. 7. Cross sectional view of hyperhemispherical optical coupling commonly used for mm-wave and THz direct detectors, mixers, and even some types of sources.

inefficiency. The problem is not so much the antenna type as it is the fundamental cause of substrate modes, which is the total internal reflection at the interfaces of the substrate. Realizing this problem, various techniques were explored to mitigate substrate mode generation and improve efficiency. Probably the most successful is to couple the radiation through the substrate using a back-side hemispheric lens whose dielectric constant is close to that of the substrate. For device integration purposes, planar antennas are often fabricated on quartz or semi-insulating GaAs. Therefore, a good choice for the lens material is quartz or a low-cost, high-resistivity semiconductor, such as float-zone silicon.

The common application of the lens-coupled antenna has been to locate a THz device, such as a mixer or direct detector diode, at driving gap of the planar antenna, and locate the driving gap at the center of curvature of the lens. If the lens is thick enough so that the spherical surface is in the far-field of the antenna, then all the radiation from the antenna reaches the spherical surface at normal incidence and passes in to free space with a reflection coefficient of $(n_L-1)^2/(n_L+1)^2$. This is a great improvement over the slab substrate, but yields a highly diverging beam from planar antennas having large beam solid angle. In such cases, a common practice is to locate the antenna behind the center of curvature at a point where the radiation will be refracted in the forward direction. If the hemispherical surface and set-back are fabricated in the same dielectric material, the resulting optic element is called a *hyperhemispherical* lens.

If n_L is perfectly matched to n_S and the antenna is located behind the center of curvature by a distance r/n , then the lens focusing is *aplanatic*. This is a property from geometric optics which states that all rays from a given point in the object plane (in this case, the planar antenna) are refracted to a parallel bundle after the lens, as shown in Fig. 7. Clearly, most planar antennas can not be accurately analyzed by geometric optics. A treatment of the focusing problem by modal (Gaussian-beam) analysis shows that the optimum set-back is just short of r/n , as might be expected.²⁴ This point is addressed further in Sec. III.E.

B.2. THz aperture-like antennas: Dishes, Lenses, and Feedhorns

Parabolic dishes: the diffraction limit

Perhaps the most universal of the aperture antenna types is the parabolic dish – a two-dimensional surface consisting of a parabola rotated 360° about one (optical) axis. The surface of the paraboloid is generally made from a high-conductivity material, such as a good metal. Parabolas are just as effective at THz frequencies as in the RF bands. And like their application in RF systems, parabolic dishes tend to get used as the primary reflector in the front end of THz sensors. This is because they are straightforward to design and manufacture, and display relatively predictable performance in terms of radiation coupled to free space. Their performance is much more easily predicted than practically any wire-like antenna.

To quantify the performance, we imagine a parabolic dish of diameter D lying at the origin of a spherical coordinate system whose origin is located at the center of the dish and whose polar axis coincides with the axis of symmetry. Then, as with any wire-like antenna, the electric field transmitted by the dish at a far distance from the origin will tend to display a modified spherical-wave form

$$E(r, \theta, \phi) \propto \sqrt{F(\theta, \phi)} \frac{e^{-jkr}}{r} \quad (16)$$

where k is the free-space propagation constant ($= \omega/c = 2\pi/\lambda$) and F is the (normalized) intensity pattern function, $F \equiv |\mathbf{S}(r, \theta, \phi)|/S_{\max}$ where \mathbf{S} is the Poynting vector and S_{\max} is its maximum magnitude, wherever in space that occurs. The range of r for which Eqn 16 is valid is called roughly the “far-field limit.” Parabolic dishes, like all wire-like antennas, display a limited region in space where $F(\theta, \phi)$ is large and other regions where it is negligible, in contrast to isotropic (point) sources. Therefore, a useful metric is the directivity

$$D = 4\pi \left(\iint_{4\pi} F(\theta, \phi) d\Omega \right)^{-1} \equiv 4\pi / \Omega_B$$

where Ω_B is the beam solid angle. Conceptually D defines how much greater the intensity is at the peak of F compared to the isotropic radiator emitting the same total power, for which $\Omega_B = 4\pi$ and $D = 1$.

Like other aperture-type antennas used in the mm-wave and THz regions, parabolic dishes generally have a pattern function that displays a predominant, symmetric or quasi-symmetric peak (i.e., “major lobe”) in a single direction of space θ_p, ϕ_p . In this case it is useful to approximate $F(\theta, \phi)$ by an equivalent spherical cone or sector having a symmetry axis along θ_p, ϕ_p , and polar angular width (or widths) equal to the full-widths at the half-maximum points $\beta(\phi)$ of the real major lobe. And throughout the cone or sector, $F(\theta, \phi) = 1.0$. If the pattern has perfect conical symmetry (generally true for parabolic dishes and lenses, and often the design goal for feedhorns), then one finds

$$\Omega_B \equiv \iint_{4\pi} F(\theta, \phi) d\Omega \approx \int_0^{2\pi} d\phi \int_0^{\beta/2} \sin \theta \cdot d\theta = 2\pi[1 - \cos(\beta/2)]$$

and

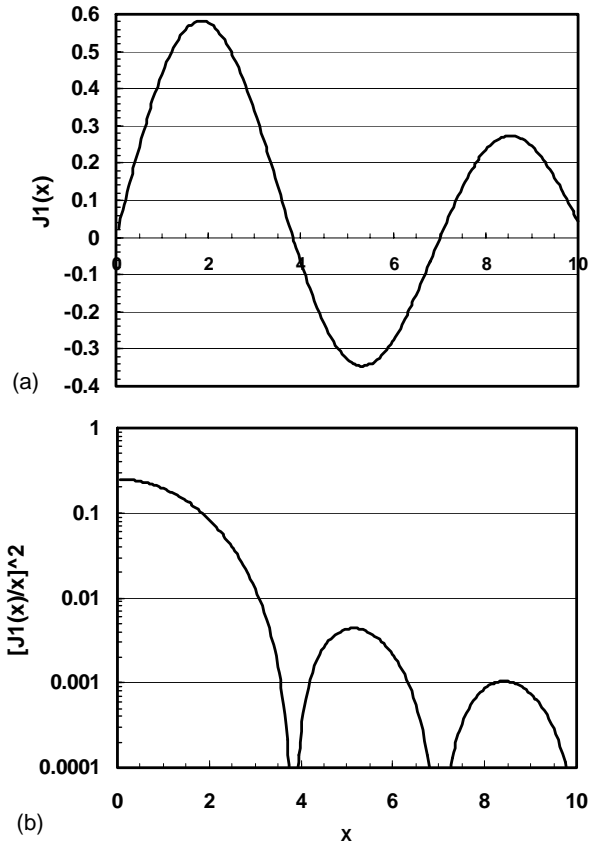


Fig. 8. (a) Bessel's function of the first kind of order 1. (b) Plot of the function $[J_1(x)/x]^2$, which according to scalar diffraction theory, is the angularly dependent part of the radiation pattern from a uniformly illuminated circular aperture.

$$D \approx \frac{2}{1 - \cos(\beta/2)} \quad (17)$$

In the limit of a narrow "pencil" beam where β is small ($\ll 1$ rad), one can Taylor expand the denominator, yielding

$$D \approx \frac{16}{\beta^2} \quad (18)$$

Note that in most books make the simpler approximation $\Omega_B \approx \beta^2$, so that $D \approx 4\pi/\beta^2$ - a less precise expression but one easier to remember.

The characterization of the parabolic dish then reduces to knowing the -3-dB full-width of the main lobe. This is where dishes and some other aperture antennas, such as lenses, are simpler to analyze than wire-like antennas because they are often physically large in lateral dimension compared to a free-space wavelength and have simple boundary conditions. In this case, one can approximate the radiation pattern with scalar diffraction theory - a formalism that provides an approximate solution to the vector electromagnetic wave (Helmholtz) equation for radiation passing through the aperture.²⁵ The scalar formalism results in the famous Kirchoff-Fresnel integral which, in essence, approximates the radiation pattern as the superposition of point sources filling the aperture, each point source radiating a spherical wave. A key issue in using this integral is the amplitude distribution of the point sources inside the aperture. In the special case of *uniform* illumination, scalar diffraction predicts a far-field pattern that goes as²⁵

$$|F(\theta, \varphi)|^2 \propto \left[\frac{J_1(ka\theta)}{ka\theta} \right]^2$$

where J_1 is the ordinary Bessel function of 1st order, a is the radius, and θ is the angle of the measurement point relative to the optical axis. Historically, this function has generally been avoided or grossly approximated in textbooks because of its lack of tabulation. With modern computation tools (e.g., Excel, Matlab), it is simple to compute and plot, and yields much useful information.

Like the more familiar sinc $[\sin(x)/x]$ function, $J_1(x)/x$ peaks at $x = 0$. But its peak value is 0.25, not 1.0 as for the sinc(x). The first null occurs at the first zero of the J_1 function, $x = 3.835$ or $\theta = 3.835 \lambda / (2\pi a) = 0.610 \lambda / a$. A secondary peak of magnitude 0.00437 occurs at approximately $x = 5.14$, corresponding to $\theta = 0.818 \lambda / a$. Note that this secondary peak (first sidelobe) has a value of 0.0175 or -17.6 dB relative to the main lobe. This is to be contrasted to the more familiar value of -13.2 dB for the relative size of the first sidelobe for a square aperture of uniform illumination.

The undulating radiation pattern of Fig. 8 and the associated minima and maxima, are collectively associated with *diffraction* - a phenomena that tends to be as important at millimeter-wave and THz frequencies as it is in the lower RF bands. This is because the size of components, such as parabolic dish antennas, tends to shrink with operating wavelength, making the important ratio λ/a approximately constant. For non-circular antennas or other components, this ratio gets replaced by λ/d where d is the maximum lateral extent of the component.

From an antenna-design standpoint, the most important feature of the diffracted radiation pattern is often the -3-dB beam width. From the uniform-illuminated case of Fig. 8, the -3-dB point is at $x = 1.616$, so that the beam full-width is given by $\beta = 2\theta = 3.232 \lambda / (2\pi a)$. This can be substituted into Eqn 17 or approximated by Eqn 18 if the dish is large enough that $\beta \ll 1$ rad. In this case $D \approx 16/\beta^2 = [16(2\pi a)^2] / (3.232 \cdot \lambda)^2 = 1.53 \cdot [4\pi A / \lambda^2]$ where $A = \pi a^2$ is the circular area. The last quantity arises frequently in the analysis of aperture and wire-like antennas and so has a special place in the electromagnetic field as the maximum or "diffraction-limited" directivity

$$D_{\max} = 4\pi A / \lambda^2 \quad (19)$$

The circular-aperture example shows this expression is not precise. But it has long been used as a metric in component and system engineering, and is often assumed to be more

precise than it really is. This is because it is so useful, applying also to wire-like antennas when the physical aperture is replaced by the “effective” aperture A_{eff} .

Another useful result that comes out of scalar diffraction theory is an estimate of the far-field-limit criterion. For the generic aperture antenna of uniform illumination, or a wire-like antenna, this is given roughly by the range expression

$$r > 2d^2/\lambda \quad (20)$$

where d is again the maximum lateral extent. Note, however, that this expression is even rougher than the diffraction-limited directivity, and r might have to be much greater than this, perhaps up to $10d^2/\lambda$, depending on the exact aperture shape and boundary conditions.²⁶

Given Eqns 19 and 20 one can quickly estimate the maximum directivity and far-field criterion as a function of frequency. Those accustomed to dealing with antennas in the microwave region are often impressed with how high the directivity can be in the THz region per unit aperture. For example, a round dish of 1-inch diameter operating at 1 THz ($\lambda = 0.3$ mm) has a “diffraction-limited” directivity of just over 7×10^4 , or 48.4 dB!

By the same token, as the frequency goes up one must go out to an increasingly further range to apply the far-field criterion. For this same dish the far field criterion becomes $r > 430$ cm = 14.1 ft, which seems surprisingly far considering the size of the aperture.

If a THz antenna is separated from its target by a distance less than indicated by Eqn 20, then the far-field form of the electric field can not be approximated by Eqn 16. An important consequence of this violation is that the magnitude of electric field at a given θ can not be assumed to fall as $1/r$ or, equivalently, the intensity can not be assumed to fall as $1/r^2$. This is a motivation for the derivation and application of the Gaussian-beam solutions to Maxwell’s equations.²⁷ Such modes are commonly applied to the free-space radiation transmitted or received by THz antennas, and are addressed in Sec. III.E.²⁸

An interesting aspect of parabolic dishes in the THz region is the limitations imposed by surface roughness or nonparabolicity. Up to approximately 100 GHz, parabolic dishes can be made very accurately by simple mechanical means such as molding. But in the THz region, a surface error of just a few thousandths of an inch can impact the performance of the dish. So surface machining and polishing becomes necessary to achieve diffraction-limited directivity – a requirement well known to people from the optical end of the spectrum. Surface machining has been scaled up to produce large area (10-m diam) parabolic dishes operable up to at least 600 GHz in astronomical observatories.²⁹ Even so, a THz parabolic dish is far easier to fabricate than one in visible or IR regions where submicron-scale smoothness is required.

Lenses

Almost since the beginning of the THz field, researchers have recognized that the wavelengths in this region become a small fraction of the diameter of common-sized lenses and other transmissive optics. For example the free space wavelength at 1 THz, $\lambda = 333$ micron, is 75 times less than the diameter of a 1-inch lens. In fact, lenses often are used as a paradigm for the field of “quasi-optics” – a broad term that means roughly the free-space control of THz (or lower frequency) radiation by conventional optical

elements. This does not imply, however, that optical techniques for analyzing the propagation are a good approximation with THz lenses or mirrors. To get acceptable accuracy, diffraction effects usually must be accounted for either by the uniform-illumination or Gaussian-beam approximations (Sec. III.E).

An advantage of transmissive optics in the THz region compared to the visible or infrared is that they can be made from plastics – perhaps the best material choice of all from the perspective of manufacturability and low-cost. Good examples of plastics having low-enough loss and good manufacturability for lenses and windows are Teflon, polyethylene, and TPX. All of these are readily machined, and simple machine-shop surfacing methods leave small enough tool marks and shallow-enough scratches to cause negligible scattering of electromagnetic radiation in the THz region.

Feedhorns

Probably the best performance of all antennas up to about 1 THz is displayed by well-designed feedhorns. Through the microwave and millimeter-wave bands, pyramidal feedhorns are very popular because of their compatibility with rectangular waveguide and good control over the beam pattern. But in the THz region, rectangular-waveguide standards are not prevalent and, because of size scale, it becomes much easier to fabricate conical feedhorns rather than pyramidal ones. Unfortunately, conical feedhorns are not as flexible to design because of their one fewer degrees-of-freedom. They are prone to inefficient and asymmetric patterns that can not be easily designed by analytic means because of their required representation by special (Bessel) functions.

With the advent of full-wave numerical simulation tools, it has become possible to design THz conical feedhorns with excellent antenna characteristics. One of the most popular numerical codes presently available– High Frequency Structure Simulator (HFSS) – is based on the method of finite elements. Given such a simulation capability, there are several design tricks that can produce excellent patterns from conical feedhorns. One such trick is to add a step in the throat region to excite just enough of the first higher-order (TM_{11}) mode to cancel the sidelobes in the fundamental (TE_{11}) mode.³⁰ A second trick is to fabricate grooves down the length of the conical horn to excite the mode exclusively. The resulting design, called a scaler feedhorn, produces outstanding beam patterns with practically no sidelobes – arguably the best possible pattern from any antenna in the THz region. Unfortunately, scaler feedhorns are difficult to design and even more difficult to fabricate.

As more and more modes begin to propagate in the circular waveguide to which the feedhorn is connected, the radiation pattern begins to resemble the ray transfer function of a condensing cone, sometimes called a light cone. In other words, the feedhorn begins to follow the laws of geometrical optics. In the 1970s infrared researchers began investigating other possible condensing cone tapers besides the simplest (linear) case. For example, Winston discovered that a dual-parabolic taper can produce a transfer function near the ideal (step) function.³¹ Not surprisingly, the Winston cone concept works very well at THz frequencies provided that the throat of the horn is heavily overmoded.

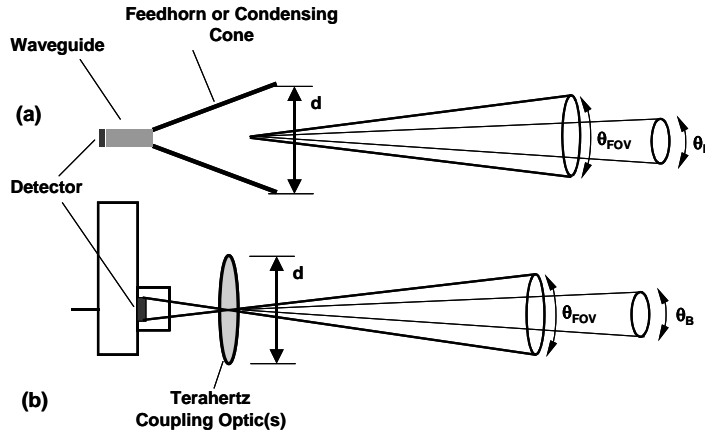


Fig. 9. (a) Feedhorn and (b) Lens coupling of mm-wave and THz detectors and devices

C. The Number of Modes: Antenna Theorem and Interpretation of Planck's Law

Whether the sensor is collecting coherent or thermal radiation, the number of spatial modes collected is an important issue from a system standpoint, is difficult to estimate, and is even more difficult to measure. In the limited space here, we only provide an approximation.

The approximation technique is based on a result from electromagnetic theory called the antenna theorem, which derives from two separate definitions of directivity given above,

$$D_{\max} = \frac{4\pi A}{\lambda^2} = \frac{4\pi}{\Omega_B} \quad (21)$$

Since D_{\max} is, in principle, the maximum possible value of the directivity as predicted by diffraction theory, then the corresponding Ω_B is the minimum possible beamwidth and, therefore, corresponds to the *fundamental spatial mode* of the antenna. So if we imagine rotating the Ω_B beam in spherical coordinates to just fill up surface of a sphere, it would take approximately $4\pi/\Omega_B$ rotations to do this. And since Ω_B is the fundamental spatial mode, D_{\max} represents the number of spatial modes required to fill the entire sphere. Of course, rarely, if ever, is a THz sensor designed to operate over an entire sphere. More often, they are designed to respond to a much smaller solid angle, called the field-of-view, Ω_{FOV} . And the number of spatial modes then becomes

$$M = \frac{\Omega_{FOV}}{\Omega_B} \quad (22)$$

Note that the correspondence represented by Eqn 21 is often stated as the following “antenna theorem”

$$A \cdot \Omega_B \geq \lambda^2$$

Conceptually, this means that if the antenna is diffraction-limited, the product will reach its minimum value of λ^2 . But it reminds us that practical antennas have electromagnetic or mechanical limitations which usually cause the fundamental beamwidth to grow beyond this minimum value.

Illustrations are given in Fig. 9 for the Ω_B and Ω_{FOV} for the two most common millimeter-wave and THz coupling structures: feedhorns (or coupling cones), and quasi-optical components (e.g., lenses). In both cases, one can always define the beam angle Ω_B . For example, for the simple round lens, it is predicted from scalar diffraction theory $\theta_B \approx 2.44 \lambda/D$, and $\Omega_B \approx (\theta_B)^2$ if $\theta_B \ll 1$. For the feedhorn or collection cone, there will always be a fundamental electromagnetic mode defined as that mode collected by the sensor at the maximum possible wavelength. Generally, this mode and its cutoff wavelength are defined by the dimensions of the throat of the feedhorn or cone.

An interesting application of the antenna theorem comes in a practical interpretation of Planck's radiation law for RF and THz systems. We can now re-write Eqn 8 ($\epsilon = 1$) as

$$\frac{dP}{d\nu} = \frac{A \cdot 2\pi h \nu^3}{c^2 (e^{h\nu/k_B T} - 1)} = \frac{2\pi A}{\lambda^2} \frac{h\nu}{(e^{h\nu/k_B T} - 1)} = \frac{2\pi}{\Omega_B} \frac{h\nu}{(e^{h\nu/k_B T} - 1)}$$

where D_{diff} is the diffraction-limited directivity of the element of area A . We now recognize the factor $2\pi/\Omega_B$ as the number of spatial modes coupled into an antenna whose orthogonal-mode beamwidth is Ω_B . So the Planck law is consistent with the antenna theorem. Of course, rarely, if ever, is a THz sensor designed with a hemispherical field-of-view. So a more practical form of Planck's law is simply

$$\frac{dP}{d\nu} = \frac{\Omega_{FOV}}{\Omega_B} \frac{h\nu}{(e^{h\nu/k_B T} - 1)}$$

This is particularly useful for collecting cones, integrating cavities, and other antenna-like structures.

D. Illustrative Examples of Modal Function

For WR-10 rectangular waveguide, we can compute the number of modes at any frequency knowing its dimensions (0.10 inch wide x 0.05 inch high) and Eqn 15. The number of modes is found by incrementing m and n , sorting by the cut-off frequencies and then counting all entries including the degenerate cases. The resulting distribution is plotted in Fig. 10. Clearly, it is approaching a quadratic dependence as the frequency increases. For analytic purposes, it is very useful to fit the curve to the following form,

$$M = \text{Int}(\nu/\nu_R)^2 + 1$$

where Int is the integer function (rounding down) and ν_R is a fit frequency somewhat greater than the lowest modal cut-off frequency of 59 GHz for the TE_{10} mode. The fit

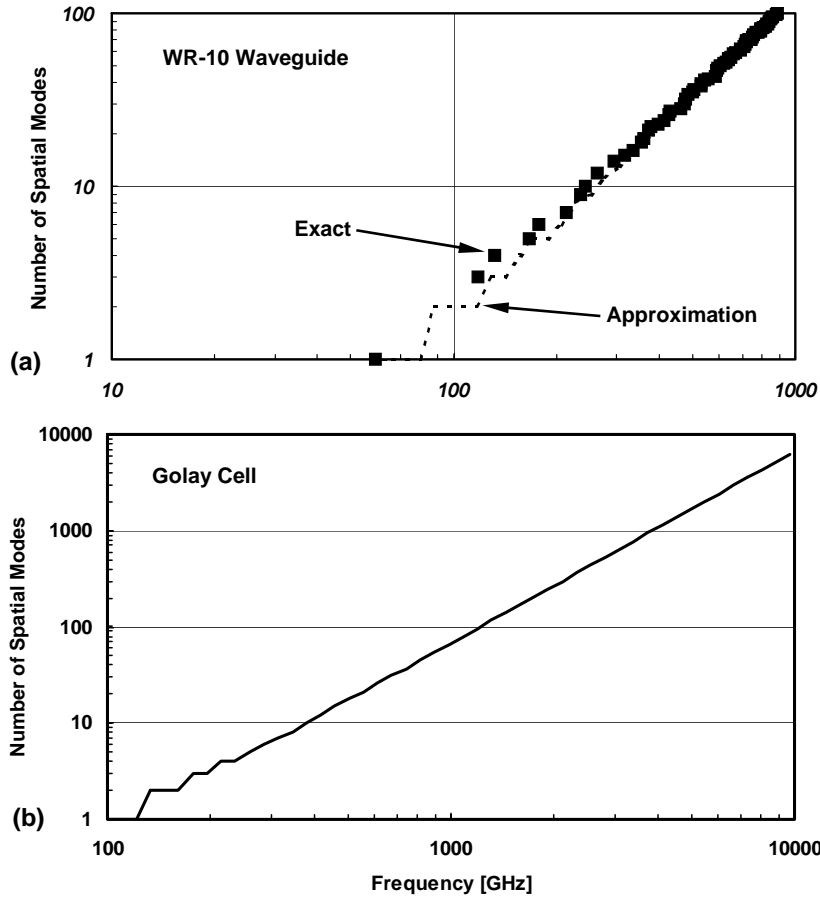


Fig. 10. Density of electromagnetic spatial modes for (a) WR-10 waveguide and (2) Golay cell having a 6-mm diameter (circular) aperture.

function is shown in Fig. 10(a) for $\nu_R = 85$ GHz.

For the lens-coupled Golay cell, we use the definition given in Eqn 22 and assume a cell diameter of 6 mm, and a lens diameter and focal length of 2 inch, respectively. The resulting modal distribution vs frequency is shown in Fig. 10(b). Because the analysis does not account for cut-off effects, the curve is perfectly quadratic and is fit very well by the same fit function as above but with $\nu_R = 0.125$ THz.

E. Gaussian beams

E.1. Gaussian-beam methodology

A key assumption behind the circular-aperture diffraction results in Sec. III.B from scalar diffraction theory is that the illumination across the aperture is uniform. This is a good assumption in some circumstances such as predicting the power collected by a receive antenna from a distance source whose pattern beam-width measured at the receive antenna is much larger than the lateral extent of the receive antenna. But there are other times when the uniform-intensity assumption is inaccurate, such as in describing the radiation transmitted or received by a fundamental-mode feedhorn.

One of the useful features of scalar-diffraction theory is its ability to predict what happens to the radiation once the uniform-illumination assumption is violated. For example, if the aperture is circular and if the illumination distribution is a Gaussian in the lateral plane with respect to the axis of symmetry, then the radiation pattern is also Gaussian, at least in the far-field limit. Intuitively, this makes sense since in this limit the Fresnel-Kirchoff integral reduces to a Fourier transform, and the Fourier transform of a Gaussian is always a Gaussian. It turns out that this result, commonly known as the Gaussian beam pattern, also applies to the near-field behavior with increasing accuracy as d/λ increases far beyond unity.

Although the Gaussian-beam result was known early in the history of electromagnetics, it was apparently not fully appreciated until the advent of the laser. The gain media in gas and solid-state lasers typically have very large values of d/λ , and generally emit much greater intensity at the center than at the lateral edges. A useful way to analyze the Gaussian behavior is to model the gain medium with a quadratic complex refractive-index lateral profile.³² The resulting intensity, described mathematically below, displays a Gaussian dependence on the lateral (r in cylindrical coordinates) dimension, and azimuthal symmetry about the propagating (z) axis.

In the mm-wave and THz region, the applicability of Gaussian beams is less obvious on first glance, but becomes plausible when one considers the coupling between antennas and circuits. MM-wave and THz antennas are often operated in their fundamental spatial modes for which the radiation intensity is maximum but rather slowly varying along the propagation axis. Then at the characteristic angle $\theta \approx \beta/2$ away from the axis the radiation begins falling rapidly in the lateral directions, with some radiation inevitably occurring at larger angles in the sidelobes because of diffraction. All these properties except the undulation of the sidelobes are described rather well by a Gaussian function vs r with perfect azimuthal symmetry about the z axis. This is the so-called fundamental Gaussian mode. Other possible symmetric functions, such as a sech^2 or Lorentzian, are either too steep about the propagation axis or decay too slowly at large angles.

MM-wave and THz antennas are often operated in fundamental mode for practical reasons. One reason is that the fundamental mode is generally the most symmetric and has the smallest beam width of all possible antenna modes. Another reason is that the devices and circuits to which the antenna is coupled are designed for their own fundamental mode, be it in high-frequency transmission line or waveguide. This is usually the easiest and most effective way to design mm-wave and THz active devices and circuits, but it generally makes the coupling between the circuits and antenna

efficient only for one antenna mode, usually the fundamental mode. These considerations break down, of course, if none of the electronics or components coupled to the antenna need to process radiation at high frequencies. Such is the case, for example, in mm-wave and THz bolometers which merely rectify any incident power absorbed. Hence, bolometers can be and often are mounted in multimode antenna-like structures, such as integrating cavities, to maximize the sensitivity and spectral bandwidth.

Finally, it is important to realize that any Gaussian beam is just one of an infinite number of modes forming an orthonormal TEM basis set. This basis set is as applicable to representing arbitrary radiation in free space as any other orthonormal basis, and must comply with the antenna theorem defined above in Sec. III. C. Each function of the basis set differs from the others in the degree of azimuthal symmetry about the propagation direction. As alluded to above, the fundamental or TEM₀₀ mode is the only one with perfect azimuthal symmetry, and is therefore the most popular and useful in solving free-space propagation problems in mm-wave and THz systems.

E.2. Gaussian-beam formulation

One benefit of the Gaussian-beam approach is it tends to add improved accuracy in mm-wave and THz design with only a minor increase in difficulty. This is because Gaussian beams, much like Gaussian distributions in probability theory, behave well mathematically under system-level operations. Given a propagation direction along the z axis, the fundamental TEM₀₀ Gaussian beam is given in cylindrical coordinates (r,θ,z) by

$$E(r, z) = E_0 \frac{\omega_0}{\omega(z)} \exp[-r^2 / \omega^2] \cdot \exp[-j(kz - \phi + kr^2 / 2R)] \quad (23)$$

where E₀ is the maximum electric field amplitude, η is the intrinsic impedance of the medium of propagation, φ is a phase constant, ω is the radius where the intensity drops by e⁻¹ relative to the on-axis intensity (i.e., the “spot size”), or “beam waist,” and R is the radius of curvature. Fortunately, all of these quantities are inter-related through simple algebraic expressions:

$$\begin{aligned} \omega^2(z) &= \omega_0^2 \cdot \left[1 + \left(\frac{\lambda \cdot z}{\pi \cdot \omega_0^2} \right)^2 \right] \\ R(z) &= z \cdot \left[1 + \left(\frac{\pi \cdot \omega_0^2}{\lambda \cdot z} \right)^2 \right] \\ \phi(z) &= \tan^{-1} \left(\frac{\lambda z}{\pi \omega_0^2 n} \right) \end{aligned} \quad (24)$$

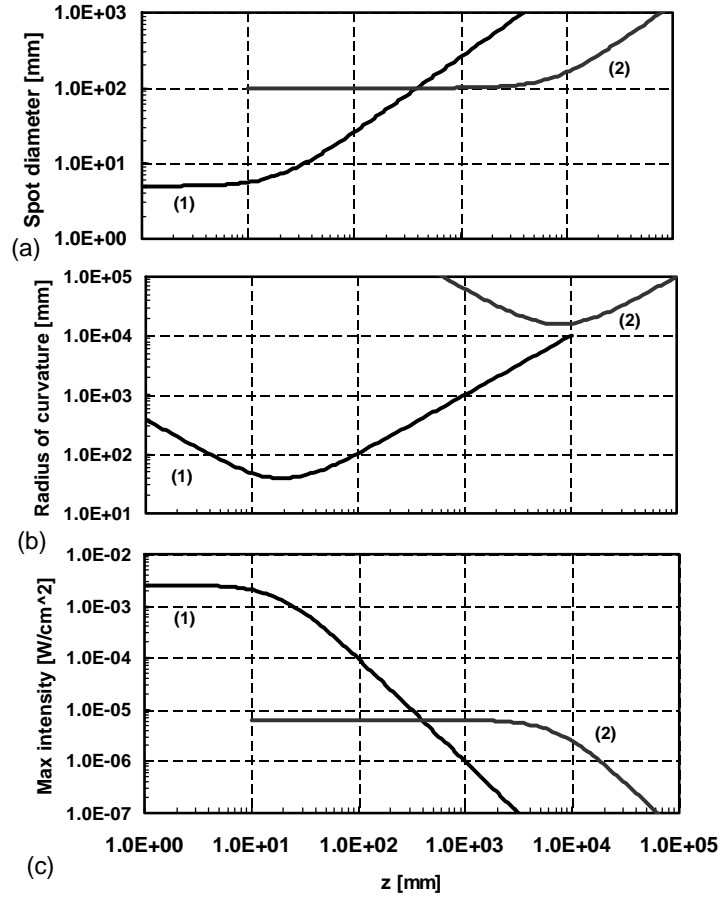


Fig. 11. Gaussian parameters: (a) spot diameter ($2 \cdot \omega_0$) (b) radius of curvature, and (c) on-axis intensity for two typical mm-wave or THz situations: (1) $\lambda = 1$ mm, $P = 1$ mW, minimum spot size at beam waist = 2.5 mm; (2) $\lambda = 1$ mm, $P = 1$ mW, minimum spot size at beam waist = 50 mm.

From these expressions it is clear that $\omega(z) \geq \omega_0$, so that ω_0 is the minimum spot radius and the maximum electric field occurs in the plane of constant z where $\omega(z) = \omega_0$ which defines the “beam waist.” In this plane the following simple form of the electromagnetic intensity is valid:

$$I(r) = \frac{E_0^2}{2\eta} \cdot \exp[-2r^2 / \omega_0^2]$$

where η is the intrinsic impedance of the propagation medium. Because the Gaussian integral is analytic if taken from $r = 0$ to ∞ , the following useful relationship exists between the total propagating power and on-axis intensity at the beam waist:

$$I(r = 0) = P \cdot \left(\frac{2}{\pi \cdot \omega_0^2} \right)$$

Further satisfying properties of the Gaussian beam follow from Eqns 23 and 24 in the limit of positive $z \gg \frac{\pi \cdot \omega_0^2}{\lambda} \equiv z_0$. In this case, $R(z) \rightarrow z$, $\omega^2 \rightarrow$

$$\left(\frac{\lambda \cdot z}{\pi \cdot \omega_0} \right)^2 = \left(\frac{\lambda \cdot R}{\pi \cdot \omega_0} \right)^2, \theta \rightarrow \pi/2, \text{ and the complex exponential approaches}$$

$\text{jexp}[-\text{j}kz(1+r^2/2z^2)]$. Hence, provided that the observation point is such that $r < z$ or R , the resulting electric field behaves as

$$E(r, z) \rightarrow E_0 \frac{\text{j}\pi\omega_0^2}{\lambda R} \exp[-(\pi\omega_0 r^2 / (\lambda R)^2)] \cdot \exp[-j(kR)]$$

which is the expected form of a spherical wave weighted by a beam-pattern function. The important distance parameter z_0 is called the Rayleigh length. It also happens to be the distance for which $R(z)$ equals its minimum value.

Another satisfying property is found by tracking the $r = \omega$ or $1/e$ profile of the beam in the “far-field” limit $z \gg z_0$. In this case the $r = \omega$ locus asymptotically approaches a full angle (centered about the z axis) of

$$\theta_B = 2 \cdot \tan^{-1} \left(\frac{\omega}{z} \right) \rightarrow 2 \cdot \tan^{-1} \frac{\lambda}{\pi \cdot \omega_0} \approx \frac{2 \cdot \lambda}{\pi \cdot \omega_0}$$

Redefining ω_0 as the lateral extent d of the Gaussian beam at its minimum aperture, we see once again a dependence of the far-field behavior on the ubiquitous “diffraction” ratio, λ/d .

All of these properties are exemplified in the curves shown in Fig. 11 for two representative Gaussian beams propagating in free space, one with a minimum spot diameter $2\omega_0 = 5$ mm (0.2 inch) and the other with $2\omega_0 = 100$ mm (4 inch). The wavelength of both is 1 mm and beam waists occur at $z = 0$. For the smaller-waist beam, the Rayleigh length z_0 is only ≈ 20 mm so the beam quickly diverges to a full divergence angle of about 14° . For the larger-waist beam, z_0 is about 8 m, so the beam remains highly collimated out to this distance and then begins to approach a divergence angle of just under 1° . Note that this larger beam would not be too difficult to support by a relatively simple telescope of aperture about 8 inches or more in diameter – that is, a man-portable instrument.

The behavior of the larger beam in Fig. 11 illustrates an important potential advantage of mm-wave and THz propagation over that in the lower RF bands. Namely, in applications where the remote sensor supports a Gaussian beam and the object or target is at a *short range* not much greater than the Rayleigh length, the divergence of the beam between the two can be very small. And thus the intensity will drop far slower than the $1/r^2$ spherical-wave behavior in the “far-field” of every common antenna. As will be shown later in Sec. X, this has important implications on the sensitivity and spatial resolution of practically any remote sensor.

E.3. Transformation of Gaussian beams: An representative system example

A second benefit of the Gaussian beam approach is its tendency to remain Gaussian through transformation by various optical two-port components, such as lenses and mirrors. As in microwave network theory, passive optical two-ports can be represented by a number of different 2x2 matrix formulations depending on the physical formulation of the propagating electromagnetic mode. A common formulation in optics is the “ray,” represented by a column or row vector $[r(z), r'(z)]$, where r is the distance from the propagation axis and r' is the slope of the ray with respect to this axis. Optical components are represented by 2x2 ABCD matrices, and an input ray is transformed according to

$$\begin{pmatrix} r_{out} \\ r'_{out} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_{in} \\ r'_{in} \end{pmatrix}$$

Some good examples of such matrices are: (1) free space path of length L , $A=1, C=0$, and $D=1$; (2) thin lens of focal length f : $A=1, B=0, C=-1/f, D=1$. The accuracy of this formulation is best for “paraxial” rays, i.e., those propagating close to the optical axis.

Remarkably, the ABCD representation also applies to Gaussian beam propagation through the definition of a complex Gaussian beam parameter

$$\frac{1}{q(z)} = \frac{1}{R(z)} - j \frac{\lambda}{n\pi\omega^2}$$

The transformation equation is then given by

$$q_{i+1} = \frac{Aq_i + B}{Cq_i + D}$$

where q_i is the beam parameter in a plane $z = z_i$. Note that the free-space ABCD matrix simply transforms as $q_{i+1} = q_i + L$.

As an illustrative example of the Gaussian beam approach, Fig. 12 shows the modeling of the radiation propagation in a system familiar to the author: the THz photomixing spectrometer. In this case, coherent mm-wave and THz radiation is generated selectively by beating two frequency-offset lasers in a small photoconductive element mounted at the driving point of a planar antenna. The planar antenna is located on a semi-insulating GaAs or InP substrate – both having very low absorption in the THz region but difficult to make antennas with because of their high dielectric constant, $\epsilon_r \approx 13$. Therefore, the photomixer substrate is abutted to the back-side of a high-resistivity Si hyperhemisphere as discussed in Sec. III.B. The radiation coming out of the hyperhemisphere will likely be diverging, so a second focusing lens (e.g., plastic) is added at some distance to focus the radiation down to a beam waist. Since the purpose of the spectrometer is to provide radiation to a sample cell for THz spectroscopic analysis, an interesting question is if and where this beam waist will occur, and how big the minimum spot size will be.

The solution is found by first estimating the pattern coming out of the planar

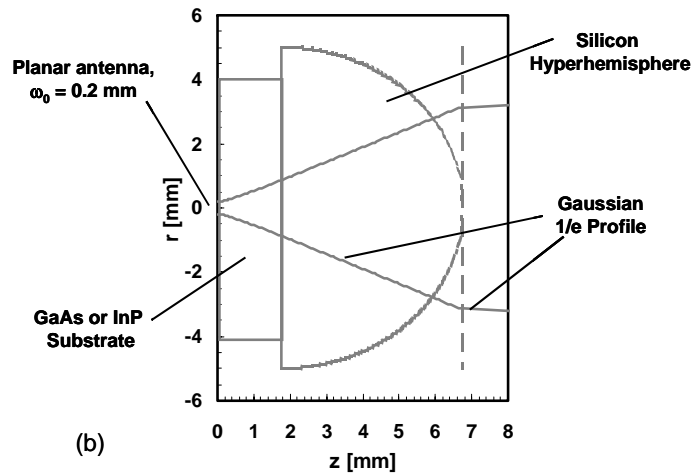
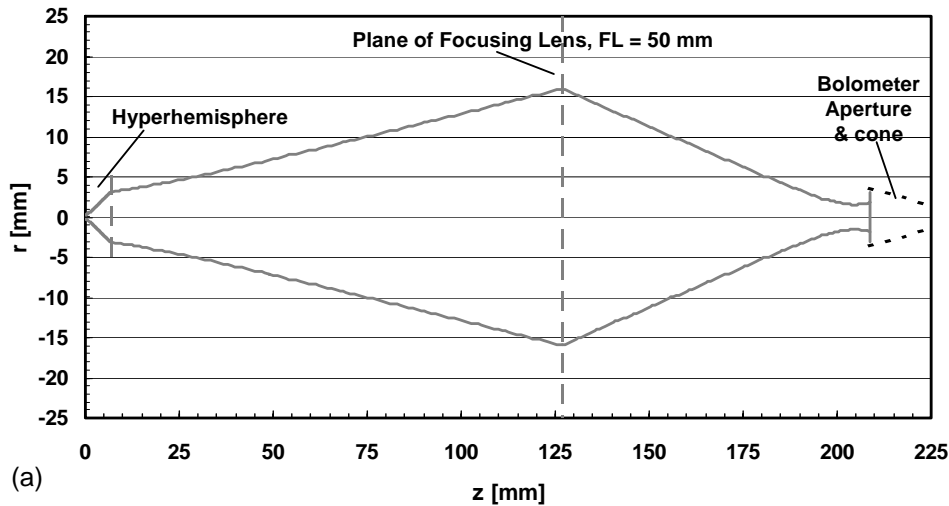


Fig. 12. (a) Results of Gaussian-beam design of THz photomixer spectrometer as a representative problem in system-level free-space radiation transformation and coupling. (b) Exploded view of photomixer region consisting of planar antenna, semi-insulating substrate and Si hyperhemisphere.

antenna as an equivalent Gaussian beam as shown in the exploded view of Fig. 12 . The two curved loci in this view represented the $r = \omega$ points. The beam propagates through the GaAs or InP substrate, into the Si hyperhemisphere, and is then transformed into free space using an ABCD matrix appropriate to a spherical-dielectric interface.³³ To avoid significant total-internal reflection, the hyperhemisphere can provide only a slight transformation of the beam, which remains diverging after passage through the Si-air interface.

The free-space Gaussian beam then propagates to the (plastic) plano-convex lens which transforms the diverging Gaussian beam to a converging beam through the application of the thin-lens ABCD matrix. By judiciously varying the hyperhemisphere-to-plano-convex separation, we achieve the beam waist shown in Fig. 12. The waist shown has a minimum spot size $\omega_0 \approx 1.5$ mm, consistent with the given frequency (300 GHz), the planar-antenna spot size (0.2 mm), the radius and set-back of the Si hyperhemisphere (5.0 and 1.76 cm, respectively), and the focal length and diameter of the plastic lens (2.0 and 2.0 inch, respectively).

F. Friis' Free-Space Transmission Equation

As discussed above, sensors in the THz region often use antennas or antenna-like structures to couple radiation to and from free space. If the antennas are unimodal, as is often the case, one can apply the concepts and terminology of RF free-space link theory, developed originally by Friis for communications systems operating at much lower frequencies. The theory assumes that the target or sample the sensor is detecting is in the “far field” of the transmit antenna, and the receive antenna is in the far field of the target or sample. In this case, even if the radiation is in the form of a Gaussian beam, the transmitted and receive waves both approach a spherical wave form, $E = E_0 e^{-jkr}/r$, so that the Poynting-vector magnitude goes as $1/r^2$. And then because of the inherent reciprocity of all antennas, it makes sense to represent the receive antenna by an effective aperture A_{eff} , which is defined by

$$P_{\text{rec}} = A_{\text{eff}} |\bar{\bar{S}}_r(\theta_r, \phi_r)| \cdot \varepsilon_p$$

where P_{rec} is the power available to the antenna for delivery to a load, $\bar{\bar{S}}_r(\theta_r, \phi_r)$ is the average Poynting vector for incoming radiation along the direction (θ_r, ϕ_r) in the spherical coordinates centered at the *receiving* antenna, and ε_p is the polarization coupling efficiency. Although this expression is prevalent in electromagnetics and communications texts, it is strictly valid *only when* $\bar{\bar{S}}_r(\theta_r, \phi_r)$ *is aligned with the direction of the beam-pattern maximum*. When there is mis-alignment, another factor is required which is the just the receive beam-pattern

$$P_{\text{rec}} = A_{\text{eff}} \cdot F_r(\theta_r, \phi_r) |\bar{\bar{S}}_r(\theta_r, \phi_r)| \cdot \varepsilon_p .$$

Now suppose that this *received* Poynting vector is generated by a second, *transmitting* antenna. We can relate the received power to the properties of the transmitting antenna by

$$|\bar{\bar{S}}_r(\theta_r, \phi_r)| \equiv |\bar{\bar{S}}_t(r, \theta_t, \phi_t)| = \tau \frac{P_{\text{rad}} D_t \cdot F_t(\theta_t, \phi_t)}{4\pi r^2} \equiv \tau \frac{P_{\text{inc}} G_t \cdot F_t(\theta_t, \phi_t)}{4\pi r^2}$$

where the subscript “t” is for transmitting, P_{rad} is the total radiated power, P_{inc} is the power used to drive the transmitting antenna (in the matched case, equal to P_{rad}), θ_t and ϕ_t are the spherical angles in the spherical coordinate system centered at the transmitting antenna, r is the distance between transmitter and receiver, and τ is the path power transmission factor. In writing this expression it is understood that F_t is taken in the direction (θ_t, ϕ_t) pointing towards the receiver, which is not necessarily the direction of the

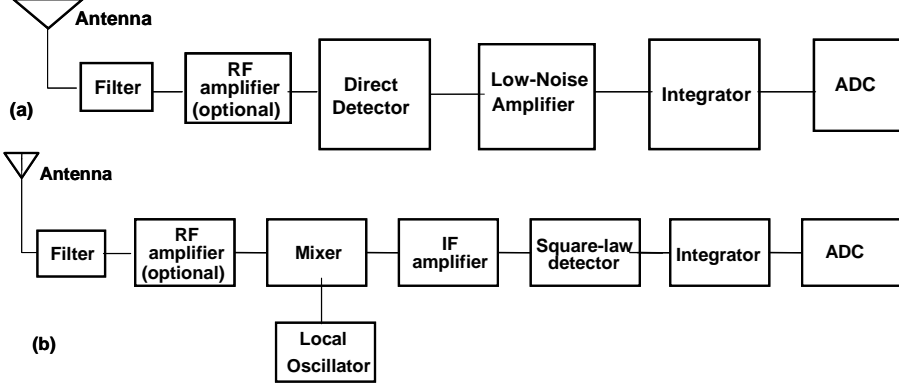


Fig. 13 (a) Direct detection receiver and (b) heterodyne (or homodyne) receiver.

maximum of F_t . Substitution into the expression for P_{rec} yields the relationship

$$P_{rec} = A_{eff} \frac{P_{in} G_t \cdot F_r(\theta_t, \phi_t) F_t(\theta_r, \phi_r)}{4\pi r^2} \tau \cdot \epsilon_p$$

This can be simplified further in terms of the (ostensibly) known parameters of the receiving antenna using the relationships,

$$P_{out} = P_{rec} \frac{G_r}{D_r} \equiv P_{rec} \frac{G_r}{4\pi A_{eff} / \lambda^2}$$

where P_{out} is the power delivered to the load of the receiving antenna. Substitution of P_{rec} into P_{out} yields

$$P_{out} = P_{in} \left(\frac{\lambda}{4\pi r} \right)^2 G_r G_t \cdot F_r(\theta_r, \phi_r) F_t(\theta_t, \phi_t) \cdot \tau \cdot \epsilon_p$$

an expression commonly known as Friis' formula, in which the *dependence on A_{eff} is cancelled out* by going from the input port to the output port. Friis' formula effectively treats the combination of antennas like a two-port network with the pattern angular dependence, polarization dependence, and path transmission included explicitly. The term $(\lambda/4\pi r)^2$ is called the free-space loss factor, which is of considerable practical and historical importance.

IV. THz Receiver Types and Performance Metrics

A. Architectures

As discussed in the Introduction, an important defining criterion for THz sensor

modality is receiver architecture. There are two types, incoherent and coherent, both shown schematically in Fig. 13. The incoherent architecture is as old as RF technology itself and the coherent came soon thereafter, dating back to the early part of the 20th century. Like many other system architectures, they persist largely by the ability of engineers to continually improve performance by perfecting the components.

A.1. Direct

The block diagram of a generic direct receiver is shown in Fig. 13(a). The incoming radiation from the target, be it thermal emission or transmitted power from the sensor itself, is collected by the receiver where it is rectified from RF (THz) to baseband by a “direct” detector. In most practical cases the baseband is defined by amplitude or frequency modulation of the incoming signal to reduce the effect of gain drifts and 1/f noise that occurs in the THz electronics. The rectified THz signal is then amplified and demodulated down to DC using synchronous detection. For AM modulation the synchronous detection is often carried out using a lock-in amplifier.

In the THz region the direct detector is almost always a power-to-voltage or power-to-current converting device. That is, it is a device that puts out a voltage or current in proportion to the incoming power. There are many examples of such devices, but the most popular are field detectors and bolometers. Field detectors, such as Schottky diodes, respond directly to the THz electric field and generate an output current or voltage through a quadratic term in their current-voltage characteristic. Bolometers are composite devices consisting of a THz absorber and a thermistor. The THz absorber is generally isolated thermally from the environment so that the absorbed THz power raises the temperature both of the absorbing layer and an attached thermistor. The thermistor is, by definition, a device that displays a large change of resistance to a small change of temperature. In some bolometers, such as the composite type, the absorber and thermistor are separate elements. In other bolometers, such as the hot electron type, they are integrated into the same device.

A key factor in all direct detectors is spectral bandwidth. As in most analyses of signal processing, we assume here that this is band limited between ν_0 and $\nu_0 + \Delta\nu$. This can be a real bandwidth defined by a THz bandpass filter, or it might be an approximation to a real spectrum.

A.2. Preamplified Direct

One of the most successful areas of RF electronics during the past decade has been monolithic microwave integrated circuits (MMICs). By integrating active devices, some passives, and matching circuits on the same semi-insulating substrates, it has become possible to fabricate low-noise amplifiers (LNAs) up to frequencies of 100 GHz and beyond.^{34,35} For example, LNAs having a gain of 17 dB, bandwidth of 30 GHz, and a noise figure of 6 dB have been fabricated and tested around 94 GHz.³⁶ The advantage of an LNA is that if it has adequate gain, it can dominate the noise figure of the following square-law detector, leading to a much lower NEP than can be achieved by direct conversion using the same square-law detector. As will be shown later, the sensitivity of the preamplified direct receiver can then approach the photon-noise (quantum) limit.

A.3. Heterodyne

In the coherent system of Fig. 13(b), incoming radiation from the target, be it thermal emission or transmitted power from the sensor itself, is combined with power from a local oscillator (LO) on a THz mixer. If the signal and LO frequency are different, there will be a beat-note generated at an intermediate frequency (IF) between the two. This is called *heterodyne* conversion. If the signal and LO frequency are equal, the beat tone degenerates to dc, and the process is called *homodyne* conversion. Independent of the conversion process, all coherent detectors require a device that can generate an *efficient* conversion of the RF power to the IF band. The most popular mixers are field-type devices having a strong quadratic nonlinearity. Good examples are Schottky diodes, superconductor-insulator-superconductor (SIS) tunnel junctions,^{37,38} and superconducting hot-electron bolometers.³⁹

Coherent down-conversion has several unique features that distinguish it from direct detection. First, mixing a weak signal with a relatively strong LO effectively amplifies the received signal relative to the receiver noise floor, which can greatly improve the sensitivity compared to direct detection. Second, for the typically weak signals in the THz region, the mixing process is linear. That is, the signal power at the IF frequency is linearly proportional to the signal power at the input. Therefore, the receiver passband can be defined by an IF band pass filter, which is generally much lower in cost and has much higher performance than any THz filter. This feature tends to make coherent receivers the favored approach in applications requiring high spectral resolution, such as molecular spectroscopy.⁴⁰ But as will be shown below, the direct receiver tends to be preferable in wideband applications such as thermal imaging because of its superior spectral bandwidth and simplicity.

A.4. Pre-Amplified Heterodyne

An intriguing possibility in the millimeter-wave band and lower end of the THz range is a preamplifier feeding a mixer element. With recent advances in MMIC solid-state power amplifiers and the possibility of integrating them with high-frequency Schottky mixers monolithically, one can envision a receiver in which an antenna couples radiation to an LNA that, in turn, is coupled to a high-frequency mixer. The mixer then down-converts the radiation to whatever IF band makes sense, be it narrowband for spectroscopic applications, or wider band for thermal imaging applications. Intuitively, one could design the LNA with just enough gain so that the overall receiver sensitivity was not affected significantly by the mixer or following IF electronics. As will be shown later, this provides excellent overall performance if the LNA noise figure is acceptably low. And not surprisingly, it is the same architecture used at lower frequencies in communications and radar receivers alike, perhaps most commonly in the handsets of nearly every mobile telephone made today at PCS wireless frequencies.

B. Signal-to-Noise Ratio and Noise Equivalent Power

Once the THz signature is defined and the free-space link is established, it is the job of the sensor to process the signal with maximum fidelity and minimum

contamination or masking by noise. From the above discussion it should be clear that the signal is the information-bearing portion of the incident THz average power, whether the power is coherent or incoherent. The noise is then all other electromagnetic mechanisms that cause the power to fluctuate about the average and, therefore, become less certain. And from Sec. II such noise is always present, even in the absence of electronic noise. So as in all other electromagnetic sensors, the analysis of sensor performance must be carried out by accounting for the noise, and it is important to define a metric for the sensitivity in the presence of the noise.

In principle, at any point in either sensor architecture of Fig. 13, it is possible to define the average power associated with the signal and the rms fluctuations in power associated with the noise. So a useful metric of the ability of the sensor to distinguish the signal from the noise is the *power* signal-to-noise ratio (SNR),

$$\frac{S}{N} = \frac{\langle P \rangle}{\sqrt{\langle (\Delta P)^2 \rangle}} = \frac{\langle P \rangle}{S_p \cdot B_{ENB}}$$

where S_p is the power spectral density and B_{ENB} is the equivalent noise bandwidth at that point in the sensor. B_{ENB} is generally dictated by sensor phenomenology, such as the resolution requirements and measurement time.

As discussed in Secs. I and IV above, there are two types of receivers used in the THz region: (1) incoherent, or direct, and (2) coherent, or heterodyne. From a signal processing standpoint, the *detector* in both types of receivers is the device that converts the signal from the RF (i.e., THz) region to baseband (usually around DC) where it can be positively identified in comparison to the residual baseband fluctuations that comprise the noise. For incoherent receivers, the *detector* is often the first device in the RF front-end, connected directly to the antenna. In coherent receivers, the *detector* is generally a device in the IF part of the circuit, after an RF front-end that uses a mixer to down-convert the RF signal from the THz region to an IF band at much lower frequency. The commonality between the incoherent and coherent detectors is that they are generally the first *nonlinear* element in the receiver chain – at least the first nonlinear element by design. This may appear contradictory to the definition of a mixer until we realize that the signal power in any coherent receiver is almost always much weaker than the local oscillator power. Under this condition, a mixer acts as a linear (in amplitude) transducer that simply translates the frequency.

As in any RF system, it is the signal-to-noise ratio *after detection* that matters most. This is what determines the bit-error-rate in communications systems and the probability of detection in radar systems, for example. This will become more clear later in the Sec. VIII.B on receiver operational characteristics. So for the present discourse, we will always seek the SNR after detection. Clearly, the signal-to-noise ratio depends on the signal strength so is not by itself a good metric for comparing detector or sensor types.

C. Noise Equivalent Power

A good figure-of-merit for overall sensor sensitivity comes from the fact that, generally speaking, only post-detection SNR values of order unity or higher are useful. This is intuitively obvious, but can also be proven mathematically by statistical decision theory. A more useful metric for sensor performance is to fix the SNR at some value and then solve for the signal power that achieves this. The fixed value of SNR universally

accepted by sensor engineers and scientists is unity. The resulting metric is the noise-equivalent power spectral density, NEP, which by definition is the *input* signal power to the sensor required to achieve a SNR of unity after detection (AD). As we will show later, square-law detection of a signal buried in additive white Gaussian noise has the effect of increasing the SNR by

$$\left(\frac{S}{N}\right)_{AD} = \left(\frac{S}{N}\right)_{BD} \sqrt{\frac{\Delta\nu}{2\Delta f}}$$

where $\Delta\nu$ and Δf are the pre- and post-detection bandwidths, respectively, and $(S/N)_{BD}$ is the SNR before detection. Thus, the NEP is given by (in units of W)

$$NEP_{AD} = N_{BD} \cdot \sqrt{\frac{2\Delta f}{\Delta\nu}} = NEP_{BD} \cdot \sqrt{\frac{2\Delta f}{\Delta\nu}}$$

For the purpose of comparing different sensor technologies, it is conventional to divide out the post-detection bandwidth effect (or equivalently, setting it equal to 1 Hz). This yields the normalized NEP'_{AD} [in units of $W/(Hz)^{1/2}$], which we define as

$$NEP'_{AD} = N_{BD} \cdot \sqrt{\frac{1}{\Delta\nu}}$$

Generally, the pre-detection NEP scales with $\Delta\nu$, so we introduce a normalized NEP'_{BD} ,

$$NEP'_{AD} = NEP'_{BD} \cdot \sqrt{\Delta\nu}$$

Note that these definitions are not spelled out explicitly in the most of the literature on remote sensing. But we do so here to help students or new researchers to the field since the NEP is one of the most bewildering of all system metrics.

A useful feature of the NEP is its additivity. If there are N mechanisms contributing to the noise at a given node in the receiver, if the mechanisms are uncorrelated to the signal and to each other, if they obey Gaussian statistics, then the total NEP is the uncorrelated sum

$$NEP_{TOT}^2 = NEP_1^2 + NEP_2^2 + \dots + NEP_N^2$$

This property applies to any node, pre- or post-detection, and will be used explicitly when we discuss the contribution from electronic noise. In reality there are cases where a noise mechanism is correlated to the signal (e.g., radiation noise) or to another noise mechanism (current and voltage noise in transistors), so one must be careful in applying this addition formula. In such cases, one can always fall back on the SNR as a useful measure of overall system performance.

D. Noise Equivalent Delta Temperature

For radiometric and thermal imaging systems, it is sometimes convenient to express the sensitivity in terms of the change of temperature of a thermal source that produces a post-detection SNR of unity. The resulting metric is the noise equivalent delta temperature, or NE ΔT , which is given by⁴¹

$$NE\Delta T = \frac{NEP_{AD}}{dP_{inc}/dT|_{P_B}}$$

where P_{inc} is the incident power. In the Rayleigh-Jeans limit and assuming that the source has unity emissivity and fills the field-of-view of the sensor, we find $P_{inc} = M k_B T_B \Delta\nu$, so the NE ΔT becomes

$$NE\Delta T = \frac{NEP_{AD}}{M \cdot k_B \cdot \Delta \nu}$$

E. Noise Figure and Friis Formula

For the linear elements of a sensor, including both active and passive components, a more universal metric is the noise figure, F . This is defined for all linear two-ports as

$$F = \left(\frac{(S/N)_{IN}}{(S/N)_{OUT}} \right); 1 < F < \infty$$

In other words, the noise figure quantifies the degradation in SNR as a signal passes through a component in a linear chain. When combined with the gain of each component, the noise figure can be calculated simply from a sequence of two-ports through the expression

$$F_{TOT} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 \cdot G_2} + \dots + \frac{F_n - 1}{\prod_{i=1}^{n-1} G_i}$$

where G_i is the power gain of the i^{th} element. Note that this gain accounts for impedance mismatch between elements. Physically, this means that components located further down a chain tend to be less important if the earlier components have high gain.

V. THz Signal and Noise Processing

Once radiation associated with the object signature is coupled into the sensor receiver, it generates the signal – a voltage or current in a network that must be processed from the antenna front-end to the data-collection or decision-making (logic) devices in the back end. Independent of the coherence of the incoming radiation, there are five functions in this process that are rather generic in THz sensors, and all electromagnetic sensors for that matter, and will be discussed in some detail here: (1) rectification, or detection; (2) integration or averaging (3) matched filtering; (4) frequency conversion; and (5) amplification; either before or after detection. Rectification is always necessary to extract the information about the object of interest, which that usually resides on the THz envelope. Therefore, it is discussed first and in some detail. The two common means of rectification in THz receivers are square-law and envelope detection. Integration is almost always carried out after the rectifier to add multiple samples of the information extracted and therefore improve the quality of the data or the confidence of the decision. Matched filtering is always carried out in the linear part of the sensor before detection. Frequency down-conversion is always necessary because logic circuits do not (at least not yet) work at THz frequencies. Amplification can be carried out either in the linear part of the receiver before detection or in the IF or baseband portion after detection.

Remarkably, all five of these functions are carried out in the two common receiver architectures discussed in the Introduction: (1) incoherent, and (2) coherent. We start with a discussion of the detection function since this is the centerpiece of most THz sensors.

A. Classical Square-Law Detection and Integration

Simply stated, a square-law detector is a device or circuit that takes an input signal and produces an output that is proportional to its square, $X_{out} = AX_{in}^2$, where X_{out} could be a current or voltage and A is a proportionality constant. The utility of such a device, just like a mixer, is in rectification, or frequency down-conversion. This is most easily seen in the special case of a coherent signal $X_{in}(t) = B\cos(\omega t + \phi)$. If put through a square-law detector, the output becomes $X_{out}(t) = AB^2 (1/2)\{1 + \cos[2(\omega t + \phi)]\}$. So if the square-law detector is followed up by a low pass filter (i.e., a time-domain integrator) with integration time $\tau \gg 2\pi/\omega$, then the second term will not contribute and the output of the filter will be

$$X_{out} = AB^2/2 \equiv \mathfrak{R}P_s \quad (25)$$

where \mathfrak{R} is a constant and P_s is the average absorbed input signal power. In the language of sensor theory, \mathfrak{R} is usually called the *responsivity* and is measured either in A/W or V/W. Presumably, A is known, so that B can be determined from the X_{out} dc term. Square-law detectors are preferred in RF systems over cubic and other possible detectors for this reason. The proportionality constant A is dependent only on the detector characteristics and not on the power level, at least up to a level where saturation and higher-order effects begin to occur. A single calibration of A in the “small-signal” regime is all that is required to use the square-law detector over a wide range of input power.

We assume that the output noise from the detector is caused by an input white noise spectrum unrelated to the signal. We note that Fourier components at different frequencies correspond to different longitudinal modes, which in a random-noise (e.g., thermal) power spectrum are uncorrelated. Thus, such components will not produce an average dc component, or signal. But just as in the case of quantum noise, they will produce a fluctuation of the dc term about its average, i.e., noise. The magnitude of this noise can be determined as a sum over all possible mixing terms whose difference frequency lies within the passband of the low pass filter. If the input power spectrum S_X is flat across the pass band $\Delta\nu$, it can be shown that⁴²

$$S_{X_{out}}(f) = A^2 \int_{\nu}^{\nu+\Delta\nu} S_{X_{in}}(\nu+f) \cdot S_{X_{in}}(\nu) \cdot d\nu \approx 2A^2 (S_{X_{in}})^2 \cdot \Delta\nu \cdot (1 - \frac{f}{\Delta\nu}) = 2\mathfrak{R}^2 (\Delta P_N)^2 \frac{1}{\Delta\nu} \cdot (1 - \frac{f}{\Delta\nu}) \quad (26)$$

P_N is the noise power. The factor of two here comes from the fact that two spectral components in S_X contribute to $X(\nu')$ - one above and one below ν .

At this point, we see the merit of an integrator in the signal processing. Such a device can be analyzed as a low-pass filter of bandwidth Δf . If $\Delta f \ll \Delta\nu$, which is almost always the case in practice, then $S_{X_{out}}$ can be considered flat over the integrator passband and we get, by linear signal processing,

$$(\Delta X_{OUT})^2 \equiv \int_0^{\infty} H(f) S_{X_{out}}(f) df \approx 2A^2 (S_{X_{in}})^2 \cdot \Delta\nu \cdot \Delta f = 2\mathfrak{R}^2 \cdot (\Delta P_N)^2 \cdot \Delta f / \Delta\nu \quad (27)$$

And by taking the ratio of Eqn 25 to Eqn 27, we end up with the output signal-to-noise ratio

$$(SNR_{out})^2 = \frac{X_{OUT}^2}{\langle (\Delta X)^2 \rangle} = \frac{\Re^2(\bar{P}_S)^2}{\Re^2(\Delta P_N)^2} \frac{\Delta \nu}{2\Delta f} \equiv (SNR_{in})^2 \cdot \frac{\Delta \nu}{2\Delta f}$$

Square-law detectors also get used for *detection* when the input signal is *random noise*, as in the case of passive radiometers or noise radars. Now the input signal has a white power spectrum superimposed on the noise spectrum. In this case, the “signal” is just the portion of the power spectrum of Eqn 26 that gets converted by the detector down to baseband, assumed in this context to be dc. We know that any Fourier component of the noise spectrum is, by definition, self-correlated. The square law detector will produce a dc output from each component in accordance with the following expression

$$X_{out} = A \int_{\nu_0}^{\nu_0+\Delta\nu} \sqrt{S_{X_{in}}(\nu)} \sqrt{S_{X_{in}}(\nu)} d\nu = A \int_{\nu_0}^{\nu_0+\Delta\nu} S_{X_{in}}(\nu) d\nu$$

If the input power spectrum is flat across the passband, this results in

$$X_{out} = A \cdot S_{X_{IN}} \cdot \Delta \nu = \Re \cdot \bar{P}_N \quad (28)$$

just as for coherent signal, but now written in terms of the average noise power. Hence

$$(\Delta X_{OUT})^2 \equiv \int_0^{\infty} H(f) \cdot S_{X_{OUT}}(f) df \approx 2A^2 (S_{X_{IN}})^2 \cdot \Delta \nu \cdot B_{int} = 2\Re^2 \cdot (\Delta P_N)^2 \cdot \Delta f / \Delta \nu \quad (29)$$

And by taking the ratio of Eqn 28 to Eqn 29, we end up with the output signal-to-noise ratio

$$(SNR_{OUT})^2 = \frac{X_{OUT}^2}{\langle (\Delta X)^2 \rangle} = \frac{\Re^2(\bar{P}_N)^2}{\Re^2(\Delta P_N)^2} \frac{\Delta \nu}{2\Delta f} \equiv (SNR_{in})^2 \cdot \frac{\Delta \nu}{2\Delta f} \quad (30)$$

which is the same as for the coherent input signal.

The utility of Eqns 29 and 30 are manifest. They get applied in nearly every mm-wave and THz sensor whose *signal* is actually *random noise*. They can be understood in a different way by noting that the sampling time of the integrator is $\tau_s = 1/(2\Delta f)$, so that there are $\Delta \nu \tau_s$ independent samples per second. Because noise is random, the effect of adding these multiple signals becomes the sum of the powers, or the square root of the number of samples. Hence we can re-write Eqn 30 as

$$SNR_{OUT} = SNR_{in} \cdot \sqrt{\Delta \nu \cdot \tau_s} = SNR_{in} \cdot \sqrt{N_s}$$

From the standpoint of modern electronics, it is important to recognize that the operation of a square-law detection is a form of autocorrelation. This is easy to see in the case of thermal-noise, or AWGN in general. The Fourier components at different frequencies corresponding to different longitudinal modes can then mix together, so the output must be considered as a sum over all possible mixing terms whose difference frequency lies within the pass band of the low pass filter, $1/\tau$. A little thought shows that this is given by

$$s_{out} \propto \int s_{in}(t) s_{in}(t + \tau) dt$$

the autocorrelation function.⁴³

B. Quantum Square-Law Detection and Integration

There exist a class of THz detectors that display quantum-mechanical response to incoming radiation. They tend to be the devices that have, for some reason, a band gap in their electronic density-of-states whose energy is of order $h\nu$. Two examples are hydrogenic-impurity semiconductors (i.e. extrinsic photoconductors), and low- T_C superconductors. Their quantum mechanical response is described by the photoelectric effect, whereby the detector absorbs photons that liberate charge-carrying particles at a rate

$$r = \frac{P}{h\nu}$$

If these charged particles are electrons or holes that are liberated in an electric field, they can contribute to a net photocurrent given by

$$i = \frac{egP}{h\nu}$$

where g is the photoelectric (more commonly called “photoconductive”) gain, which may be greater or less than unity. If we think of this expression in the context of a THz photon detector, the output power must be proportional to i^2 and therefore to P^2 . In other words, the photoelectric effect naturally obeys a square-law behavior. If such a “photodetector” is used as the first stage of a THz sensor, directly coupled to the incoming radiation, one will get a response to the signal power of

$$I_{dc} = \frac{eg\bar{P}}{h\nu} \quad (31)$$

But the response to the radiation fluctuations is more complicated. Following the same line of reasoning as applied in Sec. V.A above, the photodetector can mix fluctuations of different spectral components by virtue of its square-law behavior. We can duplicate the analysis of Sec. V.A simply by making the analogies $X \rightarrow i$ and $\mathfrak{R} \rightarrow eg/h\nu$.

$$S_I(f) = A^2 \int_{\nu}^{\nu+\Delta\nu} S_{X_{IN}}(\nu+f) \cdot S_{X_{IN}}(\nu) \cdot d\nu \approx 2 \cdot \left(\frac{eg}{h\nu}\right)^2 (\Delta P_{in})^2 \frac{1}{\Delta\nu} \cdot \left(1 - \frac{f}{\Delta\nu}\right)$$

In most sensors at THz frequencies or above, the frequency f occurs in the electronic circuit that has a maximum frequency much less than $\Delta\nu$. In this case, the power spectrum of current fluctuations is given by

$$S_I(f) \approx 2 \left(\frac{eg}{h\nu}\right)^2 \frac{(\Delta P_{in})^2}{\Delta\nu} \quad (32)$$

so that processing by a low-pass filter (integrator) yields on output mean-square current of

$$(\Delta I)^2 \approx 2 \cdot \left(\frac{eg}{h\nu}\right)^2 \frac{(\Delta P_{in})^2}{\Delta\nu} \Delta f \quad (33)$$

Taking the ratio of Eqn 31 to Eqn 33 leads to the signal-to-noise ratio

$$SNR_{OUT} = \frac{I_{OUT}^2}{\langle (\Delta I)^2 \rangle} = \frac{(eg/h\nu)^2 (\bar{P}_{in})^2 \Delta\nu}{(eg/h\nu)^2 (\Delta P_{in})^2 2\Delta f} \equiv SNR_{in} \cdot \frac{\Delta\nu}{2\Delta f}$$

the same as the classical case.

It is illuminating to examine Eqn 32 for a coherent incident signal, which using the result from Sec. III.D.2. for the mean-square input power fluctuations leads to

$$S_I(f) \approx 2 \cdot \left(\frac{eg}{h\nu} \right)^2 \left(\frac{\sqrt{h\nu \bar{P}_{in}}}{\sqrt{\Delta\nu}} \right)^2 \cdot \Delta\nu$$

Substitution of Eqn 31 then yields

$$S_I(f) \approx 2egI_{dc}$$

This is the same form as the famous Schottky expression for electron shot noise in an electronic device, discussed later in Sec. VII.A. Therefore, it is called the *photon shot noise* expression. “Full” photon shot noise corresponds to $g = 1$. “Suppressed” and “enhanced” shot noise correspond to $g < 1$ and $g > 1$, respectively. This result lends credibility to our analysis of Sec. II in terms of power fluctuations.

This result has been the source of confusion to students and professionals alike because it is identical in form to the current fluctuations in electronic devices in which shot noise occurs, a good example being bipolar transistors. This has led many to believe that it is the fluctuating nature of the electrons in the photodetector that give rise to the photon shot noise. Hopefully, the present derivation makes it clear that the source of photon shot noise is fluctuations from measurement of the radiation itself and the inherent ability of photon detectors to mix down these fluctuations into an electronic circuit.

$$(\Delta i)^2 \equiv \int_0^{\infty} H(f) S_{X_{OUT}}(f) df \approx 2A^2 (S_{X_{IN}})^2 B_{int}$$

C. Classical Heterodyne and Homodyne Conversion and Detection

As discussed in Sec. IV, heterodyne and homodyne detection entail the mixing of an input signal, be it coherent or incoherent, with a local oscillator. The mixing device is usually one that, although not necessarily perfect square-law, has a large quadratic coefficient such that the current or voltage in the output port is proportional to the input current or voltage squared, $X_{out} = AX_{in}^2$, where A is a proportionality constant. This is exactly the relation we had for a rectifier. For a mixer, we change X_{in}^2 to $X_{in}X_{LO}$, and for $X_{in} = B\cos(\omega_{in} t)$ and $X_{LO} = C\cos(\omega_{LO} t)$, we find

$$X_{out}(t) = A \left[\frac{B^2}{2} (1 + \cos 2\omega_{in} t) + \frac{C^2}{2} (1 + \cos 2\omega_{LO} t) + BC \cos(\omega_{LO} - \omega_{in})t + BC \cos(\omega_{LO} + \omega_{in})t \right]$$

The factor $\cos(\omega_{LO} - \omega_{in})t$ is the intermediate-frequency (IF) term. In the THz region the mixer is generally operated such that $\omega_{LO} \gg |\omega_{LO} - \omega_{in}|$ and $\omega_{LO} + \omega_{in} \gg |\omega_{LO} - \omega_{in}|$, so that the only remaining terms are

$$X_{out}(t) = A \left[\frac{B^2}{2} + \frac{C^2}{2} + BC \cos(\omega_{LO} - \omega_{in})t \right] = \eta \Re \left[P_{in} + P_{LO} + 2\sqrt{P_{in}P_{LO}} \cos(\omega_{LO} - \omega_{in})t \right] \quad (34)$$

where P_{in} and P_{LO} are the incident input and LO powers, and η is the fraction of each that is usefully absorbed. The output power at the IF frequency for the heterodyne case is given by

$$\overline{P}_{out} = D \overline{X}_{out}^2 = D \cdot (\eta \mathfrak{R})^2 \cdot 2P_{in} \cdot P_{LO} \quad (35)$$

since the long-term time average of $\cos^2(\omega t) = 1/2$ for any ω . The quantity D is the mixer-to-IF circuit matching factor that accounts for any mismatch between the mixer and IF load impedance. This expression clearly displays a linear input-output relationship. It is customary to lump all parameters other than P_{in} and η into one quantity called the mixer conversion gain, G_{mix}

$$\overline{P}_{out} = D \cdot (\eta \mathfrak{R})^2 \cdot 2P_{in} \cdot P_{LO} \equiv \eta^2 \cdot G_{mix} \cdot \overline{P}_{in} \quad (36)$$

In other words, G_{mix} represents the fraction of all of the absorbed incident power in one sideband that is converted to the IF. In all mixer types, \mathfrak{R} decreases as P_{LO} increases, and the impedance match can never be made perfect. Hence the maximum attainable single-sideband conversion gain is usually less than unity and the gain factor really represents a loss. Special types of mixers, such as parametric converters and certain quantum mixers can display a conversion gain > 1 , but it is rare and difficult to achieve. Similarly, the output power for the homodyne case is

$$\begin{aligned} \overline{P}_{out} &\propto \overline{X}_{out}^2 = (\eta \mathfrak{R})^2 \cdot 4P_{in} \cdot P_{LO} \\ \overline{P}_{out} &\equiv 2 \cdot \eta^2 G_{mix} \cdot \overline{P}_{in} \end{aligned}$$

which, again, is a linear input-output relationship.

The noise processing of the coherent converter is a bit more subtle. One noise process that must always be considered is the fluctuations in X_{out} caused by the quantum-mechanical fluctuations in the two *direct-detection* terms in Eqn 35, P_{in} and P_{LO} . In most if not all THz mixers, $P_{LO} \gg P_{in}$, so from Sec. V we can write

$$(\Delta X_{OUT})^2 = 2\mathfrak{R}^2 \cdot (\Delta P_{in})^2 \cdot B_{IF} / \Delta\nu$$

where the IF bandwidth B_{IF} replaces the integration bandwidth for direct detection. For practical reasons it is often true that the LO power is unimodal, so that from Eqn 12 we can write $(\Delta P_{in})^2 = \sqrt{\eta \cdot h\nu_{LO} \cdot P_{LO} \cdot \Delta\nu}$, where P_{LO} is the incident LO power. Thus we find

$$(\Delta X_{OUT})^2 = 2\mathfrak{R}^2 \cdot \eta \cdot h\nu_{LO} \cdot P_{LO} \cdot B_{IF} \quad (37)$$

or

$$(\Delta P_{OUT})^2 = 2D \cdot \mathfrak{R}^2 \cdot \eta \cdot h\nu_{LO} \cdot P_{LO} \cdot B_{IF} \equiv G_{mix} \cdot \eta \cdot h\nu_{LO} \cdot B_{IF} \quad (38)$$

Taking the ratio of Eqns 38 and 36, we get the heterodyne radiation-noise-limited signal-to-noise ratio:

$$\left(\frac{S}{N}\right)_{IF} = \frac{\eta^2 \cdot G_{mix} \cdot \overline{P}_{in}}{G_{mix} \cdot \eta \cdot h\nu_{LO} \cdot B_{IF}} = \frac{\eta \cdot \overline{P}_{in}}{h\nu_{LO} \cdot B_{IF}} \quad (39)$$

for the heterodyne case, and

$$\left(\frac{S}{N}\right)_{IF} = \frac{\eta^2 \cdot G_{mix} \cdot \bar{P}_{in}}{G_{mix} \cdot \eta \cdot h\nu_{LO} \cdot B_{IF}} = \frac{2\eta \cdot \bar{P}_{in}}{h\nu_{LO} \cdot B_{IF}}$$

for the homodyne case.

D. Quantum Coherent Heterodyne and Homodyne Conversion and Detection

As in the case of direct detection, the coherent conversion with quantum (photo) mixers follows from the classical analysis by the substitutions $X \rightarrow i$ and $\Re \rightarrow ge/h\nu$. The LO oscillator and input electric fields are treated classically. So from Eqn 34 we can write

$$i_{out}(t) = \eta \frac{eg}{h\nu} \left[P_{in} + P_{LO} + 2\sqrt{P_{in}P_{LO}} \cos(\omega_{LO} - \omega_{in})t \right]$$

This leads to the heterodyne IF signal output

$$\bar{i}_{out}^2 = \left(\eta \frac{eg}{h\nu}\right)^2 \cdot 2P_{in} \cdot P_{LO} \quad (40)$$

and the homodyne IF signal output

$$\bar{i}_{out}^2 = \left(\eta \frac{eg}{h\nu}\right)^2 \cdot 4P_{in} \cdot P_{LO} \quad (41)$$

The same substitutions into Eqn 37 lead to a mean-square IF current fluctuation arising from LO radiation fluctuations of

$$(\Delta i_{out})^2 = 2(eg/h\nu)^2 \cdot \eta \cdot h\nu_{LO} \cdot P_{LO} \cdot B_{IF} = 2egI_{LO} \cdot B_{IF} \quad (42)$$

for the heterodyne and homodyne cases.

Taking the ratio of Eqn 40 or 41 to 42 we get the “quantum-limited” SNR expressions

$$\left(\frac{S}{N}\right)_{IF} = \frac{\eta \cdot \bar{P}_{in}}{h\nu_{LO} \cdot B_{IF}}$$

for the heterodyne case, and

$$\left(\frac{S}{N}\right)_{IF} = \frac{2 \cdot \eta \cdot \bar{P}_{in}}{h\nu_{LO} \cdot B_{IF}}$$

for the homodyne case. Note that the quantum mechanical nature of the radiation enters these expressions through the photoelectric effect and the photon fluctuations – not through the mixing process. And because neither expression depends on the photoelectric gain or on the LO power, both are identical to the analogous expressions for classical coherent detection.

VI. Radiation-Noise Limits on Sensitivity

A. Radiation-Noise-Limited NEP_{BD} of THz Direct-Detection Sensors

Before radiation ever enters a receiver and gets converted to information, there are fundamental limits imposed on the sensitivity simply by the fact that every sensor must carry out a measurement. Measurements perturb the radiation field and, as such, introduce uncertainty. This uncertainty is evident through fluctuations of the radiation described in this section. These fluctuations are rather independent of the sensor architecture and depend, instead, on the nature of the radiation being detected.

A.1. Coherent Signal

In practice a THz sensor that receives a coherent signal must contend with radiation noise at least at the level of photon shot noise from the signal. For radiation in a single mode, this leads to a maximum power signal-to-noise ratio of

$$\frac{S}{N} \equiv \frac{P_{abs}}{\sqrt{(\Delta P_{abs})^2}} = \frac{\eta \cdot P_{inc}}{\sqrt{\eta \cdot h\nu \cdot P_{inc} / \Delta\nu \cdot \Delta\nu}}$$

Solving for P_{inc} with $S/N = 1$, we find the before-detection NEP_{BD} of

$$NEP_{BD} = \frac{h\nu}{\eta} \Delta\nu,$$

which is the famous photon- or quantum-noise limited expression.

The NEP after classical or quantum noise-free square-law detection follows from Sec. V.A or V.B,

$$NEP_{AD} = NEP_{BD} \cdot \sqrt{\frac{2\Delta f}{\Delta\nu}} = \frac{h\nu}{\eta} \sqrt{2\Delta\nu \cdot \Delta f} \quad (43)$$

and the specific NEP is simply

$$NEP'_{AD} \equiv \frac{NEP_{AD}}{\sqrt{\Delta f}} = \frac{h\nu}{\eta} \sqrt{2\Delta\nu}$$

A.2. Thermal Signal

A common application in THz and infrared sensors alike is remote detection of a thermal signal. In this case the maximum signal-to-noise ratio occurs when the only source of noise is the radiation noise of the signal. Again, we assume that the passband is narrow enough that the Planck factor is the same for all modes, and that each mode is equally well matched to the sensor. We then find

$$\bar{P}_{abs} = \left[\sum_m^M \eta_m \right] \cdot h\nu_0 \Delta\nu \cdot f_P$$

$$\langle (\Delta P_{abs})^2 \rangle = \left[\sum_m^M \eta_m \right] \cdot (h\nu_0 \Delta\nu)^2 f_P (f_P + 1)$$

$$\begin{aligned}
\left(\frac{S}{N}\right)_{BD} &= \frac{P_{abs}}{\sqrt{(\Delta P_{abs})^2}} = \frac{[\sum_m^M \eta_m] \cdot h\nu_0 \cdot f_p \cdot \Delta\nu}{h\nu_0 \cdot \Delta\nu \cdot \sqrt{[\sum_m^M \eta_m] \cdot f_p (f_p + 1)}} \\
&= \frac{\sqrt{\sum_m^M \eta_m \cdot f_p}}{\sqrt{f_p (f_p + 1)}} = \sqrt{\sum_m^M \eta_m} \cdot \exp(-h\nu_0 / 2k_B T)
\end{aligned} \tag{44}$$

where the last step follows from the definition of f_p and remains valid at any temperature or passband center frequency.

This remarkable result is the basis for many important considerations about thermal sensing in the THz region. First, because $h\nu$ is generally $\ll k_B T$ in terrestrial THz sensing, it implies that the signal-to-noise ratio before detection is always less than unity for a unimodal sensor, but can be greater than unity if $\Sigma \eta_m > 1$, as normally occurs in infrared detectors. Second, the modal dependence would appear, on first glance, to support an arbitrarily high signal-to-noise ratio given a suitable coupling structure. In other words, it is not clear whether or not there is a quantum-limit.

To elucidate the quantum limit for this case, we must calculate the NEP_{BD} but do so in light of the fact that both the signal and noise derive from the same incident photons. Because $P_{inc} = M h\nu f_p \Delta\nu$, one can write

$$\left(\frac{S}{N}\right)_{BD} = \frac{[\sum_m^M \eta_m] \cdot P_{inc}}{\sqrt{[\sum_m^M \eta_m] \cdot P_{inc} (P_{inc} + M \cdot h\nu_0 \cdot \Delta\nu)}}$$

Setting the $(S/N)_{BD}$ to unity and solving for P_{inc} , it is easy to show

$$NEP_{BD} = h\nu_0 \cdot \Delta\nu \cdot \frac{M}{\sum_m^M \eta_m - 1} \equiv NEP_{OL} \quad NET_{BD} \equiv \frac{h\nu_0 \Delta\nu}{k_B} \cdot \frac{M}{\sum_m^M \eta_m - 1}$$

where NET_{BD} is the noise equivalent temperature of a single-mode thermal source. For the unimodal case, this leads to a negative value - clearly not allowed on physical grounds and a result of the fact that S/N cannot equal 1.0. It diverges for $\Sigma \eta_m = 1$, and for $\Sigma \eta_m > 1$ it goes positive. In the special case that the sensor accepts a large number of modes, each having the same value of $\eta_m \equiv \eta$ such that $\Sigma \eta_m = M\eta \gg 1$, we find $NEP \rightarrow h\nu \Delta\nu / \eta$ - the same photon-noise limit as for a coherent signal.

The quantum-limited NEP after classical or quantum noise-free square-law detection follows from Eqn 43

$$NEP_{AD} = NEP_{BD} \cdot \sqrt{\frac{2\Delta f}{\Delta\nu}} = \frac{h\nu_0 M}{\sum_m^M \eta_m - 1} \sqrt{2\Delta\nu \cdot \Delta f}$$

and the specific NEP is simply

$$NEP'_{AD} = \frac{h\nu_0 M}{\sum_m^M \eta_m - 1} \sqrt{2\Delta\nu}$$

A.3. Arbitrary Signal, Thermal Background

A third important case of radiation-noise limited sensitivity occurs when the direct detector is coupled to many more spatial modes than those of the signal. Those beyond the signal modes are called “background” modes and can occur for a variety of reasons, such as imperfect coupling between the receiver antenna and the external coupling optics. If the radiation power from these modes dominates the radiation noise from the signal modes and can be associated with a brightness temperature T_B , then one can write from Eqn 44,

$$\left(\frac{S}{N}\right)_{BD} = \frac{P_{abs}}{\sqrt{(\Delta P_{abs})^2}} = \frac{\sum_m^{M_S} \eta_m \cdot P_m}{h\nu_0 \cdot \Delta\nu \cdot \sqrt{\left[\sum_n^{M_B} \eta_n\right] \cdot f_{P,B}(f_{P,B} + 1)}}$$

where $f_{P,B}$ is the background Planck function, M_S (M_B) is the number of signal (background) modes, and P_m is the power per signal mode, be it coherent or thermal (e.g., if thermal $P_m = h\nu_0 \Delta\nu f_P$). In general this expression is difficult to evaluate, but reduces to simplicity if all modes have the same coupling efficiency, η . Then one can write the “background-limited” expression

$$\left(\frac{S}{N}\right)_{BD} = \frac{\eta \cdot P_{inc}}{h\nu_0 \cdot \Delta\nu \cdot \sqrt{M_B \cdot \eta \cdot f_{P,B}(f_{P,B} + 1)}}$$

where P_{inc} is the sum of signal power over all signal modes. This leads to the NEPs

$$NEP_{BD} = \frac{h\nu_0 \Delta\nu \cdot \sqrt{M_B \cdot \eta \cdot f_P(f_P + 1)}}{\eta} = \sqrt{\frac{h\nu_0 \cdot \Delta\nu \cdot P_B(f_{P,B} + 1)}{\eta}} \equiv NEP_{back},$$

and

$$NEP_{AD} = \sqrt{\frac{2h\nu_0 \cdot \Delta f \cdot P_B(f_{P,B} + 1)}{\eta}}$$

In the Rayleigh-Jeans limit where $f_{P,B} \gg 1$, one obtains

$$NEP_{AD} \approx \sqrt{\frac{2k_B T_B \cdot \Delta f \cdot P_B}{\eta}}$$

and in the Wien limit where $f_{P,B} \ll 1$

$$NEP_{AD} \approx \sqrt{\frac{2h\nu_0 \cdot \Delta f \cdot P_B}{\eta}}$$

This is the familiar background-limited NEP_{AD} valid in the near-infrared or visible region,⁴⁴ which confirms the present formalism.

When the background thermal noise power is comparable to the signal noise power, both mechanisms must be considered. In calculating the NEP one can assume that the radiation noise mechanisms are uncorrelated, but must be careful to account for the correlation between the signal and its quantum noise. So it is wise to start with the SNR, which can be written in the following form if all modes have the same η ,

$$\left(\frac{S}{N}\right)_{BD} = \frac{\eta \cdot P_{inc}}{h\nu_0 \cdot \Delta\nu \cdot \sqrt{M_S \cdot \eta \cdot f_{P,S}(f_{P,S} + 1) + M_B \cdot \eta \cdot f_{P,B}(f_{P,B} + 1)}}$$

where $f_{P,S}$ ($f_{P,B}$) signal and background Planck functions. By comparison to the above background-limited expression, this becomes

$$\begin{aligned} \left(\frac{S}{N}\right)_{BD} &= \frac{\eta \cdot P_{inc}}{\sqrt{h\nu_0\Delta\nu \cdot \eta \cdot P_{inc}(f_{P,S} + 1) + h\nu_0\Delta\nu \cdot \eta \cdot P_B(f_{P,B} + 1)}} \\ &= \frac{P_{inc}}{\sqrt{h\nu_0\Delta\nu \cdot P_{inc}(f_{P,S} + 1)/\eta + NEP_{back}^2}} \end{aligned}$$

where the number of background modes is implicitly contained in P_B . The Planck function $f_{P,S}$ is correlated to P_{inc} through the source temperature, so a better form is

$$\left(\frac{S}{N}\right)_{BD} = \frac{P_{inc}}{\sqrt{P_{inc}(P_{inc} + M_S h\nu_0\Delta\nu)/(M_S\eta) + NEP_{back}^2}}$$

Setting the SNR to unity and P_{inc} to the NEP, one gets the quadratic equation:

$$NEP_{BD}^2 - NEP_{BD} \frac{M_S h\nu_0\Delta\nu}{M_S\eta - 1} - \frac{M_S\eta}{M_S\eta - 1} NEP_{back}^2 = 0$$

By recognizing the coefficient of the middle term as the quantum-limited NEP_{BD} when $M_S\eta > 1$, the only physically-allowable (positive) solution is:

$$NEP_{BD} = \frac{1}{2} \left[NEP_{QL} + \sqrt{NEP_{QL}^2 + 4NEP_{back}^2 \cdot \frac{M_S\eta}{M_S\eta - 1}} \right]$$

and $NEP_{AD} = NEP_{BD} (2\Delta f/\Delta\nu)^{1/2}$. When $NEP_{back} \gg NEP_{QL}$ and $M_S\eta \gg 1$, NEP_{BD} approaches NEP_{back} . And when $NEP_{back} \ll NEP_{QL}$ and $M_S\eta \gg 1$, it approaches NEP_{QL} . So a good approximation to the exact expression is obtained by the NEP summation formula from Sec. IV.C,

$$NEP_{AD} \approx \sqrt{NEP_{QL}^2 + NEP_{back}^2} = \sqrt{\left(\frac{h\nu_0 M_S}{M_S\eta - 1}\right)^2 2\Delta\nu \cdot \Delta f + \frac{2h\nu_0 \cdot \Delta f \cdot P_B(f_{P,B} + 1)}{\eta}}$$

which is always usefully accurate if $M_S\eta \gg 1$

For detection of thermal radiation, the most useful metric is $NE\Delta T$.⁴⁵ From the definition given in Sec. IV, we can compute this easily in the Rayleigh-Jeans limit using $P_{inc} \approx M_S k_B T_S \Delta\nu$ so that $dP_{inc}/dT = M_S k_B \Delta\nu$ and one can write from the approximate expression

$$NE\Delta T \approx \sqrt{\left(\frac{h\nu_0}{k_B(M_S\eta - 1)}\right)^2 \frac{2\Delta f}{\Delta\nu} + \frac{2T_B \cdot \Delta f \cdot P_B}{k_B\eta(M_S\Delta\nu)^2}} \quad (45)$$

It is interesting that there is no dependence on the source temperature.

An interesting application of Eqn 45 is to examine the wavelength dependence of $NE\Delta T$ for a multimode THz thermal sensor of fixed effective aperture. To do this in a meaningful way we need to account for the dependence of M on frequency, as discussed in Sec. III.D and written in the convenient form, $M = \text{Int}(\nu/\nu_R)^2 + 1$. If we assume that $M_S\eta \gg 1$ we find from (45)

$$NE\Delta T \approx \frac{1}{\text{Int}(\nu_0/\nu_R)^2 + 1} \sqrt{\left(\frac{h\nu_0}{k_B \cdot \eta}\right)^2 \frac{2\Delta f}{\Delta\nu} + \frac{2T_B \cdot \Delta f \cdot P_B}{k_B\eta(\Delta\nu)^2}}$$

This expression shows how important center frequency is to the sensitivity of a direct detector. From the quantum-limited NEP_{AD} expression one sees a linear dependence on

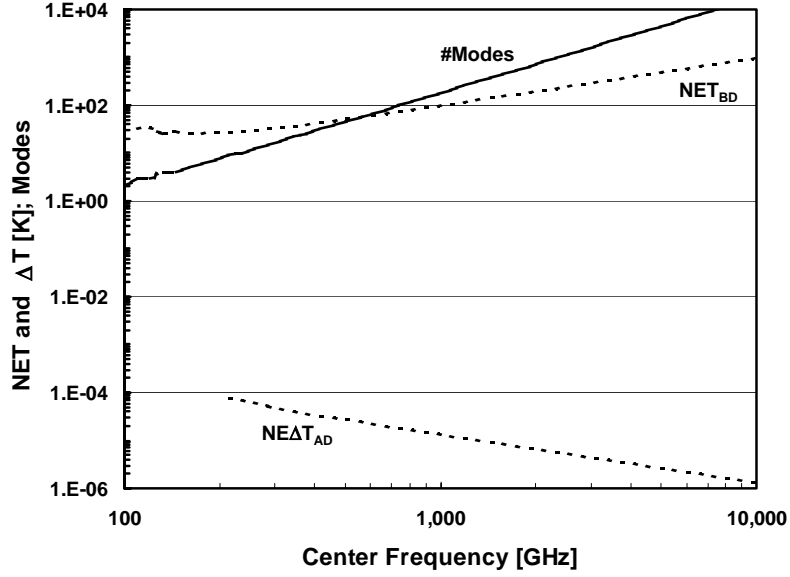


Fig. 14. Performance of a THz direct detector in the quantum limit assuming a cut-off frequency of 75 GHz and the definition $NET_{BD} = NEP_{BD}/k_B$. For $NEAT_{AD}$, an RF bandwidth of 100 GHz and a post-detection bandwidth of 30 Hz were assumed.

ν_0 and a square-root dependence on $\Delta\nu$ that appear to favor the mm-wave and THz regions over the infrared. But the NEP can be deceiving if the total signal power is also varying because of a change of mode number.

To exemplify this point, we plot in Fig. 14 the quantum-limited NEP_{BD} (expressed as a temperature NET_{BD}), the number of modes and the $NEAT_{AD}$ for a direct detector having $\nu_R = 75$ GHz. The quantum-limited curve rises linearly with center frequency as expected, but the mode density is quadratic. So the change of temperature required to match the noise decreases monotonically with frequency. This important result is sometimes misconstrued as arising from the behavior of the Planck factor instead of the mode density. And it tends to favor the infrared region in terms of ultimate sensitivity. But by the same token, the infrared direct detector often accepts so many modes that many of them couple to the background rather than the real target or the object to be imaged. This drives up the contribution from the background term in Eqn 45, leading to the common occurrence of a “background-limited infrared photodetector” or BLIP for short. In the mm-wave and THz region where the total number of modes is relatively small to begin with, the background contribution is generally easier to manage and, in good designs, make negligible.

B. Radiation-Noise-Limited NEP_{BD} of Classical and Quantum Coherent Receivers

In Secs V.C and V.D, identical expressions were derived for the IF SNR of classical and quantum coherent detection assuming that the dominant source of IF power fluctuation was radiation fluctuations in the local oscillator. Interestingly, the resulting expression, Eqn 39 depended on the input signal power but not on the LO power, so that by setting the SNR to unity, one gets for both classical and quantum coherent detection

$$NEP_{HET} = \frac{h\nu_{LO} \cdot B_{IF}}{\eta} \quad (46)$$

A similar derivation for the homodyne case would lead to

$$NEP_{HOM} = \frac{h\nu_{LO} \cdot B_{IF}}{2\eta} \quad (47)$$

These expressions are amongst the most famous in remote-sensing theory not just because of their role as a fundamental limit but because of their obvious consistency with the postulates of quantum-mechanics and classical probability theory. All sensors, even a radar operating at microwave frequencies, measure radiation by extracting photons via atomic transitions and stimulated absorption in some component of the receiver. This quantized process thus has a built-in granularity, or *uncertainty*, equal to a minimum measurable quanta, which is $h\nu/\eta$. Sensors having small η are decoupled from the environment and thus have a higher minimum quanta.

Probability theory teaches that the time rate of occurrence of an event or outcome of an experiment is just P·R, where P is the probability or “expectation value” of the event or outcome, and R is the rate of occurrence or “sampling” in the experiment. In Eqns 46 and 47, the “event” is just the stimulated absorption of a photon of energy U. From Eqn 4 the minimum expectation value is just ΔU (n=1) = $h\nu$. From sampling theory the occurrence rate is just the analog bandwidth of the measurement, B_{IF} . Now we arrive at the quantum-limited expression for the expectation-value rate, or power, associated with the measurement of radiation by any sensor at any frequency ν . Einstein is credited with the first understanding of the photoelectric effect, but apparently had little or no role in the quantum-limit of radiation measurement. Nevertheless, these expressions bear resemblance to some of his remarkable formulations – profound in concept while very simple in mathematics.

VII. Practical Limits on Receiver Sensitivity: Electronic Noise

In practice, real THz sensors are almost always limited in sensitivity by factors other than radiation noise, often by physical noise in the devices. In other cases the limit may be imposed by the atmospheric fluctuations, electromagnetic interference (EMI), power supply noise, and a myriad of other deleterious effects. Historically, it has been the environmental effects that have often made the difference between success and failure in THz deployment in the field.

A. Physical Noise of Electronics

Within every sensor system, particularly at the front end, are components that contribute noise to the detection process and therefore degrade the ultimate detectability of the signal. The majority of this noise usually comes from electronics, particularly the THz mixers or direct detectors themselves, and transistors in amplifiers that follow these devices. The majority of noise from such devices falls in two classes: (1) thermal noise, and (2) shot noise.⁴⁶ Thermal noise in semiconductors is caused by the inevitable fluctuations in voltage or current associated with the resistance in and around the active region of the device. This causes fluctuations in the voltage or current in the device by the same mechanism that causes σ - resistance, through Joule heating, couples energy to and from electromagnetic fields. The form of the thermal noise is very similar to that derived in Sec. II for blackbody radiation. And because the operational frequencies of electronics are generally well below the THz region, the Rayleigh-Jeans approximation is valid and the Johnson-Nyquist theorem applies. However, one must account for the fact that the device is coupled to a transmission line circuit, not to a free space mode, and the device may not be in equilibrium with the radiation as assumed by the blackbody model.

All of these issues are addressed by Nyquist's generalized theorem⁴⁷

$$\Delta V_{\text{rms}} = [4k_B T_D \text{Re}\{Z_D\} \Delta f]^{1/2} \quad (48)$$

where T_D , Z_D , and Δf are the temperature, differential impedance, and bandwidth of the device. Even this generalized form has limitations since it is not straightforward to define the temperature of the device if it is well away from thermal equilibrium. Also, the differential impedance of the device is not well defined for transistors or three-terminal devices in general. This topic goes well beyond the scope of this article, so suffice it to say that Eqn 48 is useful not so much for its accuracy but because of its form. For example, the thermal noise of transistors can often be written in a way that is identical to Eqn 48 with T_D being the temperature of carriers in the active region and Z_D being the transresistance or inverse transconductance of the device.

Shot noise is a ramification of the device being well out of equilibrium. It is generally described as fluctuations in the current arriving at the collector (or drain) of a three-terminal device caused by fluctuations in the emission time of these same carriers over or through a barrier at or near the emitter (or source) of the device. The mean-square current fluctuations are given by

$$\langle (\Delta i)^2 \rangle = 2e\Gamma I \cdot \Delta f$$

where Γ is a numerical factor for the degree to which the random Poissonian fluctuations of emission times is modified by the transport between the emitter (or source) and collector (or drain). If $\Gamma = 1$, the transport has no effect and the terminal current has the same rms fluctuations. When $\Gamma < 1$, the transport reduces the fluctuations, usually through some form of degenerative feedback mechanism, and the shot noise is said to be suppressed. When $\Gamma > 1$, the transport increases the fluctuations, usually through some form of regenerative feedback mechanism, and the shot noise is said to be enhanced.

B. Equivalent circuit representations

While at first appearing to add insurmountable complexity to the theory of sensor analysis, a great simplification results from the fact that radiation noise and two

forms of physical noise discussed above are, in general, statistically Gaussian. This is easy to understand in the case of thermal noise. But it remains true in the case of shot noise too provided that the number of electrons involved in the process is large. This follows from the central-limit theorem of probability theory. A very important fact is that any Gaussian noise passing through a linear component or network remains statistically Gaussian.⁴⁸ Hence, the output power spectrum S in terms of electrical variable X will be white and will satisfy the important identity

$$\langle (\Delta X)^2 \rangle = \int_{f_0}^{f_0+\Delta f} S_X(f) \cdot df \approx S_X(f) \cdot \Delta f$$

where Δf is the equivalent-noise bandwidth. Then one can do circuit and system analysis on noise added by that component at the output port by translating it back to the input port. In the language of linear system theory, the output and input ports are connected by the system transfer function $H_X(f)$, so the power spectrum referenced back to the input port becomes

$$S_X(in) = \frac{S_X(out)}{|H_X(f)|^2}$$

A related consequence of Gaussian noise is that it can be represented in circuit theory by an ac generator whose amplitude is the rms average of that variable and whose phase is perfectly random. This allows one to apply the tools of circuit theory.⁴⁹ Perhaps the simplest example is the ideal resistor. By the Johnson-Nyquist theorem, we know that the open-circuit voltage across a resistor fluctuates with rms value given by $(4k_B T \Delta f R)^{1/2}$. We can thus write a noise equivalent-circuit representation of this by adding in series with the resistor an ac voltage generator having amplitude $(4k_B T B R)^{1/2}$, a Gaussian-distributed amplitude, and a random phase. But one must remember that this Johnson Nyquist result is really based on the independent fluctuations of orthogonal photon modes, each characterized by frequency f . Hence, a more useful representation results from Fourier decomposing the total voltage generator into independent generators, each having amplitude $(4k_B T R/B)^{1/2}$, frequency f , and random phase. Then, because the generators at each f are statistically independent, we get their total voltage contribution by adding the sum of the squares. Because the noise is *white*, the total becomes

$$v_{rms}^2 = \int_{f_0-B/2}^{f_0+B/2} v^2(f) df = \left(\frac{4k_B T R}{B} \right)^{1/2} \cdot B = (4k_B T R)^{1/2}$$

as expected. The power of this decomposition comes in several techniques of signal processing, such as square-law detection of Gaussian noise, as discussed in Sec. V.

C. Electrical Noise Limitations on THz Square-Law Detector

To get a complete accounting of sensor signal and noise effects for a mm-wave or THz square-law detector, we need only combine all the noise generators with the signal processing model in to one equivalent circuit. In Secs. V and VI it was assumed that the radiation was noisy but the radiation detector was noise-free. To include the noise added by and after the detector, we assume it is AWGN and simply add a variance term ΔX^2 to account for the fluctuations in the signal (voltage or current). Hence, the mean signal is given by

$$X = \Re P_{\text{abs}} = \Re \left[\sum_m^M \eta_m P_{\text{inc},m} \right]$$

where the electrical responsivity is not mode-dependent because the detector responds the same to all absorbed power independent of how it is absorbed. The power signal-to-noise ratio after detection is given by

$$\left(\frac{S}{N} \right)_{AD} = \frac{(\Re \cdot \sum_m^M \eta_m P_{\text{inc},m})^2}{2\Re^2 \cdot (\Delta P_{\text{ABS}})^2 \cdot (\Delta f / \Delta \nu) + (\Delta X_n)^2}$$

A great simplification occurs when the total incident power, be it coherent or thermal, is divided equally over all modes. This is a good starting approximation for coupling thermal radiation into large coupling structures such as highly-overmoded feedhorns and integrating cavities. In this case $P_{\text{inc},m} = P_{\text{inc}}/M$ and for AWGN in the detector and electronics one finds

$$\left(\frac{S}{N} \right)_{AD} = \frac{(\Re \cdot \frac{P_{\text{inc}}}{M} \sum_m^M \eta_m)^2}{2\Re^2 \cdot (\Delta P_{\text{ABS}})^2 \cdot (\Delta f / \Delta \nu) + (\Delta X_n)^2} \quad (49)$$

In the common case that the detector and post-electronic noise contained in ΔX dominate the absorbed radiation noise, and assuming that none of this noise is correlated to the signal power, one can set $\text{SNR} = 1$ and solve for $P_{\text{inc}} (\equiv \text{NEP}_{AD})$ to find

$$\text{NEP}_{AD} = \frac{M \sqrt{(\Delta X_n)^2}}{\Re \cdot \sum_m^M \eta_m} \equiv \frac{M \cdot \text{NEP}_{\text{elect}}}{\sum_m^M \eta_m}$$

This allows one to write the SNR and specific NEP for the general case as

$$\left(\frac{S}{N} \right)_{AD} = \frac{(P_{\text{inc}})^2}{2 \cdot (\Delta P_{\text{ABS}})^2 \cdot (M / \sum_m^M \eta_m)^2 (\Delta f / \Delta \nu) + (M / \sum_m^M \eta_m)^2 \cdot \text{NEP}_{\text{elect}}^2}$$

As in Sec. VI.A one has to be careful in converting this to an NEP since the fluctuations in absorbed power will likely be correlated to the average signal power (e.g., radiation quantum noise). As in VI.A.C, an NEP summation formula is conveniently used as an approximation:

$$\text{NEP}_{AD} \approx \sqrt{(M / \sum_m^M \eta_m)^2 \cdot \{2 \cdot (\Delta P_{\text{ABS}})^2 \cdot (\Delta f / \Delta \nu) + \text{NEP}_{\text{elect}}^2\}}$$

When each mode has the same coupling to the detector, one has $\sum_m \eta_m = M \cdot \eta$ so that

$$\text{NEP}_{AD} = \frac{1}{\eta} \sqrt{2 \cdot (\Delta P_{\text{ABS}})^2 \cdot \Delta f / \Delta \nu + (\text{NEP}_{\text{elect}})^2}$$

Clearly, these are the same as the “unimodal” expressions one would obtain if only one spatial mode were incident and collected by the detector. These will be used later in conjunction with passive and active direct-detection sensors.

D. Electrical Noise Limitation on Heterodyne Mixer

Characterization of mm-wave and THz mixers can be done in the similar way as for square-law detectors. To account for the noise from the mixer itself and the IF electronics following, we simply add an rms term $[\Delta P^2]^{1/2}$ to the IF power:

$$\left(\frac{S}{N}\right)_{HET} = \frac{(\eta_R)^2 \cdot G_{mix} \cdot P_{inc}}{\eta_R \cdot G_{mix} \cdot h\nu_{LO} \cdot B_{IF} + \sqrt{(\Delta P_{IF})^2}}$$

where the radiation noise is all photon shot noise. Any mismatch between the antenna and the mixer is accounted for by η_R , and any mismatch between the mixer and the IF circuit is included in G_{mix} . By setting the S/N to unity and solving for P_{inc} , we find the specific NEP

$$NEP'_{HET} \equiv \frac{NEP_{HET}}{B_{IF}} = h\nu_{LO} / \eta_R + \sqrt{(\Delta P_{IF})^2} / [(\eta_R)^2 \cdot G_{mix} \cdot B_{IF}]$$

In the special case that the mixer and post-electronic noise dominate the absorbed radiation noise, and all of this noise is AWGN, the NEP reduces to the “mixer-limited” value

$$NEP'_{mixer} = \frac{\sqrt{(\Delta P_{IF})^2}}{(\eta_R)^2 G_{mix} \cdot B_{IF}}$$

Thus, we can write the overall heterodyne S/N ratio as

$$\left(\frac{S}{N}\right)_{HET} = \frac{(\eta_R)^2 \cdot G_{mix} \cdot P_{inc}}{\eta_R \cdot G_{mix} \cdot h\nu_{LO} \cdot B_{IF} + (\eta_R)^2 G_{mix} \cdot B_{IF} \cdot NEP'_{mixer}}$$

and

$$NEP'_{HET} = h\nu_{LO} / \eta_R + NEP'_{mixer}$$

Note that the units of this specific NEP is W/Hz, not W/(Hz)^{1/2} as for the square-law detector. This is because the mixer is behaving in a linear fashion with respect to the signal power.

In the mm-wave and THz regions, the NEP'_{HET} is usually measured with two loads of the same emissivity but different temperature T_L (e.g., ambient and 77 K) and satisfying $h\nu \ll k_B T_L$. Given this fact it is useful to define a new metric called the noise-equivalent temperature, or NET defined as that load which produces a $(S/N)_{HET}$ of unity. A short derivation yields

$$NET_{HET} = \frac{1}{k_B} (h\nu_{LO} / \eta_R + NEP'_{mixer})$$

In the millimeter-wave and THz literature, this NET is more commonly known as the receiver noise temperature, T_{REC} . More specifically, in the present case this is the double-sideband noise temperature because both signal bands are assumed to be down-converted. Because in the Rayleigh-Jeans limit the noise power per mode is linear in T_L , this NET is also equal to the $NE\Delta T$ that we would compute from the more common definition $NE\Delta T = NEP / [(dP_{inc}/dT)]$

For the purpose of system analysis and characterization, an even more useful metric is based on the fact that from Sec. VI.B the minimum NEP'_{HET} is just $h\nu_{LO}$ and the corresponding minimum NET is just $h\nu_{LO}/k_B$. This motivates a single parameter to represent the degradation in sensitivity caused by imperfect coupling, mixer noise, and IF electronic noise. This quantity is the “heterodyne coupling efficiency” η_{HET} , and is defined by

$$NEP'_{HET} = \frac{h\nu_{LO}}{\eta_{HET}}; NET_{HET} = \frac{h\nu_{LO}}{k_B \cdot \eta_{HET}}$$

VIII. Receiver Performance Limitations and Statistics

A. Optimum signal-to-noise ratio before detection: North's Theorem

THz sensors face many of the same challenges as faced by early microwave radars trying to detect small targets or large targets over a long range. Because of the weakness of the source or the high atmospheric attenuation, the received signal in THz sensors is usually very weak and difficult to distinguish from noise. As in the radar case, there is great interest in processing the weak signal in an optimum fashion, particularly in active sensors.

In one of the most elegant theorems in sensor theory, North showed how to construct the receiver transfer function or impulse response function to maximize the signal-to-noise ratio within a given resolution element. It is similar conceptually to Shannon's theorem in the sense that both state conditions of optimization: Shannon's stating the maximum information handling capacity on a noisy channel and North's stating the maximum signal-to-noise ratio on a noisy channel. North utilized the fact that independent of the sensor type, the maximum possible RF SNR is

$$SNR_{RF} = \frac{E_s}{N_0}$$

where E_s is the RF energy per sample and N_0 is the noise spectral density in the RF part of the receiver. When phrased for a pulsed system, E_s gets replaced by E_p , the energy per pulse. But North's theorem applies to cw systems as well by recognizing that these systems sample the RF power at a rate given by the sensor instantaneous bandwidth B , so the energy per pulse gets replaced by the energy per sampling time, or P_s/B , so that $SNR_{RF} = PS/(N_0B) = S/N$.

North's theorem showed that to achieve this SNR, the linear part of the RF sensor should have the following transfer function $H(f)$ and impulse response functions $h(t)$:

$$H(f) = A^*(f) \\ h(t) = a^*(t)$$

where $A(f)$ is the signal power spectrum, $a(t)$ is the signal waveform (current or voltage), and $*$ denotes complex conjugation. The first one makes sense intuitively since it simply states that all Fourier components of the signal spectrum should get “conjugately” coupled by the receiver. It is similar to the condition for optimum power transfer between a generator of impedance Z_G and a load of impedance Z_L , which from ac circuit theory can be proven to occur when $Z_G = Z_L^*$. The second condition is not so obvious, but becomes clearer in the special case of a sinusoidal signal since then complex

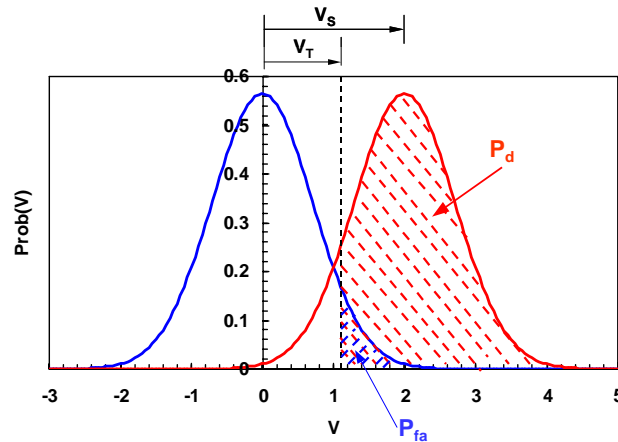


Fig. 15. Gaussian probability for the noise only and signal-plus-noise, applicable to a direct detection receiver.

conjugation is equivalent to time reversal. The combination of the two conditions is accomplished electronically with a “matched” filter – a pervasive device in radar and communications systems today at microwave frequencies and below.

B. Sensor Performance after Detection: Receiver Operating Characteristics

Given that the ultimate limits on the sensitivity any sensor depend on physical and environmental noise, the final performance can only be predicted using statistical means that properly account for each fluctuating effect. Physical noise and other fluctuations in each stage from the source of radiation to the output of the receiver detector must be ascribed a probability density function (PDF) for the relevant electrical variable at that point, and the overall PDF of the sensor in the presence of signal and noise must then be constructed.

Given the overall PDF, a common measure of the sensor performance comes from establishing a signal threshold at the output of the receiver detector.⁵⁰ The threshold approach, established early in the history of radar, is based on the observation that receivers are usually designed to “filter” the signal from the noise such that in the receiver baseband after detection the signal will have a greater amplitude than the noise. For example, if the baseband is dc, then the signal will correspond to the dc component and the noise will correspond to the residual ac power spectrum.

Given a well-define threshold, the *probability of detection* P_d , also called true positive detection, represents the probability that the sensor output PDF in the presence of the signal and all noise exceeds the threshold. The *probability of false alarm* P_{fa} , also called false positive detection or “false positive” for short, represents the probability that the PDF in the absence of the signal exceeds the threshold. The statistical assumption behind P_d is that a large number of such measurements are made so that P_d can be

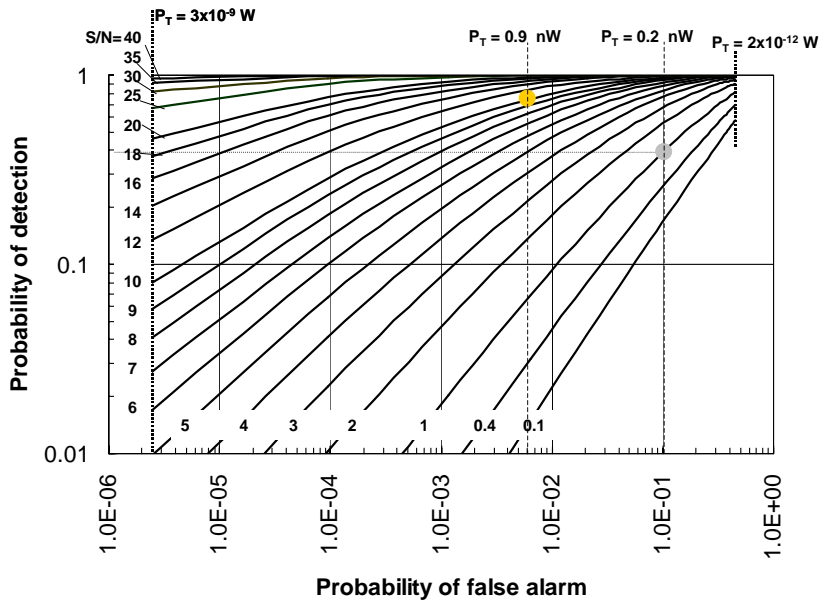


Fig. 16. Receiver-operating-characteristic diagram for direct-detection case.

thought of as either an average over time or an average over an ensemble of identical receivers processing the same signal but at all the possible noise states.

Plotted together, P_d vs P_{fa} form what is called the “receiver operating characteristic”, or ROC for short. Having its roots in radar detection, ROC diagrams are now universal and get applied to remote and point sensors alike in all parts of the electromagnetic spectrum and to other sensor domains, such as acoustic and chemical. At present their most common application appears to be in the biomedical industry where the difference between a true and false positive from a sensor often has serious human consequences.

For electromagnetic remote sensors there are two commonly used sensor PDF functions and resulting ROC diagrams that apply to the ideal direct-detection sensor and heterodyne (or homodyne) sensor, respectively.

B.1. Direct Detection Sensor

The analysis of Sec. V showed that the output noise spectrum after direct detection could be treated in a simple way if the rf bandwidth greatly exceeded the post-detection bandwidth, which is almost always true. In this case, the output consists of a dc component superimposed on white noise whose spectral density is related to the sensor PDF. In the ideal case, all of the noise is related to “physical” fluctuations of the radiation or in the electronics so has an AWGN character. The overall PDF for the noise alone is then Gaussian, meaning that the probability of measuring a particular value of X (voltage or current) at the at any point in time is given by

$$P(X) = \frac{1}{\sqrt{2\pi \cdot \overline{(\Delta X)^2}}} \exp[-(X)^2 / 2\overline{(\Delta X)^2}]$$

where $\overline{(\Delta X)^2}$ denotes the variance. The overall PDF in the presence of signal is simply

$$P(X, X_s) = \frac{1}{\sqrt{2\pi \cdot \overline{(\Delta X)^2}}} \exp[-(X - X_s)^2 / 2\overline{(\Delta X)^2}]$$

where X_s is the average value, or signal. Note that the signal strength amplitude X_s is always positive while X goes both positive and negative – consistent with the way that square-law detectors process signals.

These two functions are plotted in Fig. 15 where a threshold X_T is also drawn. Now the definitions of P_d and P_{fa} are defined graphically in terms of the cross-hatched or shaded regions below the two curves and above the threshold. Mathematically they are given by

$$P_d = 1 - \frac{1}{\sqrt{2\pi \overline{(\Delta X)^2}}} \int_{X_T - X_s}^{\infty} \exp[-X^2 / 2\overline{(\Delta X)^2}] \cdot dx$$

and

$$P_{fa} = \frac{1}{\sqrt{2\pi \overline{(\Delta X)^2}}} \int_{X_T}^{\infty} \exp[-X^2 / 2\overline{(\Delta X)^2}] \cdot dx$$

which can be written easily in terms of error functions. One can compute these integrals for various values of $(\text{SNR})_{AD} = (X_s)^2 / \overline{(\Delta X)^2}$ and X_T . One then arrives at the ROC plot of P_d vs P_{fa} shown in Fig. 16. Note that this plot is *universal* in that it does not depend on the specific type of direct detector or overall performance. The only assumptions are that the noise is entirely AWGN in nature, and that the receiver architecture is based on direct detection.

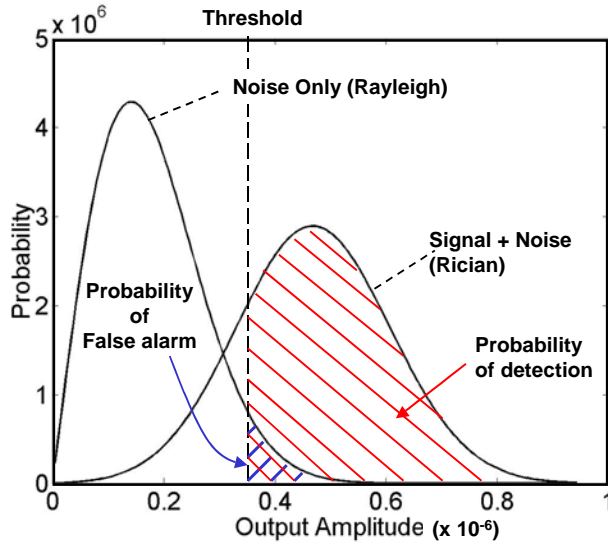


Fig. 17. Probability distribution functions for noise only and signal-plus-noise in coherent receiver.

The universal ROC curve clearly shows how the combination of SNR and X_T impacts Pd and Pfa in a trade-off fashion. It also shows that an SNR of unity is rather impractical, reaching a Pd of 0.9 only when the Pfa is greater than 0.1. This is in spite of the engineering convention of using an SNR = 1 as a criterion of good detection in sensors.

B.2. Heterodyne (or Homodyne) Sensor

The heterodyne (or homodyne) sensor is more complicated because mixing AWGN with a local oscillator will have a different result after square law detection than AWGN alone. The PDF now is the combination of two Gaussian functions, one representing the fraction of the receiver noise in phase with the LO before detection, and one representing the fraction of the receiver noise in quadrature (i.e., 90° displaced) from the LO. The resulting post-detection PDF for the noise alone, decomposed into these two components, is the classic Rayleigh distribution

$$P(X) = \frac{X}{(\Delta X)^2} \exp[-(X^2)/2(\Delta X)^2] \cdot$$

The post-detection PDF for the signal plus noise is given by the famous Rician distribution,⁵¹

$$P(X, X_s) = \frac{X}{(\Delta X)^2} \exp[-(X^2 + X_s^2)/2(\Delta X)^2] \cdot I_0[X \cdot X_s / (\Delta X)^2]$$

where I_0 is the modified Bessel function of order zero. These two functions are shown in Fig. 17 where the Pd and Pfa have similar graphical interpretations as given for the direct-detection case.

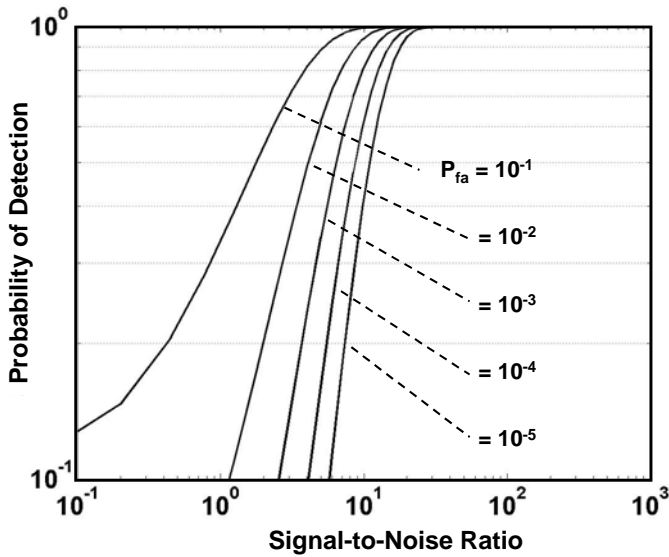


Fig. 18. Receiver operation characteristic (ROC) diagram for coherent case.

Because the Rayleigh distribution is integrable, one can easily derive an analytic form for the Pfa as

$$P_{fa} = \int_{X_T}^{\infty} P(X) dX = \exp[-(X_T^2)/2 \cdot \overline{(\Delta X)^2}]$$

but the Pd is more complicated and is best done numerically,

$$P_d = \int_{X_T}^{\infty} P(X, X_s) dX$$

One form of the numerical results are plotted in Fig. 18 where Pd is plotted SNR with Pfa as a parameter. This form eliminates the explicit dependence on a threshold. The resulting curves show that, as in the case of direct detection, an SNR of unity is not very useful, achieving values of Pd > 0.1 only when the Pfa is greater than about 0.01. The key advantage heterodyne (or homodyne) over direct detection, as will be shown below, is the much greater SNR that can be achieved for a given incident signal power. This is the same reason that heterodyne receivers are still the standard today in radio communications.

IX. Overall Performance of Four Types of Passive Sensors

In a passive sensor the “signal” is generally the time-averaged thermal radiation in some spectral bandwidth that propagates between an object of interest and the sensor. The advantage is clearly that the source radiation comes for free. The downside is that the atmospheric attenuation and fading effects can quickly degrade the sensor performance as will be shown below. It is very important to construct a realistic

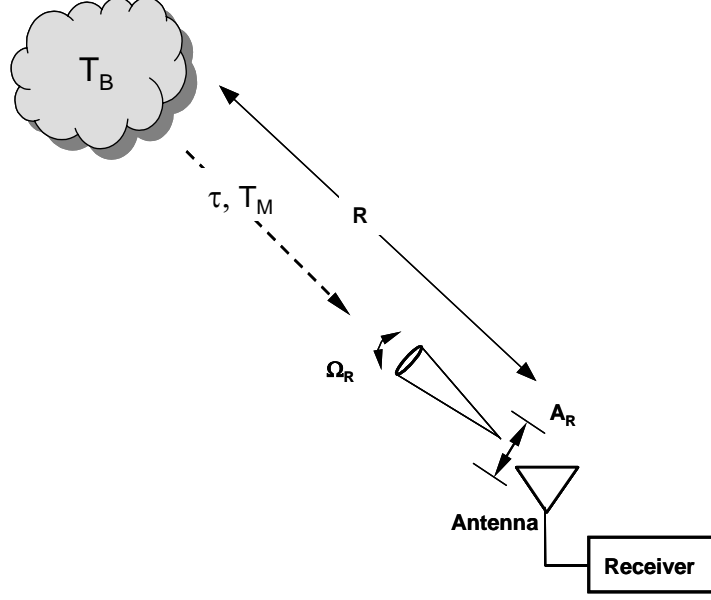


Fig. 19. Block diagram of canonical receiver with receiving-antenna effective aperture A_R , solid angle Ω_R , and separation from target R .

scenario of the object and the intervening atmosphere to properly predict the remote sensing performance.

A. Sensor Scenario

Perhaps the most common passive remote-sensing scenario in the mm-wave and THz regions is the thermal signal with a thermal background, as shown in Fig. 19. Between the target (brightness temperature T_B) and the sensor detector there is assumed to be an attenuating atmosphere whose physical temperature is T_M . In addition, telescopes or other optical components placed between the target and detector add to the attenuation, creating an overall transmission factor τ_A which can be significantly less than unity. For simplicity, the object is assumed to fill the field-of-view of the passive sensor so that each spatial mode of the receiver is subtended by the target. We assume further that the temperature of the target, the atmosphere, and the background are all sufficiently high that $h\nu \ll k_B T$. Hence the following expression can be used from Sec. II for the thermal power incident in each of the M spatial modes

$$P_{inc,m} = k_B \cdot \Delta\nu \cdot [\tau_A \cdot T_B + (1 - \tau_A) \cdot T_M] \cdot$$

where only the term dependent on T_B is actual signal. From the analysis of Sec. II.D, the fluctuations in this thermal radiation absorbed by the detector can be estimated by

$$\overline{(\Delta P_{abs})^2} = \left[\sum_m^M \eta_m \right] (h\nu_0 \Delta\nu)^2 f_P (f_P + 1) \approx M \cdot \eta \cdot (k_B T_{eff} \cdot \Delta\nu)^2$$

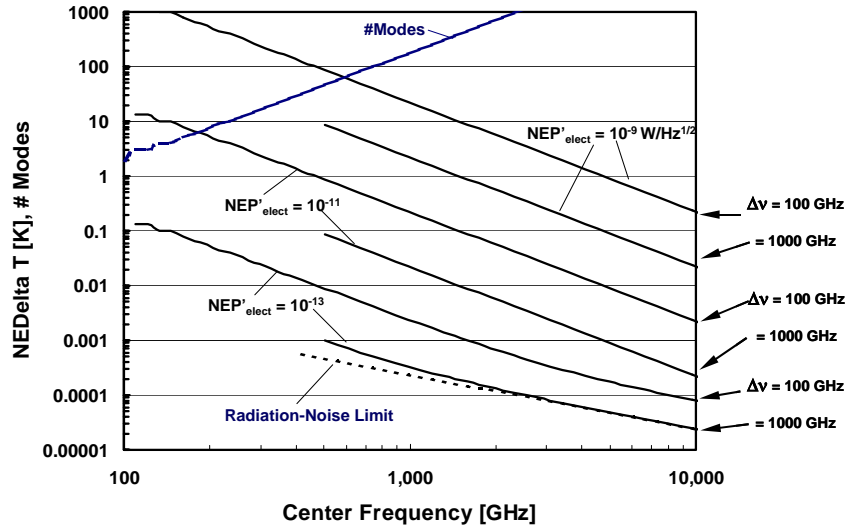


Fig. 20. Curves of NE Δ T vs ν_0 parameterized by various values of the NEP $_{\text{elect}}$ and $\Delta\nu$. Also shown is the number of spatial modes assuming $\nu_R = 75$ GHz. Other important parameters are: $\Delta f = 30$ Hz (consistent with live-video requirements), $\tau_A = 1.0$, and $\eta = 0.5$ (all modes), and $T_S = T_M = 290$ K.

where the effective temperature is defined by $T_{\text{eff}} = \tau_A T_B + (1 - \tau_A) T_M$. This expression will include the quantum-noise limit as a special case. The following sub-sections will evaluate four passive sensors with this thermal signal and radiation noise as the basis.

B. Direct Detection with Classical Square-Law Detector

B.1. General Analysis

In Sec. VII.C an approximate general expression was derived for the NEP of a mm-wave or THz direct detector in terms of absorbed radiation fluctuations and the electrical NEP. For the present passive scenario in the Rayleigh-Jeans limit and assuming all spatial modes have the same value of η , this expression is given by

$$NEP_{AD} \approx \frac{1}{\eta} \sqrt{2 \cdot (\Delta P_{ABS})^2 \cdot \Delta f / \Delta \nu + (NEP_{\text{elect}})^2}$$

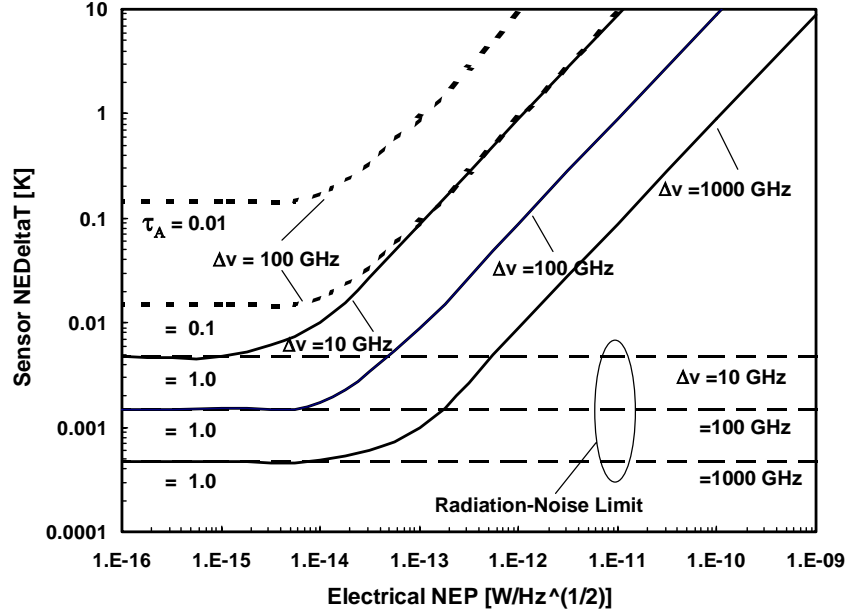


Fig. 21. Curves of NE ΔT vs NEP $_{elect}$ parameterized by various values of Δv and τ_A . Other important parameters are: $\nu_0 = 500$ GHz, $\Delta f = 30$ Hz (consistent with live-video requirements), $\eta = 0.5$ (all modes), $M = 45$, and $T_S = T_M = 290$ K.

$$= \sqrt{2 \cdot (k_B T_{eff})^2 \cdot M \cdot \Delta f \cdot \Delta \nu / \eta + (NEP_{elect} / \eta)^2} \quad (50)$$

The appropriate sensitivity metric in this case is the NE ΔT defined as in Sec. IV but now in terms of just the source brightness temperature. From the expression above we have $dP_{inc}/dT_B = M \cdot k_B \cdot \tau_A \cdot \Delta \nu$, so the NE ΔT is given by

$$NE\Delta T \equiv \frac{NEP_{AD}}{dP_{inc}/dT_B} = \tau_A^{-1} \sqrt{2 \cdot T_{eff}^2 \Delta f / (M \cdot \eta \cdot \Delta \nu) + [NEP_{elect} / (\eta \cdot M \cdot k_B \cdot \Delta \nu)]^2} \quad (51)$$

In the event that the first term in the radicand dominates, we obtain

$$NE\Delta T = \frac{T_{eff}}{\tau_A} \sqrt{2 \cdot \Delta f \cdot (M \cdot \eta \cdot \Delta \nu)^{-1}}$$

The condition on receiver sensitivity for reaching this limit is simply that the electrical NEP satisfies

$$NEP_{elect} < k_B T_{eff} \sqrt{2 \cdot \eta \cdot M \cdot \Delta f \cdot \Delta \nu}$$

Eqn 51 becomes useful in evaluating the performance of direct-detection sensors once the parameters are known. From Sec. III the number of spatial modes can be estimated using $M \approx \Omega_{FOV}/\Omega_B \approx \text{Int}(\nu_0/\nu_R)^2 + 1$ where ν_R is a reference frequency, assumed here to be 75 GHz. The external coupling efficiency per mode is assumed to be

$\eta = 0.5$. The post-detection bandwidth is assumed to be 30 Hz consistent with live-video requirements. The center frequency ν , instantaneous bandwidth $\Delta\nu$, and atmospheric transmission are best left as parameters since these are the most flexible or variable parameters from an engineering or phenomenology standpoint. Fig. 20 shows the NE Δ T vs ν_0 for three different NEP $'_{\text{elect}}$ values and two different $\Delta\nu$ values. Also shown is the number of modes vs ν_0 for $\nu_R = 75$ GHz. In all cases NE Δ T drops monotonically with ν_0 simply because M increases quadratically – there is no other ν_0 dependence in Eqn 51. An increase of $\Delta\nu$ decreases NE Δ T at any given ν_0 , but a decrease in NEP $'_{\text{elect}}$ is effective only down to roughly 10^{-13} W/Hz $^{1/2}$. This is because of the radiation noise arising from the 290-K source and atmosphere. The reader may note that the literature is replete with cryogenic mm-wave and THz bolometers having far lower values of NEP, consistent with operation in upper-atmospheric and astronomical sensing where the background radiation noise is far lower than in the present scenario.

Fig. 20 also reiterates the point made in Sec. VI.A.3 of how important center frequency is to sensitivity. A detector having NEP $'_{\text{elect}} = 10^{-9}$ W/Hz $^{1/2}$ is relatively easy to obtain for room-temperature operation, but can provide an NE Δ T only between roughly 1 and 20 K at a 1000-GHz center frequency – not considered sensitive enough for most applications. But this same detector can provide NE Δ T between about 0.01 and 0.2 K if operated at a center frequency of 10 THz ($\lambda = 30$ μm). The explanation is, again, the increase in the number of spatial modes, which also explains why “uncooled” detectors can be used as the basis for sensitive cameras at roughly $\lambda = 10$ μm and shorter wavelengths. But the reader should note that the present analysis assumes all spatial modes contribute both signal and noise, which is increasingly difficult to achieve when the frequency goes well into the infrared.

A second interesting plot from Eqn 51 is NE Δ T vs NEP $'_{\text{elect}}$, parametrized by $\Delta\nu$ and τ_A . The center frequency, number of modes, and mode coupling are fixed at 500 GHz, $M=45$, and $\eta = 0.5$, respectively. The resulting curves in Fig. 21 show, again, the asymptotic approach toward the radiation-noise-limited NE Δ T for NEP $'_{\text{elect}} \leq 10^{-13}$ W/Hz $^{1/2}$. To get a useful sensitivity of, say NE Δ T ~ 0.1 K with $\tau_A = 1.0$, the NEP $'_{\text{elect}}$ can be no worse than about 10^{-11} W/Hz $^{1/2}$ – a performance that is hard to achieve at room temperature if one also requires $M = 45$ and $\eta = 0.5$. As τ_A degrades, the required NEP $'_{\text{elect}}$ also drops until $\tau_A \sim 0.01$ (-20 dB), when the radiation noise then limits the NE Δ T to values above 0.1 K. Hence for passive thermal imaging it is imperative in the THz region to choose the center frequency carefully as a trade-off between NEP $'_{\text{elect}}$, number of modes, and atmospheric transmission. From the atmospheric transmission curve of Fig. 5(a), a compelling choice of center frequency would be the highest-frequency “good window” centered around 9 cm^{-1} (270 GHz).

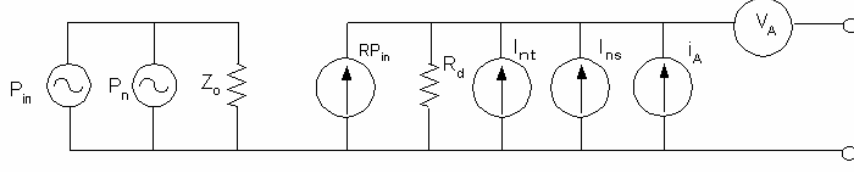


Fig. 22. Equivalent circuit of mm-wave/THz direct detector.

B.2. Detector-Specific Analysis

In many types of THz direct detectors, the more specific model of Fig. 22 is applicable which allows one to make first-principles estimates of the NEP and other metrics. The input RF impedance is represented by Z_{in} , and the rectification process is represented by power-dependent-current source of rms amplitude, i_p , which is related to the average power absorbed in Z_{in} through the responsivity \mathfrak{R} and the coupling efficiency η_m for each incident spatial mode

$$i_p = \mathfrak{R} P_{abs} = \mathfrak{R} \left[\sum_m^M \eta_m P_{inc,m} \right]$$

The thermal and shot-noise generators represent the noise contributed by the detector itself. In a more general analysis, there would be a third generator representing cross-correlation between these two. But in many THz detectors, either the thermal or shot-noise generator dominates the other so the cross-correlation becomes insignificant. The noise contributed by all of the remaining electronics after the detector is lumped into one thermal generator, T_n . This term is generally dominated by the first amplifier that occurs in the post-detection circuit. Note that this circuit model handles the RF noise as well as the signal entering the device. Hence, given the thermal model for fluctuations of the radiation absorbed by the device and assuming, again, that $P_{inc,m} = P_{inc}/M$, we can write

$$\left(\frac{S}{N} \right)_{AD} = \frac{\left(\mathfrak{R} \cdot \frac{P_{inc}}{M} \sum_m^M \eta_m \right)^2}{2\mathfrak{R}^2 \cdot \left(\sum_m^M \eta_m \right) \cdot (k_B T_{eff})^2 (\Delta\nu \cdot \Delta f) + \overline{(\Delta i_n)^2} + \overline{(\Delta v_n)^2} \cdot (G_D)^2 + 4k_B T_0 \cdot \Delta f / R_A},$$

so that the NEP_{AD} can be written

$$NEP_{AD} = M \cdot \sqrt{\frac{2 \cdot (k_B T)^2 \cdot \Delta\nu \cdot \Delta f + \overline{(\Delta i_n)^2} + \overline{(\Delta v_n)^2} \cdot (G_D)^2 + 4k_B \cdot T_0 \cdot \Delta f / R_A}{\sum_m^M \eta_m \cdot \left(\mathfrak{R} \cdot \sum_m^M \eta_m \right)^2}} \quad (52)$$

Under the common condition that the incident noise is dominated by the electrical noise terms, we get the electrically-limited NEP:

$$NEP_{AD}^{elect} = \frac{M}{\Re \sum_m \eta_m} \sqrt{(\Delta i_n)^2 + (\Delta v_n)^2 \cdot (G_D)^2 + 4k_B T_0 \cdot \Delta f / R_A} \quad (53)$$

These expressions are particularly useful when the individual noise mechanisms for a direct detector are already understood.

When the noise mechanisms are not understood but the NEP of a direct detector is already known or are readily measured experimentally, the following forms are more useful and easily derived from Eqns 52 and 53 above:

$$NEP_{AD} = \sqrt{\frac{2 \cdot M^2 \cdot (k_B T)^2 \cdot \Delta v \cdot \Delta f}{\sum_m \eta_m} + (NEP_{AD}^{elect})^2}$$

B.3. An Illustrative Example: Schottky-Diode Direct Detector

The Schottky diode is one of the oldest and best understood of the THz detectors. Detection and mixing depend on the nonlinear rectification properties of a metal-semiconductor junction – a phenomenon that has been exploited for over half a century. The intrinsic sensitivity and frequency conversion properties of the diode are well described by classical rectifier theory.² One of the strengths of the Schottky diode is that it can be made to have useful sensitivity over an enormous wavelength range extending from microwaves to about 100 μm . In addition, it is capable of very large instantaneous bandwidth (limited only by the IF circuit parameters), displays excellent performance characteristics at room temperature and can be fabricated in a variety of geometries to suit various applications. The three most prevalent geometries are the whisker-contacted honeycomb, the beam-lead diode and the planar or surface-oriented diode. Each of these will be discussed later in terms of their applications as submillimeter-wave mixers.

Independent of the particular geometry, the I-V characteristic of a Schottky diode is fit quite well by the expression⁵²

$$I = I_s \left[\exp\left[\frac{eV}{\eta k T_j}\right] - \exp\left[\frac{(\eta^{-1} - 1) \cdot V}{k T_j}\right] \right] \quad (54)$$

where η is the ideality factor (≥ 1.0), T_0 is the junction temperature, and $I_s = RT_j^2 \exp(-e\phi_{bi}/k_B T_j)A$ where R is the Richardson coefficient ($= 4.4 \text{ A/cm}^2/\text{K}^2$) and ϕ_{bi} is the built in potential. Typically in 100-GHz to 1 THz region the ideality factor of Schottky diodes is 1.2 to 1.3.⁵³ The first exponential term represents the normal thermionic emission current while the second term is mainly due to two tunneling components, field emission and thermionic field emission. The relative importance of each of these terms depends on both the temperature and the impurity concentration in the epitaxial layer. At room temperature and relatively low impurity concentration, the thermionic component usually dominates. A satisfactory fit to the I-V curve is then obtained by the more familiar expression $I = I_s \left[\exp\left\{\frac{eV}{\eta k T}\right\} - 1 \right]$.

According to the seminal theory of Torrey and Whitmer,⁵⁴ the low-frequency short-circuit voltage responsivity can be estimated directly from the I-V curve by the expression

$$S \approx \frac{1}{2} \frac{d^2 I / dV^2}{dI / dV} \quad (55)$$

The dependence on the second derivative generally leads to maximum sensitivity for forward bias at the “knee” of the I-V curve. The short-circuit physical noise after detection can be approximated by the three terms: (1) Schottky-diode “full” shot-noise $(\Delta i)^2 = 2eI_D \Delta f$, (2) the Schottky-diode thermal noise term that goes as $(\Delta i)^2 = 4k_B T_J G \Delta f$, and electronic noise all lumped into the current generator $(\Delta i)^2 = (4k_B T_0 / R_A) \Delta f$. Note that the Schottky thermal noise is intended to account for both the series resistance and differential resistance of the device. We also assume that the radiation noise power absorbed by the detector is negligible, so that the specific NEP can be approximated as,

$$NEP'_{AD} = \frac{1}{\Re} \sqrt{2eI_D + 4k_B T_J \cdot G_D + 4k_B T_0 / R_A} \quad (56)$$

It is illuminating to compute Eqns 54, 55, and 56 for a typical THz Schottky diode made of GaAs and coupled at 100 GHz to an antenna having a THz impedance of 100 Ω . We assume the diode has an area of 2 square micron and an ideality factor of 1.2. We assume the output amplifier has a noise voltage of 1 nV/Hz^{1/2} and a noise current of 1.0 pA/Hz^{1/2}. The resulting I-V curves at 77 K, 200 K, and 300 K are shown in Fig. 23(a). The short-circuit responsivity of Torrey&Whitmer is shown in Fig. 23(b), including mismatch with the antenna. The corresponding curves of NEP vs bias voltage are shown in Fig. 23(c). These curves demonstrate an important reality about direct detection: it is difficult to make the detector noise dominate the electronic noise that follows it, and it is even more difficult to make the detector noise dominate the radiation noise.

The limitations introduced by the electrical noise of direct detectors and the following electronics has been a problem since the early days of the THz field. Traditionally, this problem has been approached by the use of cryogenic detectors, particularly those operating at 4.2 K (vaporization point of liquid helium) and below. There is a wealth of literature on cryogenic direct detectors, particularly a myriad of different types of bolometers. It is not the purpose of this article to reiterate, but rather to examine the alternative receiver architectures for what may operate at room temperature. This leads us to the next three architectures, all of which have the potential to significantly out-perform the direct receiver under room temperature operation.

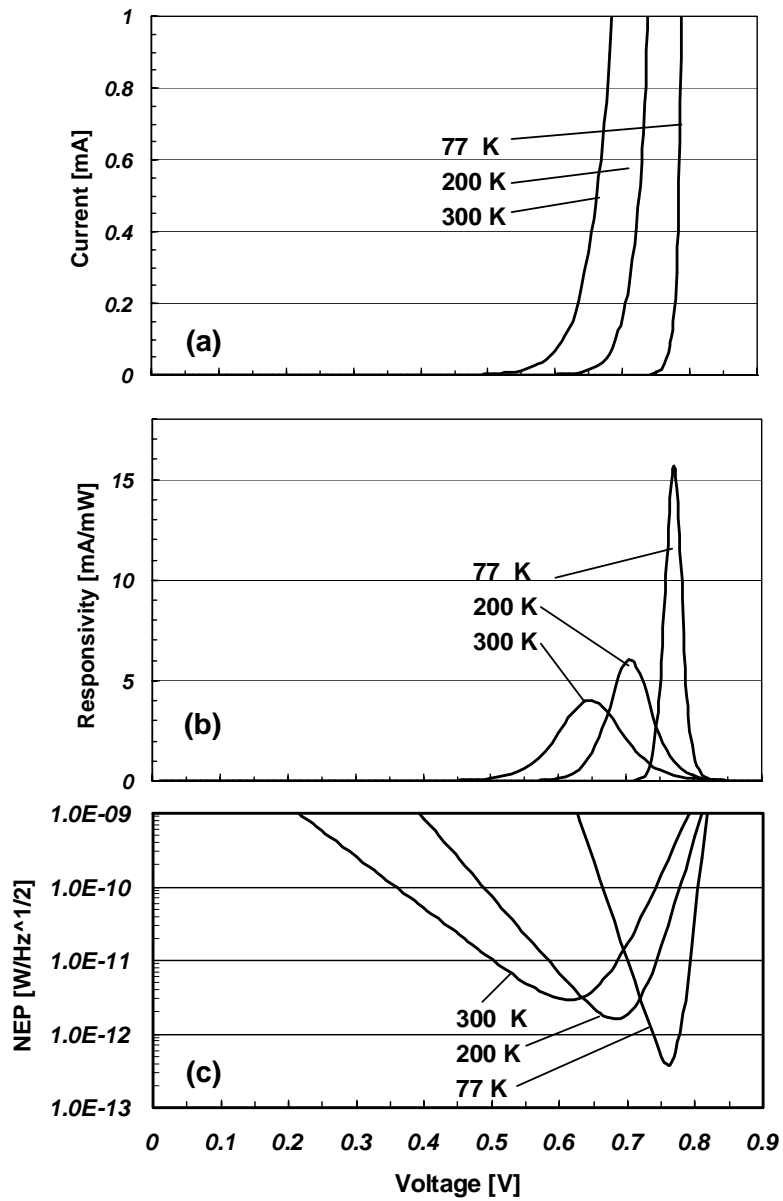


Fig. 23. Analytic results for a generic Schottky diode at 3 temperatures; (a) I-V curves, (2) electrical responsivity (including mismatch between the antenna and the diode, and (3) electrical NEP.

C. Heterodyne

In the interest of brevity, we analyze only the case of classical heterodyne detection as occurs routinely with Schottky-diode mixers, hot-electron bolometers, and other mixer types. From Sec. VII we have an expression for the SNR in the IF section immediately after down-conversion that includes the noise contribution from the mixer itself and the following IF electronics

$$\left(\frac{S}{N}\right)_{HET} = \frac{(\eta_R)^2 \cdot G_{mix} \cdot P_{inc}}{\eta_R \cdot G_{mix} \cdot h\nu_{LO} \cdot B_{IF}}$$

We multiply each term by the gain of the entire IF amplifier chain G_{IF} along with the responsivity of the square-law detector. To the LO “shot-noise” of the denominator we add the noise figure of the entire IF amplifier chain according to the Friis formula of Sec. III. The implicit assumption here is that, as for most heterodyne receivers, the IF gain is high enough that the noise contribution of the square-law detector and electronics following it are negligible. Because the IF band is generally at much lower frequency than the mm-wave or THz radiation, any losses in coupling to the square-law detector can be lumped in G_{IF} , and we obtain

$$\begin{aligned} \left(\frac{S}{N}\right)_{AD} &= \frac{[\mathfrak{R}_D \cdot G_{IF} \cdot (\eta_R)^2 \cdot G_{mix} \cdot P_{inc}]^2}{2\mathfrak{R}_D^2 \cdot \{G_{IF} \cdot \eta_R \cdot G_{mix} \cdot h\nu_{LO} \cdot B_{IF} + (\eta_R)^2 \cdot G_{mix} \cdot B_{IF} \cdot G_{IF} \cdot NEP'_{mixer}\}^2 \cdot (\Delta f / B_{IF})} \quad (57) \\ &= \frac{(\eta_R^2 \cdot P_{inc})^2}{2 \cdot \{\eta_R \cdot h\nu_{LO} + (\eta_R)^2 \cdot NEP'_{mixer}\}^2 \cdot (B_{IF} \cdot \Delta f)} \end{aligned}$$

This leads to the following useful expression for the NEP in active detection

$$NEP'_{AD} = \{h\nu_{LO} / \eta_R + NEP'_{mixer}\} \cdot (2B_{IF})^{1/2} = \frac{h\nu_{LO}}{\eta_{HET}} \cdot (2B_{IF})^{1/2}$$

Assuming once again that the incident signal is thermal noise contained in the same (single) mode as the local oscillator, we have $P_{inc} = k_B T_B \Delta f = k_B T_B B_{IF}$, so that in passive detection

$$NE\Delta T'_{AD} \equiv \frac{NEP'_{AD}}{dP_{inc} / dT} = \frac{h\nu_{LO}}{\eta_{HET} \cdot k_B} \sqrt{\frac{2}{B_{IF}}}$$

Fig. 24 shows curves of the NEP_{HET} before detection and the $NE\Delta T_{AD}$ after detection for various values of η_{HET} and B_{IF} .

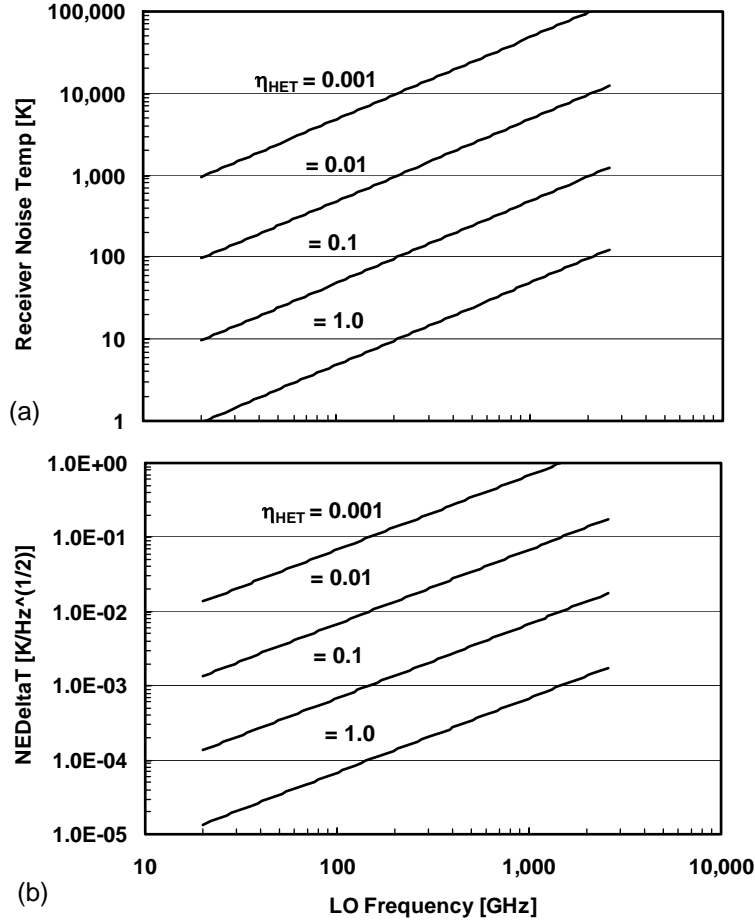


Fig. 24. Noise temperature and NEAT vs LO frequency of double-sideband heterodyne receiver having an IF bandwidth of 10 GHz.

Several of the popular detector types can operate in both as direct detector and mixers. A good example is the Schottky diode. Based on the above analysis, the following rule-of-thumb relationship between the heterodyne and direct NEP values can be found: $NEP'_{HET} \approx (NEP'_{AD})^2 / P_{LO}$. From the analysis of the generic Schottky diode in Sec. IX.B, we would have a room-temperature direct NEP'_{AD} of about $3 \times 10^{-12} \text{ W/Hz}^{1/2}$ at 300 GHz. Then given a local oscillator power of 1 mW, one would predict a heterodyne NEP of $9 \times 10^{-21} \text{ W/Hz}$, corresponding to a η_{HET} value of 0.02. These numbers are all reasonably close to the experimental reality.

D. Pre-Amplified Direct

We have seen that the electrical-noise limit in direct detection arises from the fact that the responsivity of the direct detector, at least the Schottky diode detector, is

limited. So intuitively, one would expect to be able to improve the sensitivity of a direct receiver through pre-amplification, provided that the gain of the pre-amplifier is suitably high and the noise figure is suitably low. To determine how good the amplifier has to be, we apply the noise formalism to the architecture of Fig. 13(a). To get the signal absorbed by the square law detector, we multiply the P_{inc} in each mode by the input radiation coupling coefficient $\eta_{R,m}$ (i.e., the fraction of incident power in each mode absorbed by the *amplifier*) and the amplifier available gain G_m for that mode

$$S_{BD} = \left(\sum_m^M G_m \cdot \eta_{R,m} \cdot P_{inc,m} \right)^2$$

Any mismatch between free space and the amplifier is lumped into $\eta_{R,m}$, and mismatch between the amplifier and the direct THz direct detector is lumped into G_m .

The mean-square power fluctuation absorbed by the amplifier depends, of course, on the type of radiation being detected. If the radiation being detected is thermal, the fluctuations can be approximated in the Rayleigh-Jeans limit by

$$\begin{aligned} \overline{(\Delta P_{abs})^2} &= \left(\sum_m^M \eta_{R,m} \cdot (k_B T_B \cdot \Delta \nu)^2 \cdot G_m^2 \right) + (P_{AMP})^2 \\ &= \sum_m^M \eta_{R,m} \cdot G_m^2 \cdot (k_B T_B \cdot \Delta \nu)^2 + [G_m \cdot (F_m - 1) \cdot k_B T_{300} \cdot \Delta \nu]^2 \end{aligned}$$

where P_{amp} is the rms noise power contributed by the amplifier, F_m is amplifier noise figure for mode m , and T_B is the background temperature. We can substitute this into the mean-square power fluctuation term in Eqn 49 to get the signal-to-noise ratio at the output of the square-law detector:

$$\left(\frac{S}{N} \right)_{AD} = \frac{(\Re \cdot \sum_m^M G_m \cdot \eta_{R,m} \cdot P_{inc,m})^2}{2\Re^2 \cdot \left\{ \sum_m^M \eta_m \cdot G_m^2 \cdot (k_B T_B \cdot \Delta \nu)^2 + [G_m \cdot (F_m - 1) \cdot k_B T_{300} \cdot \Delta \nu]^2 \right\} \cdot (\Delta f / \Delta \nu) + (\Delta X)^2}$$

We also assume that the signal fluctuations in the detector and the following electronics are Gaussian, so that detector electrical NEP scales with the post-detection bandwidth Δf and we can write

$$\frac{S}{N} = \frac{[\Re \cdot (\sum_m^M G_m \cdot \eta_{R,m} \cdot P_{inc,m})]^2}{2\Re^2 \cdot \left\{ \sum_m^M \eta_m \cdot G_m^2 \cdot (k_B T_B \cdot \Delta \nu)^2 + [G_m \cdot (F_m - 1) \cdot k_B T_{300} \cdot \Delta \nu]^2 \right\} \cdot (\Delta f / \Delta \nu) + (\Re \cdot NEP'_{elec})^2 \cdot \Delta f}$$

At the present point in time, solid-state amplifiers only operate in the sub-THz region and, like local oscillators, are designed only for one spatial mode. This simplifies the performance analysis to the “unimodal” expression,

$$\left(\frac{S}{N} \right)_{AD} = \frac{(\Re \cdot P_{inc} \cdot \eta_R \cdot G)^2}{2\Re^2 \cdot \left\{ \eta_R \cdot G^2 \cdot (k_B T_B \cdot \Delta \nu)^2 + [G \cdot (F - 1) \cdot k_B T_{300} \cdot \Delta \nu]^2 \right\} \cdot (\Delta f / \Delta \nu) + (\Re \cdot NEP'_{elec})^2 \cdot \Delta f}$$

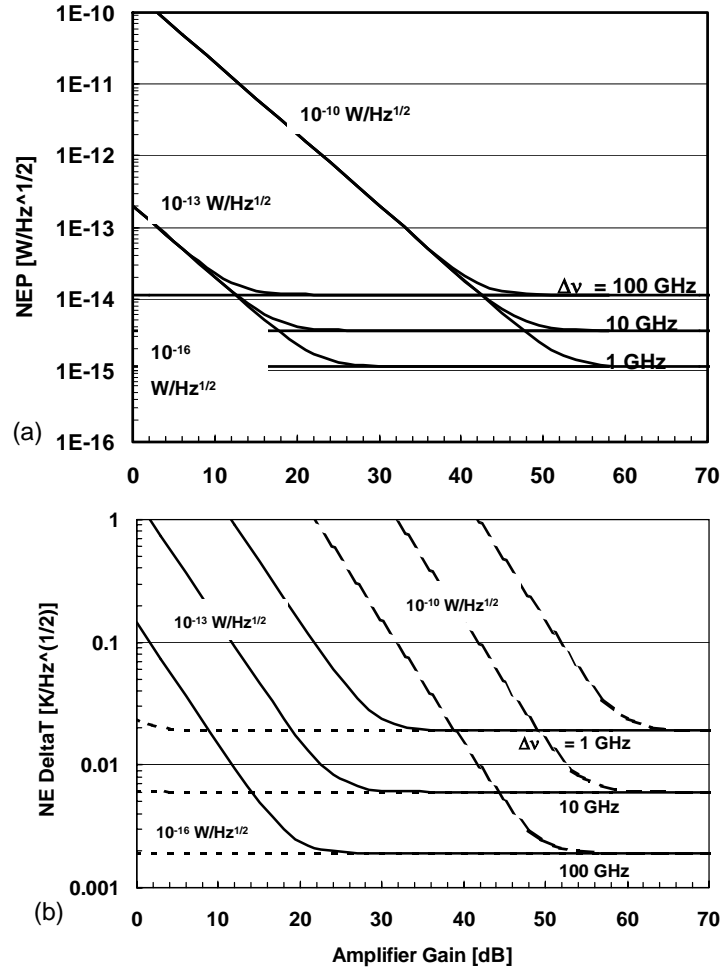


Fig. 25. (a) NEP and (b) NE Δ T for pre-amplified direct detection as a function of amplifier gain and parametrized by direct detector NEP'_{elec} and RF bandwidth $\Delta\nu$. The amplifier noise figure is fixed at 6.0 dB.

Thus the NEP is given as follows:

$$NEP_{AD} = \sqrt{2 \cdot \{\eta_R^{-1} \cdot (k_B T_B)^2 + [\eta_R^{-1} \cdot (F-1) \cdot k_B T_{300}]^2\} \cdot (\Delta\nu \cdot \Delta f) + (NEP'_{elec} / (\eta_R \cdot G))^2 \cdot \Delta f}$$

or

$$NEP'_{AD} = \sqrt{2 \cdot \{\eta_R^{-1} \cdot (k_B T_B)^2 + [\eta_R^{-1} \cdot (F-1) \cdot k_B T_{300}]^2\} \cdot (\Delta\nu) + (NEP'_{elec} / (\eta_R \cdot G))^2}$$

The unimodal assumption and the Rayleigh-Jeans approximation also facilitate the computation of the NE ΔT using the fact that the incident power per mode is $\bar{P}_{inc} \approx k_B T \cdot \Delta \nu$ and $d\bar{P}_{inc} / dT \approx k_B \cdot \Delta \nu$. We find

$$NE\Delta T \equiv \frac{P_{inc}}{dP_{inc} / dT} = \sqrt{2 \cdot \{\eta_R^{-1} \cdot (T_B)^2 + [\eta_R^{-1} \cdot (F-1) \cdot T_{300}]^2\} / (\Delta \nu) + (NEP'_{elec} / (\eta_R \cdot G \cdot k_B \Delta \nu))^2}$$

Fig. 25 shows plots of NEP and NE ΔT for the pre-amplified direct receiver under the following conditions. Fig. 25(a) shows the NEP vs the amplifier gain parametrized by three values of RF bandwidth: 1 GHz, 10 GHz, and 100 GHz and three values of NEP for the direct detector 10^{-9} W/Hz $^{1/2}$, 10^{-11} W/Hz $^{1/2}$, 10^{-13} W/Hz $^{1/2}$. Fig. 25(a) clearly shows that the NEP curves saturate with increasing amplifier gain at levels consistent with the amplifier noise. The quantum-limited NEP under each condition is also shown according to the expression derived in Sec. VI for direct detection of a thermal signal,

E. Pre-amplified Heterodyne

An intriguing possibility in the millimeter-wave band and lower end of the THz range is a preamplifier feeding a mixer element. As in Sec. IX.D we will analyze this only for the unimodal case since that is, by far, the most effective way to design and construct amplifiers and local oscillators alike. To simplify the expression, we will also assume that the local-oscillator photon noise dominates the incoming radiation noise – an assumption that will obviously become more questionable as the amplifier gain increases

$$\left(\frac{S}{N}\right)_{AD} = \frac{[\Re_D \cdot G_{IF} \cdot \eta_R \cdot G_{mix} \cdot G_{RF} \cdot P_{inc}]^2}{2\Re_D^2 G_{IF}^2 G_{mix}^2 \{\eta_R \cdot (G_{RF} \cdot h\nu_{LO} \cdot B_{IF})^2 + [G_{RF} (F-1) \cdot k_B T_{300} \cdot B_{IF}]^2 + (B_{IF} \cdot NEP'_{mixer})^2\} \cdot (\Delta f / B_{IF})}$$

$$= \frac{[\eta_R \cdot G_{RF} \cdot P_{inc}]^2}{2 \cdot \{\eta_R \cdot (G_{RF} \cdot h\nu_{LO})^2 + [G_{RF} (F-1) \cdot k_B T_{300}]^2 + (NEP'_{mixer})^2\} \cdot (B_{IF} \cdot \Delta f)}$$

This leads to a performance in active sensing of

$$NEP'_{AD} = \frac{1}{\eta_R} \sqrt{\{\eta_R (h\nu_{LO})^2 + [(F-1) \cdot k_B T_{300}]^2 + (NEP'_{mixer} / G_{RF})^2\} \cdot (2 \cdot B_{IF})} \quad (58)$$

and a performance in unimodal passive sensing of

$$NE\Delta T = \frac{1}{\eta_R \cdot k_B} \sqrt{\{\eta_R (h\nu_{LO})^2 + [(F-1) \cdot k_B T_{300}]^2 + (NEP'_{mixer} / G_{RF})^2\} \cdot (2 / B_{IF})}$$

Eqn 58 has the following satisfying properties. First, if the gain of the preamplifier is high enough to overcome the mixer noise, the resulting NEP has the form

$$NEP'_{AD} = \frac{1}{\eta_R} \sqrt{\{\eta_R (h\nu_{LO})^2 + [(F-1) \cdot k_B T_{300}]^2\} \cdot (2 \cdot B_{IF})}$$

If the noise figure is also high enough that the amplifier noise dominates the photon shot noise

$$F > 1 + \frac{\eta_R (h\nu_{LO})}{k_B T_{300}} = 1 + \frac{\eta_R \nu_{LO}}{6.248 \cdot 10^{12}}$$

and we get the ‘‘amplifier-limited’’ expression

$$NEP'_{AD} = \frac{(F-1) \cdot k_B T_{300}}{\eta_R} \sqrt{2 \cdot B_{IF}}$$

Presently, all known solid-state room-temperature amplifiers operating in the millimeter or THz regions satisfy this condition. The plots of amplified-heterodyne receiver performance are shown in Fig. 26. For active sensing, the NEP in Fig. 26(a) clearly increases monotonically with increasing B_{IF} for any value of G or NEP_{mixer} . And each case shows a monotonically decreasing value of NEP' down to a minimum value defined by the amplifier noise limit given above. The NEP also drops inversely with receiver coupling efficiency.

In a passive sensor, the amplified-heterodyne receiver provides the performance shown in Fig. 26(b). The $NE\Delta T$ now decreases with increasing B_{IF} just as in the canonical heterodyne receiver. Again, the $NE\Delta T$ saturates at a minimum level corresponding to the amplifier limit. In comparison to the amplified-direct or canonical heterodyne, the amplified-heterodyne provides one advantage in terms of reaching the amplifier-limited performance with a relatively poor mixer performance. As shown in Fig. 26(a), the limit defined by a 6-dB-noise-figure amplifier can be achieved with a mixer having an $\eta_{\text{HET}}=0.01$ ($NEP = 6 \times 10^{-18}$ W/Hz) and an amplifier gain of just ~ 10 dB. But the receiver can still reap the benefits of a superheterodyne receiver in terms of frequency selectivity defined in the IF rather than the RF. At the present time, this seems particularly advantageous for all-room-temperature operation where LO power is either difficult to generate or difficult to distribute to the mixer. One example is imaging arrays where LNAs would be simpler to implement than local oscillators. The reason for this is that LNAs have become monolithic integrated circuits whereas local oscillators in the mm-wave and THz regions are still discrete components. Worse yet, the local oscillators often require low frequency synthesizers to multiply up harmonically and act as the reference for the fundamental oscillator at the LO frequency.

X. Issues and Performance of Active Sensors

By the definitions given in Sec. I, active sensors are those that provide their own illumination of separated targets and objects using a sub-system called the transmitter. The job of the transmitter is to provide enough coherent or incoherent radiation at the target to get a measurable reflection or transmission of this radiation at the position of the receiver. The source of the radiation has long been the bane of the mm-wave and THz regions, tending to become more difficult and weaker in power as the frequency increases above 30 GHz. Historically, the best sources for this radiation have been vacuum-tube devices, such as oscillators driving traveling wave tube amplifiers below about 100 GHz, and fundamental vacuum oscillators (e.g., backward wave oscillators) above 100 GHz. Such sources are notoriously expensive and cumbersome once the high-voltage power supply is factored in. Researchers have long sought after solid-state sources to provide the transmit power. Until recently, the power levels and limited

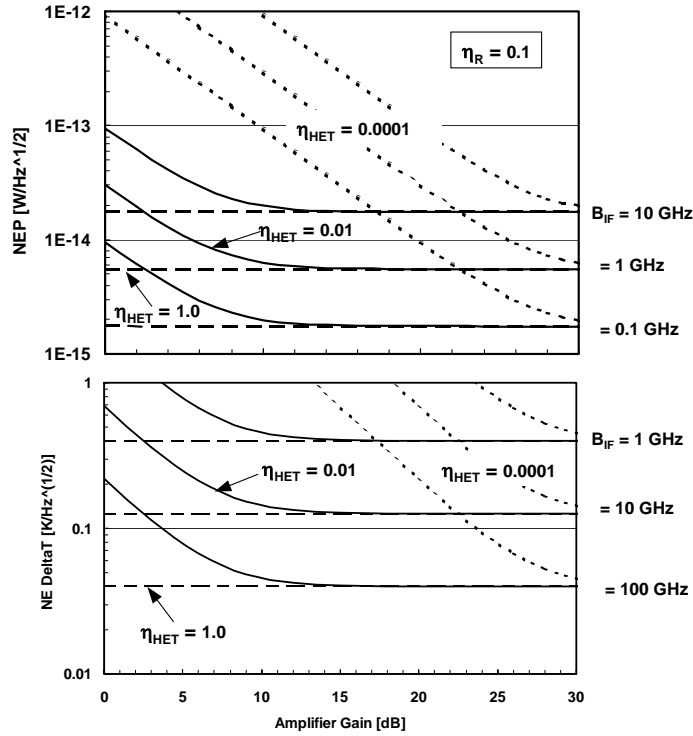


Fig. 26. (a) NEP and (b) NE Δ T for pre-amplified heterodyne receiver as a function of amplifier gain and parametrized by IF bandwidth and η_{HET} . The frequency is 100 GHz, the pre-amplifier noise figure is fixed at 6.0 dB, the quantum-limited NEP is $\approx 6.6 \times 10^{-21}$ W/Hz, and the external coupling efficiency is 0.1.

tunability have generally been sufficient to make only local oscillators, not free-space transmitters.

A. Active Sensor Scenario

By combining a transmitter (Tx) with a receiver (Rx) oriented as in Fig. 19, a coherent or quasi-coherent signal is sent out by the Tx, is reflected or transmitted by the object or target, and is then collected by the Rx where it is rectified from RF (THz) to baseband either by a direct or coherent receiver. The target might be a cloud or a partially transparent object, for example. For purpose of maximizing the received signal, the relative orientation of the Tx and Rx is very important.

If the object or target is best observed in transmission, then for good performance the Rx is best located along the direction of the transmitter radiation but on the opposite side of the object. An alternative (see Sec. X.D below) is to place a retroreflector on the opposite side of the object and co-locate the Rx next to the Tx. If the

object or target is best observed in reflection because of a strong back-scatter effect, then Tx is best configured to produce a quasi-collimated beam, similar to the “large-waist” beam plotted in Fig. 11.

A common goal of all three of these Tx-Rx orientations is that the received power, as defined in the Friis analysis of Sec. III.F, should fall with separation r between the Tx and Rx by $\sim 1/r^2$ or slower. This is in contrast to the typical $1/r^4$ decay that occurs in conventional radar, where the $1/r^2$ factor from the far-field Tx radiation is multiplied by a second $1/r^2$ factor based on assumed isotropic scattering from the target or object. Although the $1/r^4$ factor is considered accurate for long-range radar, particularly at microwave frequencies and below, it is unnecessarily pessimistic for the relatively-short range applications being pursued with mm-wave and THz sensors. Two good examples – concealed-weapons detection and all-weather navigation – are addressed later in Sec. XI.

To get the receiver SNR and ROC performance, one first needs to know how much of the transmitted power is collected by the receiver. For scatter-free atmospheric propagation, the instantaneous signal power at the receiver (Rx) can be estimated analytically given reasonable assumptions about the Tx power and beam patterns. For simplicity, we make the following three assumptions: (1) the Tx radiates only the one fundamental spatial mode of its antenna, (2) the width of the Tx and Rx beams are much smaller than the lateral extent of the target at every point where they pass through, and (3) the Tx and Rx antennas are perfectly aligned such that the peaks of their patterns are anti-collinear. The Rx incident power can then be approximated by the Friis formula (III.18) in the case of perfect beam and polarization alignment:

$$P_{inc}(\nu) = \frac{P_T G_T G_R \lambda^2}{(4\pi)^2 R^2} \tau(\nu) \quad (59)$$

This can be considered as a worst-case estimate if the Tx fundamental mode is a Gaussian beam having a Rayleigh length comparable or greater than R .

In the active sensor with direct- or pre-amplified direct detection, the Rx noise power will generally be the sum of at least two terms in addition to the electrical noise of the front-end electronics: (1) radiation fluctuations in the coherent received power and (2) radiation fluctuations of the background thermal radiation. In the Rayleigh-Jeans limit these become:

$$\overline{(\Delta P_{abs})^2} = \sum_m^M \eta_m \cdot \{(k_B T_{eff} \cdot \Delta \nu)^2 + (h\nu_0 P_{inc} \Delta \nu)\}$$

In the active sensor with heterodyne- or preamplified-heterodyne down-conversion, these must be added to the local-oscillator photon fluctuations, which is done below.

It is interesting to solve for the incident power that makes the photon noise dominate the thermal noise in the above expression. Under this condition one would expect to achieve a photon shot noise limit similar to that discussed in Sec. VI. Substitution from Eqn 59 yields the condition

$$P_T > \frac{(k_B T_{eff})^2 \cdot \Delta \nu (4\pi)^2 R^2}{h\nu_0 \cdot G_T G_R \lambda^2 \cdot \tau(\nu)}$$

We estimate this for a typical terrestrial case: $T_{eff} = 300$ K, $\Delta \nu = 1$ GHz, $R = 100$ m, $\nu_0 = 300$ GHz, $G_T = G_R = 1000$, $\lambda = 1$ mm, and $\tau = 0.1$. The result is $P_T > 1.4$ mW – an available power level, even from solid-state technology.

A facilitating consequence of the limited THz Tx power is that the coherent radiation term can usually be neglected in the above expression. As shown below, this greatly facilitates the sensitivity analyses and leads to the “background limited” case for active mm-wave and THz sensors.

B. Transmitter Types

B.1 Continuous Wave

The preferred source of cw radiation practically anywhere in the RF bands is that derived from solid-state electronics. The situation for solid-state sources in the mm-wave and THz region has changed dramatically in recent years with the development and continual improvement of MMIC-based solid-state power amplifiers operating around 100 GHz. The reader is referred to a review article for more details.⁵⁵ Although outside the scope of this chapter, it is important to recognize that SSPAs are a great enabler for sensors in the mm-wave and THz regions alike. Around 100 GHz they can be used as the basis for power-combining networks to produce power levels far above that possible from fundamental oscillators. Around 300 GHz and above, SSPAs can produce useful power levels (> 1 mW) and high tunability by acting as the driver for varactor multiplier chains.

Besides cost, size, and ability to integrate, another benefit of solid-state sources is modulation. Sensors, like communications systems, generally provide some form of analog or digital modulation on the carrier so that the transmitted signal has, for the analog case, the form $S(t) = A(t)\cos[\omega t + \phi(t)]$, where A is the amplitude and ϕ is the phase. Pure amplitude modulation implies $\phi(t)$ is a constant and pure phase (or frequency) modulation implies that $A(t)$ is a constant. In either case, modulation allows improvement of performance or mitigation of the atmospheric and fading effects described in Sec II.

The standard methods of detecting modulated radiation exist as readily at mm-wave and THz frequencies as they do at microwave frequencies and below. For an AM-encoded carrier, the standard demodulation technique is either envelope detection or square-law detection. It can be shown that an envelope detector is the better performer when the pre-detection signal-to-noise ratio is greater than unity, and the square-law detector is better when the pre-detection SNR is below unity.⁵⁶ Because mm-wave and THz sensors have generally operated under the latter condition, square-law detection as discussed in Sec.V should be optimal or nearly optimal in most demodulation applications.

B.2. Pulsed: T-Rays

Pulses transmitters were first developed for radar to get range information through the classic time-of-flight and range-gating techniques. This has not become an issue yet in THz systems made with solid-state sources. However, pulsed transmitters have become interesting for a different reason – if the width of a pulse becomes significantly less than 1 ps, then Fourier analysis shows that it contains significant spectral density in the THz region. Mode-locked lasers, particularly those based on dyes or Ti:sapphire crystals have been developed to the point where they routinely generate ~ 200 fs pulses or shorter. If such a pulse is used to excite an ultrafast- photoconductive

switch (i.e., Auston switch), then the photocurrent from this switch will have THz components. And if the photocurrents are, in turn, coupled to a THz antenna, there will be significant free-space THz radiation.

Important questions are how much power a T-ray transmitter can produce and what the sensitivity of the receiver can be. The Tx power can be estimated roughly from the way photoconductive switches work, which is the application of enough peak energy from the mode-locked laser pulse to short-circuit the switch for some fraction of the laser pulse width t_p . If the switch is biased with voltage V_B and is coupled to a THz antenna with radiation resistance R_A , then the peak THz power available for radiation from the antenna will be roughly $(V_B)^2/R_A$ and the THz pulse energy will be roughly $(t_p/2)(V_B)^2/R_A$. The average THz power P_{ave} will then be $P_{ave} \approx f_{rep}(t_p/2)(V_B)^2/R_A$ where f_{rep} is the laser repetition frequency. Note that the mode-locked laser power does not enter this estimate provided that it is great enough to short the photoconductive switch at the peak point.

For example, the typical ultrafast photoconductive switch is made of a short-lifetime (~ 1 ps or less) material, such as low-temperature-grown GaAs, exposed between two low-capacitance electrodes, such as a square gap in a THz planar transmission line. To make the switch easily shorted by a mode-locked laser pulse, it can not be made too large in area, 100 square microns being typical. The safe bias voltage is then ~ 20 V. A Ti:sapphire mode-locked laser typically produces 200 fs pulses at $f_{rep} \approx 100$ MHz. These values result in a peak THz pulse power of 4 W, a THz pulse energy of 0.4 pJ, and an average THz pulse power of 40 μ W. This is not too much greater than the maximum values that have been measured by several researchers in the field, which is typically in the range of 1 to 100 μ W.⁵⁷ While very useful for laboratory spectroscopy and imaging, these power levels are not yet considered to be high enough for remote sensing applications. With the continuous advances being made in photonics and optoelectronics, T-rays are an intriguing prospect for future pulsed THz applications.

C. Receiver Types

C.1. Direct Conversion with Classical Square-Law Detector

To get a useful estimate of the sensitivity of an active sensor in the present scenario, we start with the generic expression for the SNR of a direct-detection receiver from Sec. VII, written for one spatial mode. To this we add the two above expressions for the incident power and the radiation fluctuations, resulting in

$$\left(\frac{S}{N}\right)_{AD} = \frac{\left[\eta \cdot \frac{P_T G_T G_R \lambda^2}{(4\pi)^2 R^2} \tau(\nu)\right]^2}{2 \cdot \eta \cdot \{(k_B T_{eff} \cdot \Delta\nu)^2 + (h\nu_0 P_{inc} \Delta\nu)\} \cdot (\Delta f / \Delta\nu) + (\eta \cdot NEP'_{elect})^2 \cdot \Delta f}$$

As discussed above, the range, RF bandwidth, and antenna-gain requirements in mm-wave and THz sensors along with the limited Tx power often lead to the condition where the thermal-background radiation noise dominates the coherent incident radiation noise. This allows us to simplify the SNR to the following form

$$\left(\frac{S}{N}\right)_{AD} = \frac{\left[\eta \cdot \frac{P_T G_T G_R \lambda^2}{(4\pi)^2 R^2} \tau(\nu)\right]^2}{2 \cdot \eta \cdot (k_B T_{eff} \cdot \Delta\nu)^2 \cdot (\Delta f / \Delta\nu) + (\eta \cdot NEP'_{elect})^2 \cdot \Delta f} \quad (60)$$

This expression has great utility, but for the sake of brevity only one form will be given here. This is the noise equivalent transmit power, NEP_T , analogous to the NEAT for a passive sensor observing a thermal source. Setting the SNR to unity, we find

$$NEP_T = \frac{(4\pi R)^2}{\tau \cdot G_T G_R \lambda^2} \sqrt{2 \cdot (k_B T_{eff})^2 (\Delta f \cdot \Delta\nu) / \eta + (NEP'_{elect})^2 \cdot \Delta f}$$

or in the more convenient specific form

$$NEP_T' = \frac{(4\pi R)^2}{\tau \cdot G_T G_R \lambda^2} \sqrt{2 \cdot (k_B T_{eff})^2 \Delta\nu / \eta + (NEP'_{elect})^2} \quad [\text{W/Hz}^{1/2}]$$

Note that this expression could also be arranged in terms of the range at which the sensor SNR would be unity at each possible transmit power. This is the normal way to think about active sensors (i.e., radars) at microwave frequencies where transmit power is not so hard to come by.

In the event that the first term in the radicand dominates in either expression above, we obtain a “background-limit” analogous to that of passive sensor:

$$NEP_T = \frac{(4\pi R)^2 k_B T_{eff}}{\tau \cdot G_T G_R \lambda^2} \sqrt{2 \cdot \Delta\nu \cdot \Delta f / \eta}$$

Fig. 27 shows a plot of the NEP_T as a function of range parametrized by the electrical NEP and the atmospheric transmission factor. It is clear that even at moderate range, the worst-case direct detector having an NEP'_{elect} of 10^{-12} can not provide useful sensitivity under any atmospheric condition unless the transmit power is $\gg 10$ mW. So one would be advised in an active receiver with direct detection to use an NEP no worse than roughly 10^{-13} W/Hz^{1/2}. According to Eqn 60 this produces an SNR of 1.0 with approximately 1 mW of transmit power at a range of 1 km.

C.2. Heterodyne and Homodyne Conversion with Classical Square-Law Detector

In passive sensors the source spectrum is generally thermal radiation so is naturally very broad. Hence, mixing with a local oscillator will generate a broad IF spectrum and heterodyne conversion works very well. By contrast, with an active

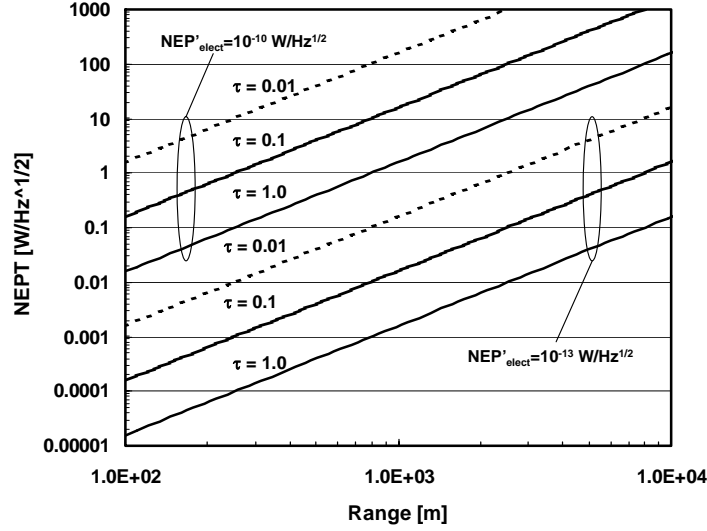


Fig. 27. Noise equivalent transmit power for the active sensor with a direct-detection receiver as a function of range separating the transmitter and receiver. The lines are parametrized by the net atmospheric transmission factor τ and the direct-detector NEP'_{elect} . The other parameters in the curve are $G_T = G_R = 100$, $\eta = 0.5$.

sensor heterodyning can occur only if the Tx and Rx local oscillator are deliberately offset in frequency, which then requires a more sophisticated receiver (e.g., in-phase and quadrature processing) to maintain good sensitivity. This is practical in the millimeter-wave region but becomes expensive and cumbersome at THz frequencies. In the latter region there is a tendency to build sensors that use the same oscillator for both the Tx and LO functions. This results in a homodyne receiver, as described briefly in Sec. IV.

To get a useful estimate of the sensitivity of an active sensor in the present scenario, we note that both heterodyne and homodyne receivers can be described by the generic Eqn 57 for the post-detection SNR in a single spatial mode with radiation noise dominated by the LO photon shot noise. To this we add the above expression for the incident power to get the surprisingly simple results

$$\left(\frac{S}{N}\right)_{AD} = \frac{\left\{H \cdot \eta_R^2 \cdot \frac{P_T G_T G_R \lambda^2}{(4\pi)^2 R^2} \tau(\nu)\right\}^2}{2 \cdot \{\eta_R \cdot h\nu_{LO} + (\eta_R)^2 \cdot NEP'_{\text{mixer}}\}^2 \cdot (B \cdot \Delta f)} \equiv \frac{\left\{H \cdot \frac{P_T G_T G_R \lambda^2}{(4\pi)^2 R^2} \tau(\nu)\right\}^2}{2 \cdot \{h\nu_{LO} / \eta_{HET}\}^2 \cdot (B \cdot \Delta f)} \quad (61)$$

where B is the post-mixer electrical bandwidth (centered about zero for the homodyne case) and the factor H = 1 and 2 for heterodyne and homodyne conversion, respectively.

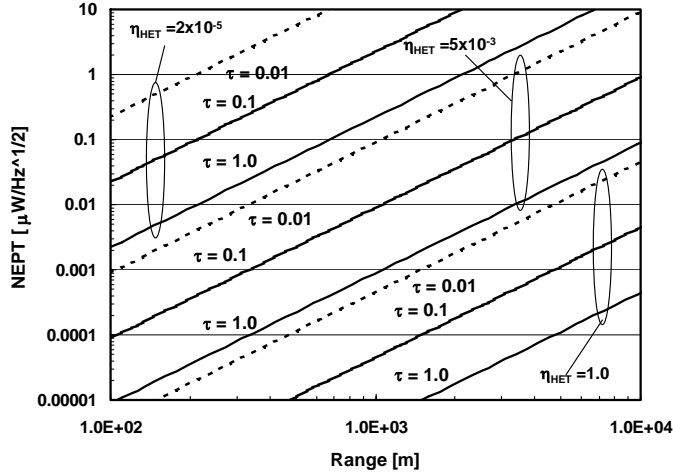


Fig. 28. Noise equivalent transmit power for the active sensor with a heterodyne receiver as a function of range separating the transmitter and receiver. The lines are parametrized by the net atmospheric transmission factor τ and the heterodyne quantum efficiency η_{HET} . The other parameters in the curve are $G_T = G_R = 100$, $\eta = 0.5$.

This then leads to the noise equivalent transmit power

$$NEP_T^i = \frac{(h\nu_{LO} / \eta_{\text{HET}})(4\pi R)^2 \sqrt{2 \cdot B}}{\tau \cdot G_T G_R \lambda^2}$$

Similar to the direct-detection case, this can be plotted vs range with τ and η_{HET} as parameters. The results are shown in Fig. 28 where the vertical axis is now shown in units of $\mu\text{W}/\text{Hz}^{1/2}$. Here we see a very large improvement in performance compared to the direct detection case. On first inspection this seems a bit incredible with even relatively poor heterodyne or homodyne mixers resulting in outstanding sensitivity. For example, a sensor in which the receiver mixer has an η_{HET} of 0.5% achieves an SNR of unity for a transmit power of 0.01 μW per $(\text{Hz})^{1/2}$ of post-detection bandwidth over a range of 1 km and an atmospheric transmission of 0.1. This means that for a more practical transmit power of 1 mW and a post-detection bandwidth of 1 MHz – enough to accommodate AM or FM modulation to help mitigate atmospheric and gain fluctuations – the SNR would be approximately $(10^{-3}) / (0.01 \times 10^{-6}) / (10^6)^{1/2} = 100$.

D. An Illustrative Example: Design of an active sensor to measure absorption signatures

Recent measurements of the electromagnetic transmission through *Bacillus subtilis*, an anthrax surrogate, has revealed absorption signatures at sub-THz

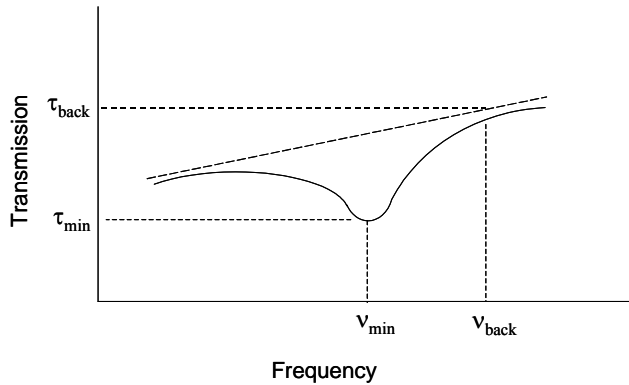


Fig. 29. Representative absorption signature in the terrestrial atmosphere.

frequencies.⁵⁸ The physical origin of these signatures is still being investigated, but from the measured absorption strength and signature one can begin to estimate their detectability in a remote sensor. This is not to imply that the resulting THz sensor will be better than competitive sensors in other (e.g., IR and visible) spectral regions. Rather, this problem serves as an interesting example of how the THz region may offer unique phenomenology in the form of unique absorption *signatures* not present in other electromagnetic regions. This is similar to the recent use of THz radiation for medical imaging. In that case, the THz provides unique contrast between tissue types that is proving useful in diagnostic imaging.

D.1. Absorption Signature Characteristics

A representative absorption signature is shown in Fig. 29 as it might appear in the power transmitted through a cloud of bioparticles. As discussed above, the background transmission in the THz region generally has a complex behavior. Although shown as a slowly increasing transmission with frequency, the background could display the opposite slope or be resonant depending on the proximity to water absorption lines. In any case the bioparticle absorption introduces a broad “dip” in the transmission spectrum that is characterized by a minimum transmission τ_{\min} at frequency ν_{\min} . The goal is to characterize the signature in a way that can be used in the active sensor scenario.

Table I. THz absorption signatures of <i>B. Subtilis</i>			
Freq, ν_{\min}	α_0 (ν_{\min}) (cm^{-1})	$\nu_{\text{back}}, \Delta\nu$ (GHz)	τ (ν_{back})
421.5 GHz	0.7	430.5, 10.0	0.25
619.5 GHz	1.3	600.0, 19.5	$\sim 10^{-11}$
940.05 GHz	1.7	930.0, 10.5	$\sim 5 \times 10^{-4}$
1075.5 GHz	1.2	1057.5, 180	$< 10^{-30}$

To do this we characterize the dip by a depth $\Delta\tau$, and a half-width $\Delta\nu$. $\Delta\tau$ is the absolute difference between τ_{\min} and the background transmission τ_{back} measured at a frequency ν_{back} on whichever side of ν_{\min} that τ is greater (e.g., in Fig. 29 $\nu_{\text{back}} \gg \nu_{\min}$). Hence, $\Delta\nu \equiv |\nu_{\min} - \nu_{\text{back}}|$. We assume further that the concentration of bioparticles ρ is low enough for their attenuation to be described by a coefficient α that is linearly dependent on ρ and that affects the transmission through the Lambert-Beer law $\tau(\nu) = \exp[-\alpha_0(\nu)L\rho/\rho_0]$, where L is the thickness of the bioparticle cloud and α_0 is a reference attenuation coefficient measured at a concentration ρ_0 that may be much different than the actual ρ . Then if the background is slowly varying over $\Delta\nu$, we can write

$$\begin{aligned}\Delta\tau &\equiv \tau_{\text{back}} - \tau_{\min} = \exp[-\alpha_0(\nu_{\text{back}})L\rho/\rho_0] - \exp[-\alpha_0(\nu_{\min})L\rho/\rho_0] \\ &\approx \tau_{\text{back}} \{1 - \exp[-\alpha_0(\nu_{\min})L\rho/\rho_0]\}\end{aligned}$$

Listed in Table I are the values of ν_{\min} , $\alpha_0(\nu_{\min})$, ν_{back} and $\Delta\nu$ derived from laboratory transmission measurements through dry films of *B. subtilis* (an anthrax surrogate) containing $\rho_0 \approx 1 \times 10^{12} \text{ cm}^{-3}$ – a density that is necessary to get an accurate measure of $\Delta\alpha$, but is much larger than expected in airborne bio-warfare agents. The results listed in Table I describe four different absorption features having center frequencies between 421 and 1075 GHz. In the THz region, the remote detection of these signatures depends critically on the atmospheric transmission $\tau(\nu_{\text{back}})$. Upon analysis of the results of PCLnWin shown in Fig. 5, the most transparent frequency by far is 421.5 GHz through which the single pass attenuation is approximately 0.25 (i.e., -6.0 dB) over a path length of 1 km at sea level and at 60% relative humidity .

D.2. Sensor Design: Direct and Homodyne Differential Absorption Radars

To take advantage of the $\Delta\tau$ we imagine an active sensor such as that shown in Fig. 30. The transmitter contains a frequency agile source that can hop between the two frequencies, keeping the transmit power constant. To make the scenario a bit more practical, we suppose the Tx and Rx are co-located and used with a retrodirective mirror that is perfectly specular. We can thus write for the difference in incident power between the two frequencies:

$$\Delta P_{\text{inc}}(\nu) = \frac{P_T G_T G_R \cdot \lambda^2}{(8\pi R)^2} \Delta\tau(\nu)$$

In practice one can readily do this frequency hopping periodically (e.g., square wave) and then synchronously demodulate in the post-detection portion of the receiver, assuming of course that Δf is greater than the modulation frequency. In this case, the post de-modulator signal-to-noise ratio will be approximately the difference in the SNRs given by Eqn 61 between the two frequency states with a reduction factor f_m for the inefficiency of simple (e.g., AM) demodulation schemes. This results in the post-detection SNR for a direct receiver of

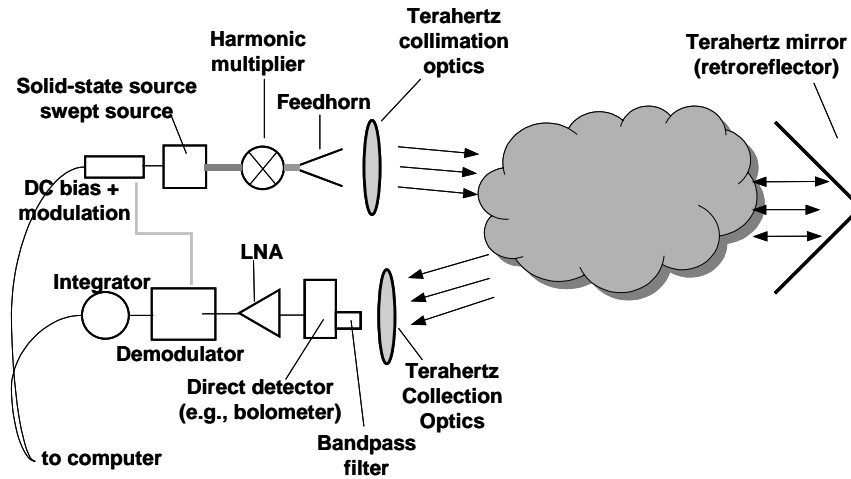


Fig. 30. Block diagram of active sensor designed for differential absorption detection of a cloud of weak bioparticles at a short standoff (~ 1 km). The receiver is based on direct detection.

$$\left(\frac{S}{N}\right)_{AD} = \frac{\left[\eta \cdot \frac{P_T G_T G_R \lambda^2}{(8\pi)^2 R^2} f_m \cdot \Delta\tau(\nu) \right]^2}{2 \cdot \eta \cdot (k_B T_{eff} \cdot \Delta\nu)^2 \cdot (\Delta f / \Delta\nu) + (\eta \cdot NEP'_{elect})^2 \cdot \Delta f}$$

and an SNR for a heterodyne (or homodyne) receiver of

$$\left(\frac{S}{N}\right)_{AD} = \frac{\left\{ \frac{P_T G_T G_R \lambda^2}{(8\pi)^2 R^2} f_m \cdot \Delta\tau(\nu) \right\}^2}{2 \cdot \{h\nu_{LO} / \eta_{HET}\}^2 \cdot (B \cdot \Delta f)}$$

As a practical matter, note that one can readily do this frequency hopping periodically at a much higher frequency than the rate of change of the atmospheric fading effects. Then by synchronously demodulating the signal at the receiver and taking the ratio between the two frequency states, one can cancel the effect of atmospheric fading.

Note that in a homodyne version of Fig. 30, because the transmitter and receiver are co-located at the same point in space, only a few components would have to be changed in the THz front-end. First, the direct detector would be replaced by a mixer, such as a Schottky diode. Then, a portion of the transmit power would be coupled through a THz transmission line to the mixer where it would be coupled to the mixer as an LO. Finally, an IF circuit would be added that could process both the in-phase (I) and

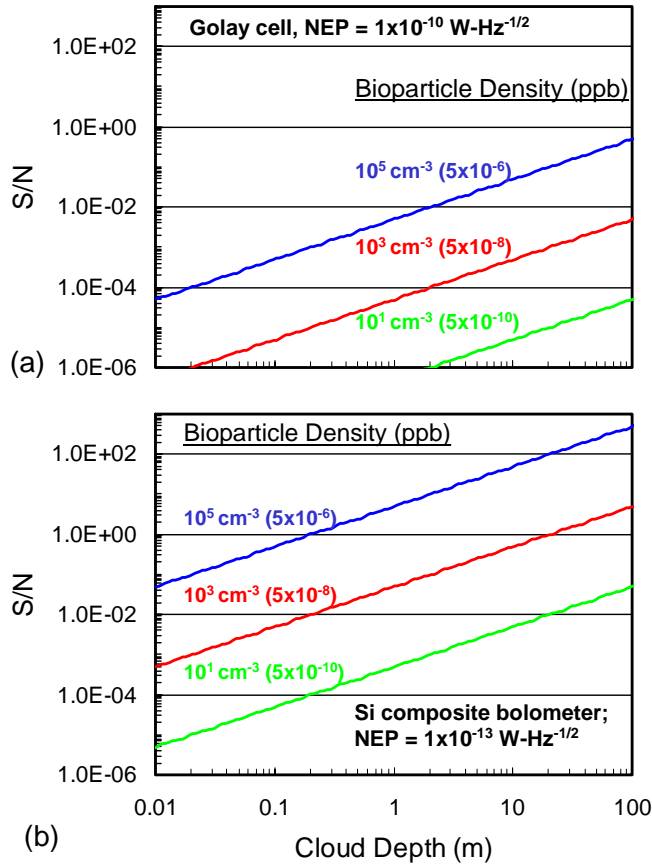


Fig. 31. SNR performance of differential-absorption radar with direct-detection receiver. In (a) the detector is a room temperature Golay cell, and in (b) it is a cryogenic (4.2-K) bolometer.

quadrature (Q) components of the mixer output. The sum of the I- and Q-detected channels would then, in principle, recover all of the signal power of the received coherent signal, independent of phase difference(s) between the received signal and the LO.

D.3. Direct and Homodyne Sensor Performance at 425 GHz

Figs. 31 (a) and (b) show the post-demodulation SNR at three different concentrations of *B. subtilis*, $\rho = 10, 10^3, \text{ and } 10^5 \text{ cm}^{-3}$ for the direct-detection transceiver. The results are plotted vs cloud depth for a fixed Tx-to-cloud separation of 0.5 km and a transmit power of 1 mW. The other scenario parameters are listed in Table II. In Fig.

31 (a), the Rx detector is assumed to have $NEP=1 \times 10^{-10}$ W-Hz^{-1/2} corresponding to a state-of-the-art room-temperature bolometer (e.g., Golay cell). In Fig. 31(b) the Rx detector is assumed to have $NEP = 1 \times 10^{-13}$ W-Hz^{-1/2} corresponding roughly to a state-of-the-art cryogenic (4.2 K) bolometer (e.g., silicon composite).

The results on overall sensor performance are now determined by comparing these SNR curves to the universal ROC curves of Fig. 16. For example, according to Fig. 31(b), in a cloud 20 m thick at a stand-off of 0.5 km, the SNR is 1.0 for a concentration of 10^3 cm⁻³. Assuming that the signal threshold at 0.2 nW, we find Pd and Pfa values from Fig. 16 to be approximately 0.4 and 0.1, respectively – unsatisfactory for most applications. However, the enhanced SNR at greater cloud depth or higher concentration in Fig. 31(b) would result in more reliable detection. For example, a cloud depth of 20 m and a concentration of 10^4 cm⁻³ would yield an SNR of 10 and associated Pd and Pfa values of 0.75 and 0.006 if the threshold was also increased to 0.9 nW.

For the homodyne case, we assume the down-converting device is a Schottky-diode mixer. If driven hard enough ($P_{LO} \sim 1$ mW) so that the noise-equivalent power (NEP) is limited by LO-shot noise and the coupling efficiency η is usefully large, we find $NEP_{HET} \approx h\nu/\eta_{HET} = 5.6 \times 10^{-21}$ W/Hz ($h \equiv$ Planck's constant) for $\nu = 421$ GHz and $\eta_{HET} = 5\%$. The calculated SNR_{AD} curves for coherent detection are plotted in Fig. 32 for the same bioparticle concentrations as in Fig. 31, and the same transmit power and range. The homodyne receiver is assumed to have an IF bandwidth B of 1 MHz and the post-detection bandwidth of 1 KHz. Note that for a given cloud depth and bioparticle concentration, the coherent post-detection SNR is approximately 300 times larger than the analogous post-detection SNR of the incoherent sensor in Fig. 31(b). In other words, a homodyne receiver operating at room temperature is approximately 300 times more sensitive than a direct receiver with a cryogenic (4.2 K) detector under the same sensor scenario.

Substitution of pre-detection SNR results of Fig. 32(a) into the ROC curve of Fig. 18 leads to the following exemplary results. When the cloud depth is 20 m and the concentration is 10^3 cm⁻³, the SNR in Fig. 32(a) is approximately 16 and we find from Fig. 18 that $P_d \approx 0.99$ for $P_{fa} = 10^{-3}$, $P_d \approx 0.95$ for $P_{fa} = 10^{-4}$, and $P_d \approx 0.90$ for $P_{fa} = 10^{-5}$, each possibility corresponding to one particular threshold level. Clearly, the homodyne

Background temperature (K)	290
Detector coupling efficiency	0.5
Line center frequency (GHz)	421
Background frequency (GHz)	431
Linewidth (GHz)	10
Line center abs coeff (1/cm)	0.7
Background transmission	0.25
Spectral bandwidth (GHz)	10
Integration time (s)	1.0
Transmit power (mW)	1.0
Tx aperture (cm ²)	100
Rx aperture (cm ²)	100
Tx-Rx separation (m)	1000
Cloud thickness (cm)	variable
Cloud temperature (K)	290
Molecule density (cm ⁻³)	$10, 10^3, 10^5$

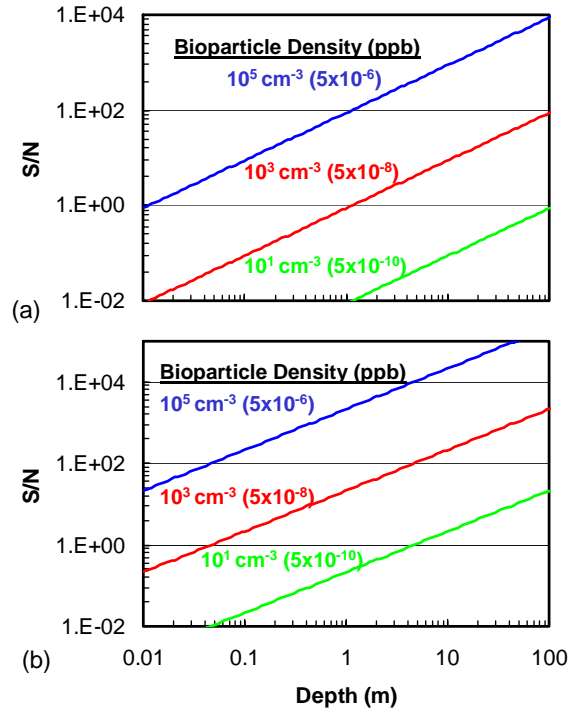


Fig. 32. (a) Pre-detection SNR of homodyne receiver in the differential-absorption radar scenario. (b) SNR after square-law detection and synchronous demodulation. The parameters are $\eta_{\text{HET}} = 0.05$, $\text{NEP} = 5.6 \times 10^{-21}$ W/Hz.

receiver yields a far superior performance in comparison with the direct receiver. With this superior SNR, one can envision an early-warning system in which the presence of bioparticles sets off an alarm based on the envelope of signal plus noise exceeding a threshold set in baseband.

XI. 2D Imaging and the Quest for Popular Applications

A. Heuristics

A holy grail of sensor technology in the mm-wave, THz or any other band of the electromagnetic spectrum is the capability to do imaging of objects at a useful stand-off in real or near-real time. This capability is motivated, of course, by the appeal and utility of live-video and television. The appeal needs no explanation. The utility goes well beyond the entertainment or practical impact of live video and involves such issues as pattern and object recognition. The fact remains that the human visual-cognitive system

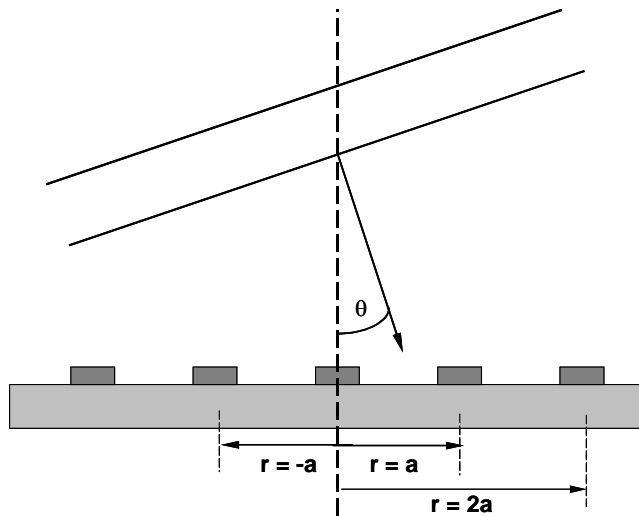


Fig. 33. Schematic diagram of optical spatial sampling

(i.e., the combination of the eyes and the brain) is still much better than computers or other information machines at recognizing objects surrounded by other objects or hidden in “clutter.” And a good, if not the best, way to couple the imagery to the visual-cognitive system is a two-dimensional live image.

The live-video advantage has driven sensor technology since the beginnings of television with continual development of better cameras operating across the visible band. The early technology, such as that of the *vidicon*, was essentially an electronic version of photographic film. The technological breakthrough that revolutionized visible imagers during the past few decades has been the discretization the imaging process. The concept is simple: emulate the behavior of photographic film by placing individual detectors in a dense two-dim array in the focal plane of an optical imaging system. The detector density required to obtain image contrast comparable to that of the photographic film are well defined by a version of Nyquist’s sampling theorem applied to free-space plane waves and imaging – the Whitaker-Shannon (W-S) theorem. This important result is discussed briefly below and in great detail in many books on Fourier optics and optical imaging.

The benefits of live video expanded significantly with the realization that discrete imagers could be readily extended into other parts of the electromagnetic spectrum where the phenomenology behind free space propagation are different. Perhaps the best example is the infrared region where it was realized quickly that objects that might be unapparent to a visible-band imager could be readily apparent to an infrared imager, particularly one constructed to operate in one of the well-known IR atmospheric windows from 3-to-5 micron or 8-to-12 micron wavelength. And the phenomenology behind this realization is as old as quantum physics itself – it is just blackbody, or

“thermal” radiation. Terrestrial objects tend to have a physical temperature between roughly 273 and 300 K. According to the blackbody spectral power distribution of Eqn 8, the peak of emission tends to occur between approximately 8 and 10 micron. If the object in question has a temperature sufficiently lower or higher than the background and has a *sufficiently high emissivity*, it will be apparent to the camera either as a dark or bright image, respectively.

The same reasoning applies equally well to making thermal imagers in the mm-wave and THz region with the obvious difference that objects at any terrestrial temperature generally emit much less power than they do in the infrared. On the other hand, as shown above, the radiation noise is a lot lower in magnitude, particularly the photon shot noise. So the receiver sensitivity can be much lower in the mm-wave and THz region. More importantly, the mm-wave and THz regions display some phenomenological differences beyond the atmospheric and materials propagation differences discussed in Sec. II.E. One such difference relates to the difference in texture and scattering caused by the large difference in wavelength. Air-solid interfaces that appear rough and, therefore, diffusive, in the infrared can appear relatively specular in the mm-wave and THz regions. A good example is a man-made wall made from a common building material, such as drywall.

B. Imaging by Spatial Sampling

Conceptually the W-S theorem states that the discrete imager can recover the same spatial information as from a continuous film if the discrete pixels sample at least twice per *spatial frequency* of the highest spatial frequency contained in the image. The definition of spatial frequency follows from the ability to decompose any spatial distribution of radiation into a Fourier series of plane waves propagating at different angles with respect to the axis of propagation. This is readily seen by the construction shown in Fig. 33, which shows an arbitrary plane wave incident on a periodic imaging array of period a . From fundamental electromagnetics, the plane wave can be represented by $E_0 \exp[-j(\mathbf{k} \cdot \mathbf{r} - \omega t)]$ where \mathbf{k} is the propagation vector and \mathbf{r} is the position of the measurement point in some reference system whose origin is conveniently chosen as the center of the array. In this case, the vector inner product is $\mathbf{k} \cdot \mathbf{r} = (2\pi/\lambda) \sin\theta (m \cdot a)$, where λ is the free-space wavelength and m is an integer. In order for this phase factor to be unambiguous between neighboring elements ($m = 1$) of the array, it must lie in the range $-\pi$ to π since this comprises one entire branch of the Argand plane. Hence, the (un)ambiguity condition becomes $(2\pi/\lambda) |\sin\theta| a < \pi$, or

$$a < \lambda / (2 |\sin\theta|) .$$

This is often written as

$$|\sin\theta| / \lambda < 1/2a$$

which has the same form as Nyquist’s sampling theorem if we recognize $1/a$ as the spatial sampling rate, $\sin\theta/\lambda$ as the spatial frequency, and $1/2a$ as the Nyquist sampling rate.

These simple equations have several interesting properties, one of which relates to the maximum allowed spacing of the elements. For an antenna or optical front-end providing a very wide angular field-of-view, θ will vary from nearly $-\pi/2$ to $\pi/2$, the maximum $|\sin \theta|$ will approach 1, and thus the element spacing must be $\leq \lambda/2$ for

unambiguous imaging. This is a familiar result from phased-array antenna theory, where it is usually stated in the reciprocal sense: i.e., the maximum spacing that array elements can have and still radiate “unambiguous” beams is $\lambda/2$. The ambiguity is manifest in the antenna performance through radiation directed into grating lobes.

Spatial sampling places just an upper limit on a , so that one is free to space the elements arbitrarily closer with no penalty on performance. This is analogous to “oversampling” in signal processing, and can be unnecessarily expensive and complicated. So the actual element spacing is usually determined, as in visible and infrared imaging arrays, by spatial resolution requirements, which are a strong function of diffraction effects in the optical or antenna components in front of the array. Once the diffraction effects are determined, which is usually done by spatial resolution requirements of the sensor as a whole, there is no point in placing the elements in a center-to-center spacing smaller than the diffraction-limited diameter for a plane-wave incident on the sensor.

Note that strictly speaking, the above considerations apply only to a sensor having phase sensitive, or coherent, elements. In other words, each element must be connected to coherent receiver. The analogous conditions for an incoherent imaging system are a factor-of-two different and discussed in great detail with an emphasis on mm-wave and THz imaging in the excellent review article by Rutledge et al.⁵⁹

C. Applications and Examples of MM-Wave Imagers

A practical advantage of mm-wave and THz imagers over visible and infrared ones is packaging. Because of the much greater maximum pixel spacing that generally applies, the former imagers can be constructed with “macroscopic” spacings of the order of 1 mm. This fact has been utilized for decades through the construction of first one- and then two-dimensional arrays of incoherent thermal detectors. Perhaps the first such array and associated camera system was engineered in the 1970s using a linear-array of pyroelectric detectors and later productized by Spiricon, Inc. The single-pixel NEP_{DIR} values were approximately 10^{-8} W/Hz^{1/2}. While good enough for some applications, such as THz laser beam diagnostics, the sensitivity was not adequate for most remote sensing. Later it was realized that the mm-wave and THz coupling efficiency to the pyroelectric elements was not very good because of low absorption coefficient in the typical (ferroelectric) materials. And the old trick of blackening the surface of the elements with graphite-loaded paint was not nearly as effective as in the mid-infrared region where pyroelectric detectors in “uncooled” bolometer arrays continue to be very useful.

By the late 1970s it was realized that to make widely useful mm-wave and THz imaging cameras, significant improvement would be required in the radiation coupling efficiency. Considerable research was conducted on planar antennas coupled to free space through dielectric substrates, cavities, and by other techniques. This culminated in the first room-temperature two-dim incoherent imaging arrays made from bow-tie antennas coupled to bismuth microbolometers.^{60, 52} Somewhat later the first coherent imager was developed and demonstrated, this time using a linear array of Schottky diode mixers coupled to bow-tie antennas.⁶¹ Both were developed and demonstrated primarily for plasma diagnostics – an area of widespread research, particularly after the energy crisis of the 1970s.

During the mid to late 1980s the strong push for plasma imaging had waned, and the primary system pull for mm-wave and THz imagers came from the radio astronomy

community.⁶² It was recognized that some of the cryogenic detector types that had been developed in the 1970s for high-sensitivity astronomy applications were amenable to fabrication in planar arrays. Two examples were LHe-cooled composite bolometers for direct detection,⁶³ and superconductor-insulator-superconductor tunnel junctions for coherent detection.⁶⁴ Although these concepts have proven to be successful, they were not quick to develop largely because of the difficulty in fabricating and packaging cryogenically cooled devices and read-out electronics.

The interest and activity in imagers increased significantly in the mid 1990s with two developments. First, monolithic microwave integrated circuits based on GaAs devices improved to the point where they could provide superior performance to discrete components in the mm-wave band. Perhaps the first example of this was the low-noise amplifier (LNA). Second, two compelling mm-wave applications were identified that offered a compelling advantage over visible and infrared imagers: (1) concealed weapons detection and (2) all-weather imaging for aircraft landing.

The sensing of weapons concealed behind human clothing is a natural one for mm-wave and THz sensors since clothing is, by design, opaque in the visible and more or less opaque in the infrared depending on the fabric. For the same reason that plastics are quite transparent to mm- and THz waves, so are most of the common fabrics. One of the first, if not the first, demonstrations of concealed weapons was made by engineers at the Millimetrix Corp^{65,66} using both active and passive mm-wave arrays. Although not highly integrated, these demonstrations showed that images of concealed guns could be of high enough quality to allow identification by a human observer.

Around this same time engineers began pursuing mm-wave imagers for all-weather imaging systems. People had long known that this was a natural application because of the size advantages of mm-wave components over lower-frequency RF ones, and the superior mm-wave propagation compared to that in the visible or infrared. The drawback of this application was always size and cost. The development of MMIC LNAs has made preamplified direct detection possible up to at least W band (75-110 GHz) in array architectures. Probably the most successful of these has been the passive camera developed and demonstrated by TRW for all-weather landing.^{67,68} Below 100 GHz, it utilizes their GaAs pHEMT process to produce LNAs having < 3 dB noise figure up to approximately 100 GHz. It has also been developed for 140 GHz using InP-based MMIC technology.⁶⁹

More recently, a two-dimensional array of Nb microbolometers has been used in an active imaging array.⁷⁰ This work has shown excellent imagery at real-time acquisition rates.

Acknowledgement and References

The author is indebted to Dr. Dwight Woolard who commissioned this article and then practiced great patience while it was being composed. The chapter was a challenging exercise since much of the information is not well known and not covered in textbooks or review articles. And some had to be derived from scratch, leading perhaps to original expressions and arguments. The author apologizes in advance for not referencing the same derivations or expressions if they have appeared previously in the open literature. Unfortunately, there is simply not enough time to review the entire embodiment of publications related to remote sensing – the plethora from the rf side of the fence and the plethora from the photonic side. One goal of this article was to

emphasize the dual nature of the THz field, drawing on concepts and techniques from both sides of the fence in a rather unpredictable way.

The author also thanks his many other DoD sponsors over the years, many of whom have supported the author in research related to this article. Among these are Dr. Jim Harvey, Dr. Dev Palmer, Dr. Edgar Martinez, Dr. James Nichter, Dr. Daniel Radack, Dr. Jim Murphy, Dr. Elias Towe, Dr. Henry Everitt, Dr. Gerald Witt, and Dr. Jane Alexander.

The author also thanks his many close colleagues in the mm-wave and THz field over the years, starting with his mentors at the Hughes Space and Communications Group (group of Mr. Frank Goodwin), his advisers and teachers at Caltech (group of Prof. T.G. Phillips), his colleagues and associates at MIT Lincoln Laboratory, his fellow program managers and government associates at DARPA, and his colleagues and students at UCLA and UCSB.

¹ T.G. Blaney, "Radiation Detection at submillimetre wavelengths," *Journal of Physics E (Sci. Inst)*, vol. 11, p. 856, 1978.

² J. C. Wiltse, "History of millimeter and submillimeter waves," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, pp.118-127, Sept. 1984.

³ P.H. Siegel, "Terahertz Technology," *IEEE Transactions on Microwave Theory and Tech*, vol. 50, no. 3, 2002.

⁴ M. Exter and D. Grischkowsky, "Characterization of an optoelectronic terahertz beam system," *IEEE Trans. Microwave Theory Tech.*, vol. 38, pp. 1684-1691, Nov. 1990.

⁵ B. B. Hu and M. C. Nuss, "Imaging with terahertz waves," *Opt. Lett.*, vol. 20, no. 16, pp.1716-1718, Aug. 15, 1995.

⁶ D. M. Middleman, R. H. Jacobsen, and M. C. Nuss, "T-ray imaging," *IEEE J. Select. Topics Quantum Electron.*, vol. 2, pp.679-692, Sept. 1996

⁷ D. Marcuse, *Principles of Quantum Electronics* (Academic Press, New York, 1980).

⁸ F. T. Ulaby, R.K. Moore, and A.K. Fung, "Microwave Remote Sensing," vol. 1, sec. 5.15 (Addison-Wesley, Reading, MA), sec. 5-15

⁹ R. Eisberg and R. Resnick, "Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles, 2nd Ed. (Wiley, New York, 1985).

¹⁰ W.B. Davenport and W.L. Root, "An Introduction to the theory of random signals and noise," (IEEE Press, New York, 1987).

¹¹ R. H. Kingston, "Detecton of Optical and Infrared Radiation" (Springer, 1978).

¹² PcLnWin, market by Ontar, Inc., Andover, MA

¹³ J.D. Kraus, "Radio Astronomy", (McGraw-Hill, New York, 1966), Sec. 3-16.

¹⁴ F.T. Ulaby, R.K. Moore, and A.K.Fung, "Microwave Remote Sensing," (Addison Wesley, New York, 1981), V. 1.

¹⁵ J-A. Kong R-T. Shin, L. Tsang, "Theory of Microwave Remote Sensing," (Wiley, New York, 1985).

¹⁶ see, www.josephson.terahertz.co.uk/QMCI/GOLAY.HTM

¹⁷ see Infrared Laboratories, Inc., www.irlabs.com

¹⁸ R.S. Elliott, "Antenna Theory and Design," (Prentice Hall, Englewood Cliffs, 1981).

¹⁹ H. R. Fetterman, P. E. Tannenwald, B. J. Clifton, W. D. Fitzgerald, and N. R. Erickson, "Far IR heterodyne radiometric measurements with quasioptical Schottky diode mixers," *Appl. Phys. Lett.*, vol. 33, no. 2, pp. 151-154, (1978).

²⁰ D. Rutledge, D. Neikirk, and Kasilingham, "Printed Circuit Antennas" in *Infrared and Millimeter Waves*, vol.10, ed. By K. J. Button (Academic, New York, 1983), p. 1.

²¹ L. Katehi and N. Alexopolous, *IEEE Trans. Ant. And Propag.*, Vol. 31, p. 34 (1983).

-
- ²² K.S. Yngvesson, D.H. Schaubert, T.L. Korzeniowski, E.L. Kollberg, T. Thungren, and J.F. Johansson, "Endfire tapered slot antennas on dielectric substrates," *IEEE Trans. Antennas and Prop.*, Vol. 33, p. 1392-1400 (1985)
- ²³ M. Kominami, D.M. Pozar and D.H. Schaubert, "Dipole and slot elements and arrays on semi-infinite substrates," *IEEE Trans. Antennas and Propagation*, vol. AP-33, pp. 600-607 (1985).
- ²⁴ D. F. Filipovic, G.P. Gauthier, S. Raman, and G.M. Rebeiz, "Off-axis properties of silicon and quartz dielectric lens antennas," *IEEE Trans. On Antennas and Propagation*, vol. 45, pp. 760-766 (1997).
- ²⁵ For an excellent review of this topic, see "Classical Electromagnetic Radiation," J. Marion (Academic, New York, 1965), Chapter 12.
- ²⁶ R. J. Mailloux "Phased Array Antenna Handbook," (Artech House, Norwood, MA, 1994), p. 15.
- ²⁷ A. Yariv, "Quantum Electronics," (Wiley, New York, 1975), p. .
- ²⁸ P. Goldsmith, "Quasi-Optical Systems: Gaussian Beam Quasi-Optical Propagation and Applications " (Wiley IEEE Press, 2001).
- ²⁹ T. G. Phillips and J. Keene, "Submillimeter Astronomy," *Proc. IEEE*, vol. 80, pp. 1662-1678, Nov. 1992
- ³⁰ P.D. Potter, "A new horn antenna with suppressed sidelobes and equal beamwidth," *Microwave J.* Vol. 6, pp. 71-78 (1963).
- ³¹ R. Winston and W. T. Welford, "High-Collection Non-Imaging Optics," (Academic, New York, 1989).
- ³² A. Yariv, "Quantum Electronics, 2nd Edition", (Wiley, New York, 1975), Chapter 6.
- ³³ A. Yariv, *IBID*.
- ³⁴ H. Wang, L. Samoska, T. Gaier, A. Peralta, H. H. Liao, Y. C. Leong, S. Weinreb, Y. C. Chen, M. Nishimoto, and R. Lai, "Power amplifier modules covering 70-113 GHz using MMICs," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 9-16, Jan. 2001.
- ³⁵ S. Weinreb, T. Gaier, R. Lai, M. Barsky, Y. C. Leong, and L. Samoska, "High-gain 150-215 GHz MMIC amplifier with integral waveguide transitions," *IEEE Microwave Guided Wave Lett.*, vol. 9, pp. 282-284, July 1999
- ³⁶ TRW reference.
- ³⁷ A. H. Dayem and R. J. Martin, "Quantum interaction of microwave radiation with tunneling between superconductors," *Phys. Rev. Lett.*, vol. 8, pp. 246-248, Mar. 1962.
- ³⁸ G. J. Dolan, T. G. Phillips, and D. P. Woody, "Low noise 115 GHz mixing in superconductor oxide barrier tunnel junctions," *Appl. Phys. Lett.*, vol. 34, pp.347-349, Mar. 1979.
- ³⁹ D. E. Prober, "Superconducting terahertz mixer using a transition-edge microbolometer," *Appl. Phys. Lett.*, vol. 62, no. 17, pp. 2119-2121, 1993.
- ⁴⁰ J. W. Waters, "Submillimeter-wavelength heterodyne spectroscopy and remote sensing of the upper atmosphere," *Proc. IEEE*, vol. 80, pp.1679-1701, Nov. 1992.
- ⁴¹ Kingston, *ibid*.
- ⁴² Davenport and Root, *ibid*.
- ⁴³ Davenport and Root, *ibid*.
- ⁴⁴ Kingston, *ibid*, Sec. 2.4.
- ⁴⁵ J. D. Kraus, "Radio Astronomy", (McGraw Hill, New York, 1966), Chap. 7.
- ⁴⁶ A. van der Ziel, "Noise in Solid-State Devices and Circuits," (Wiley, New York, 1986).
- ⁴⁷ A. van der Ziel, *ibid*.
- ⁴⁸ W.B. Davenport and W.L. Root, *ibid*.
- ⁴⁹ C.D. Motchenbacher and F.C. Fitchen, "Low-Noise Electronic Design," (Wiley, New York, 1973).
- ⁵⁰ J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, L. Ridenour and G. Collins, Eds. New York: McGraw-Hill, 1950.
- ⁵¹ S.O. Rice "Mathematical analysis of random noise," *Bell System Tech Journal*, vol. 24, pp. 46-156, January 1945.

-
- ⁵² M. Schneider, "Metal Semiconductor Junctions as Frequency Converters," in *Infrared and Millimeter Waves*, vol. 6, Chapter 4, p. 209 (1982).
- ⁵³ Dr. Peter Siegel, private communication.
- ⁵⁴ H. C. Torrey and C. A. Whitmer, *Crystal Rectifiers*, Ed. By L. Ridenour and G. Collins (New York, McGraw-Hill) 1950.
- ⁵⁵ P.H. Siegel, *ibid.*
- ⁵⁶ Davenport and Root, *ibid.*
- ⁵⁷ X-C. Zhang, "Terahertz wave imaging: horizons and hurdles," *Phys. Med. and Biol.*, vol 47, pp. 3667 – 3677 (2002)
- ⁵⁸ D. Woolard, R. Kaul, R. Suenram, A. H. Walker, T. Globus, and A. Samuels, "Terahertz electronics for chemical and biological warfare agent detection," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Anaheim, CA, June 13-19, 1999, pp.925-928.
- ⁵⁹ D. Rutledge, D. Neikirk, and Kasilingham, "Printed Circuit Antennas" in *Infrared and Millimeter Waves*, vol.10, ed. By K. J. Button (Academic, New York, 1983), p. 1
- ⁶⁰ D.P. Neikirk, P.P. Tong, D.B. Rutledge, H. Park. And P.E. Young, *Appl. Phys. Lett.* Vol. Vol. 41, pp. 329-331 (1982); D. P. Neikirk, W. W. Lam, and D. B. Rutledge, "Far-infrared microbolometer detectors," *Int. J. Infrared Millimeter. Waves*, vol. 5, no. 3, pp. 245-278, Mar. 1984.
- ⁶¹ C.-E Zah, D. Kasilingam, J.S. Smith, D.B. Rutledge, T-C Wang, and S.E. Schwarz, "Millimeter-Wave Monolithic Schottky Diode Imaging Arrays," *Int. J. Infrared and MM Waves*, 1985.
- ⁶² T. G. Phillips and J. Keene, "submillimeter astronomy," *Proc. IEEE*, vol. 80, pp. 1662-1678, Nov. 1992.
- ⁶³ N. S. Nishioka, P. L. Richards, and D. P. Woody, "Composite bolometers for submillimeter wavelengths," *Appl. Opt.*, vol. 17, no. 10, pp. 1562-1567, May 1978.
- ⁶⁴ T.G. Phillips and D.M. Watson, "Baseline Study for a Large Deployable Reflector," private communication.
- ⁶⁵ G.R. Hgueenin, "Millimeter-wave concealed weapons detection and through-the-wall imaging systems," *Proc. SPIE*, Vol. 2938, (1996).
- ⁶⁶ G.R. Huguenin, E.L. Moore, S. Bandia, and J.J. Nicholson, "A Millimeter-Wave Monolithic load switching twist reflector for compact imaging cameras," *IEEE Trans. Microwave Theory and Tech.*, vol. 44, p. 2751 (1996).
- ⁶⁷ M. Shoucri, R. Davidheiser, B. Hauss, P. Lee, M. Mussetto, S. oung, and L. Yujiri, "A Passive MM-Wave Camera for Landing in Low Visibility Conditions," *Proc. of the 1994 IEEE Digital Avionics Conference*, p. 93.
- ⁶⁸ "Passive millimeter-wave camera," L. Yujiri, M. Biedenbender, M. R. Flannery, B.I. Hauss, R. Kuroda, P. Lee, H.H. Agravante, G.S. Dow, S. Fornaca, R.L. Johnson, K. Jordan, D. Low, B.H. Quon, T.K. Samec, K.E. Yokoyama, A. W. Rowe, M Shoucri, and J. Yun, *SPIE Proceedings Vol. 3604*, April 1997.
- ⁶⁹ L. Yujiri, S. Fornaca, B. Hauss, R. Kuroda, R. Lai, and M. Shoucri, "140-GHz Passive Millimeter-Wave Video Camera," *SPIE Proceedings*, Vol. 3703, 1999, p. 20.
- ⁷⁰ E.N Grossman and A.J. Miller, "Active Millimeter-Wave Imaging for Concealed Weapons Detection," *Proc. SPIE*