SIMPLIFIED HOMODYNE DETECTION FOR FM CHIRPED LIDAR

BY

Peter Adany
B.S., Electrical Engineering,
University of Kansas, 2005

Submitted to the Department of Electrical Engineering and Computer Science and the faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

Thesis Committee:

Dr. Rongqing Hui: Chairperson

Dr. Christopher Allen

Dr. James Stiles

Date Defended: November 30, 2007

The Thesis Committee for Peter Adany certifies
That this is the approved Version of the following thesis:

SIMPLIFIED HOMODYNE DETECTION FOR FM CHIRPED LIDAR

Committee:

Dr. Rongqing Hui: Chairperson

Dr. Christopher Allen

Dr. James Stiles

Date Approved:

ABSTRACT

The investigation of global warming requires more sensitive altimeters to better map the global ice reserves. A homodyne detection scheme for FM chirped lidar is developed in which dechirping is performed in the optical domain, simplifying both the optical and the RF circuits compared to heterodyne detection. Experiments show that the receiver sensitivity approaches the quantum limit and surpasses the performance of direct and heterodyne detection. In addition, the required electrical bandwidth of the photodiode and receiver RF circuitry are both significantly reduced, facilitating the use of large area photodetector arrays. A field trial using a 5”-aperture diameter telescope and a 370-m target range verified the sensitivity estimation and demonstrates the feasibility of this technique. The problem of homodyne carrier fading is addressed by incorporating a phase diversity receiver using a 90-degree optical coupler. Finally, an outline of the future direction of research is given.

ACKNOWLEDGEMENTS

Firstly, I am highly grateful to Dr. Rongqing Hui for the opportunity to be a graduate researcher and for all the encouragement he has given me since the very start of my graduate studies. Over the last two years his positivity and outstanding enthusiasm have given me confidence and raised my aspirations greatly.

I also thank my Thesis committee members, Dr. Chris Allen and Dr. James Stiles. I would especially like to thank Dr. Allen for the many times he helped me understand the underlying principles behind this research and for the many words of support.

Finally I would like to thank my family and friends, and especially my parents for giving so much that I could be a graduate student today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

THE GLOBAL WARMING PROBLEM

Over the last century, the average global surface air temperature increased by 0.74 ± 0.18 ^{o} C while atmospheric concentrations of carbon dioxide, methane and nitrous oxide were driven far beyond pre-industrial levels by fossil fuel use and agriculture [(1-p.2, 5)(2)]. Putting these facts together suggests that human activity has for the first time impacted our very climate, and this issue has been the subject of tremendous research and debate since the late 1970's. Most recently, the Intergovernmental Panel on Climate Change (IPCC) stated in their fourth assessment report that the existence of global warming is now absolutely certain, and they claim very high (90%) confidence that anthropogenic contributions to greenhouse gases since 1750 have resulted in a net warming force [(1-p.5)].

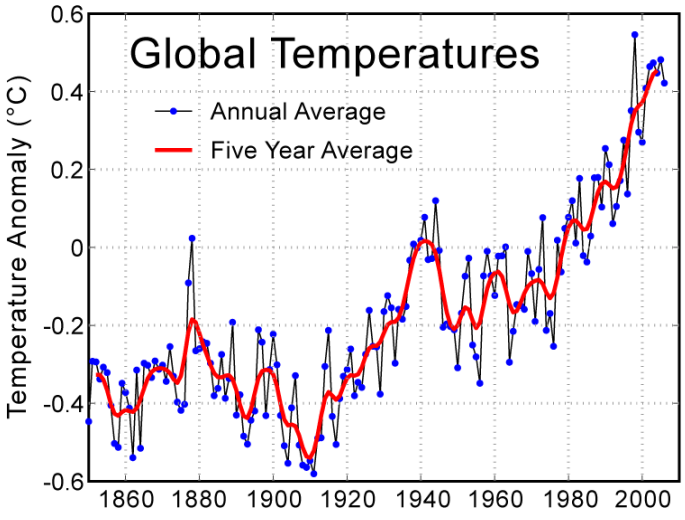


Figure 1.1 – Global temperature record.

Global warming raises serious risks of environmental damage and social and economic impact. Paleoclimatic records show that although our current climate is relatively mild, the planet's climate history is highly dynamic. Information gleaned from ice cores, fossilized tree rings and other indicators tells us that the warming trend observed over the last century is unprecedented in at least the last 1300 years, while the last extended period of comparable warmth (about 125,000 years ago) likely produced 4 to 6 meters of sea level rise [(1-p.10)]. Climate processes have large inertia and the extra load of anthropogenic warming on our climate could trigger climate feedback mechanisms. It is estimated that even if greenhouse gas production from fossil fuel use was halted at current levels the impact of anthropogenic warming would still span centuries. Recently, record setting heat waves across North America and Europe in 2006 and Asia in 2007 resulted in hundreds of deaths, demonstrating the difficulty of reacting en masse to sudden temperature change. In 2005, Hurricane Katrina devastated Louisiana and Mississippi and became the costliest hurricane in US history at a total estimated economic toll of over $150 billion [(3)]. These examples are relatively minor in comparison to the scale of global warming. Current climate models predict that in the next century the global temperature may rise by anywhere from 1.1 to 6.4 ^{o} C [(1 p.13)], and under sustained warming the recession of Earth's vast ice regions would cause between 0.34-m and 1.4-m rise in global sea levels over the 21st century [(1 p.13)]. Studies of land inundation show that a 1-m sea level rise by 2100 would eradicate anywhere between 25% and 80% of the United States coastal

wetlands alone [4]. All of these factors urge us to address the global warming problem to protect the future of our planet.

The foremost variable of concern in global warming is the sea level change which is tied to the ice mass balance. Because land ice helps to insulate the earth by reflecting solar energy, ice mass balance plays a role in the warming feedback process. These feedback processes also contribute to uncertainty about the long-term estimates of land loss and damage. Notwithstanding the IPCC conclusions, there is still disagreement and uncertainty among various findings on the current global mass balance. For instance, Rignot and Kanagaratnam reported escalating mass losses of -91 km^3/yr in 1996 and -224 km^3/yr in 2005, citing the acceleration of several outlet glaciers [5].

Table 1.1 – Reported mass loss of the Greenland ice sheet.

Year	Mass Balance (km^3/year)
1996	-91 ± 31
2000	-138 ± 31
2005	-224 ± 41

However, Zwally et al. reported a current gain of +0.05 ± 0.03 mm per year, reasoning that higher accumulation rates are a natural response to increasing temperatures and precipitation in a warming climate [6]. The IPCC fourth assessment report concludes that Greenland and Antarctic ice sheets have in fact contributed to the observed sea level rise of 1.8 mm/year between 1961 and 2003 and 3.1 mm/year

during the 1993 to 2003 decade. However, more data is still needed in many regions. The mass balance is estimated from ice height and depth data fed into physical models which are used to estimate the losses from outflow towards the oceans and the gains due to surface accumulation. These processes are now much better understood thanks to increased geographical coverage and refinement of analysis methods and uncertainties. However, the estimate of projected sea level rise before 2100 still varies $\pm 0.53m$. More dense and accurate measurement of the state of the planet's ice reserves is needed to assemble an accurate survey of global warming.

Collecting ice mass balance data poses a major challenge due to the harsh conditions and vast, remote expanses of the major ice sheets. The University of Kansas has lead ice elevation and depth measurement missions over Greenland and Antarctica and continues to develop manned and unmanned measurement systems [7][8][9]. The emphasis of research is to increase the sensitivity and efficiency of the fielded radar and lidar equipment. The proposed compact, low power lidar system would enable more dense and accurate measurements of ice sheet altitude to be made from unmanned aerial vehicles (UAVs) and satellites than previous systems, significantly benefiting the climate research objective.

IMPROVED LIDAR DETECTION CONCEPT

This work introduces a novel lidar design that improves upon previous designs in simplicity and sensitivity of the coherent receiver. The newly proposed lidar concept introduces a simplified method of coherent detection based on a lidar system

previously developed and demonstrated at KU. The novelty of the design rests in simplifying the linear FM down conversion process while increasing the sensitivity of the system by eliminating the need for beat signal recovery via RF hardware. This is achieved by performing the RF down conversion mixing using a modulated optical local oscillator. This simplified process provides a significant additional advantage by greatly reducing the bandwidth requirement of the receiver photo detector. This feature reveals exciting new possibilities for improvements to range resolution and the capturing and processing of received light.

The following chapters present the relevant background, introduce the improved receiver concept and document the successful field demonstration of this new lidar system. The background discussions include overviews of the relevant topics and also highlight some interesting related concepts to lead up to the analysis of the proposed lidar. The complete analysis of the new lidar concept follows in Chapter 4, and the hardware prototype and results are presented thereafter. Finally the conclusions and future direction of work are presented as well as references and an appendix of materials that includes a lidar simulation project.

CHAPTER 2 – RADAR AND LIDAR TOPICS

Radar is about measuring distances, position and velocity with radio waves. Every radar works by radiating electromagnetic energy and measuring the resulting response of its environment. Radars can be used to measure line of sight (e.g. altimetry), 2D images (e.g. synthetic aperture radar) and volumetric data (e.g. atmospheric gas concentrations). Lidar (light detection and ranging) is simply radar operating in the optical wavelength region of the EM spectrum. The optical power is launched via telescope, and lidars generally employ free space and fiber optic components instead of RF waveguides. The much larger center frequency compared to radar eliminates the large-wavelength diffraction issues associated with RF, such as broad radiating angle (low antenna gain) and sidelobes. At optical wavelengths it is possible to illuminate a small and well defined area so that artifacts such as radar clutter are inherently prevented. On the other hand, micron wavelengths introduce diffraction and coherence issues in lidar systems.

2.1 REVIEW OF EXISTING LIDAR SYSTEMS

High altitude lidar altimeters can be differentiated into short pulse and pulse compressed types. Short pulse systems use high peak power transmitters such as Q-switched lasers. The advantage of short pulse systems is simple transmit and receive architecture, while the disadvantage is the need for very short pulses with very high peak power to provide adequate SNR and range accuracy. The lifetime of short pulse

systems is limited because the laser source is eventually damaged by the very high internal optical intensity.

Pulse compression relieves the high peak power requirement by processing the pulses so that the range accuracy does not depend on the pulse duration. In general this is accomplished by modulating the pulse in some way (e.g. amplitude, frequency or phase modulation) to increase its bandwidth. The resulting pulse waveform is broad both in time and frequency domains, giving pulse compression an advantage over simple pulsed-carrier systems in their ability to resolve target range for a given peak power. However, compression systems usually require extra hardware and more complex transmitter and receiver design.

Some lidar systems use a detector that responds to the optical power level of the received light, referred to as direct detection. There is another class of lidar which uses coherent detection, based on the mixing of two light sources by interferometry. Optical mixing enables amplification of incoming light by the use of a strong local oscillator mixed with the incoming field. This process is feasible thanks to the square law response of the photodiode. The advantage of one detection method over the other for a given application depends mainly on the sensitivity requirement and the acceptable cost and complexity of the system.

There are many research and commercial grade lidars in existence, each with its own specializations. A good reference for comparison of recent commercial and research grade lidars is found in Appendix A of a 2004 NIST report [(10)]. Our aim is to

develop a highly specialized lidar tailored for high altitude ice sheet altimetry. This system can be compared with two high altitude lidar altimeters currently in use, GLAS and the Mars Global Surveyor MOLE. These systems utilize high power Q-switched lasers without pulse compression, while the proposed lidar uses a low power diode laser and incorporates pulse compression techniques borrowed from RF radar technology.

Geoscience Laser Altimeter System (GLAS)

The Geoscience Laser Altimeter System (GLAS) is a pulsed-carrier lidar that was launched in 2003 on ICESat [(11)]. Operating at 1064 nm wavelength, GLAS transmits 75 mJ pulses with 5 ns pulse duration, translating to 15 MW peak power. GLAS uses a Q-switched laser which is a specialized laser capable of very short and very high peak power pulses. The lifetime of Q-switched lasers is limited due to intracavity damage from high light intensity and the system life span is limited to 5 years.

Table 2.1 – GLAS parameters.

GLAS Parameter	*Value*
Wavelength	1064 nm
Pulse Energy	75 mJ
Pulse Duration	5 ns
Peak Pulse Power	15 MW
Pulse Repetition Frequency	40 Hz
Operating Altitude	600 km
Range Accuracy	10 cm
Range Resolution	75 cm
System Life Span	3~5 years

Mars Global Surveyor: MOLA

MOLA, the Mars Orbital Laser Altimeter uses similar architecture as GLAS and is in use on the Mars Global Surveyor [12] [13]. Like GLAS, MOLA uses a diode pumped, Q-switched Nd:YAG laser. Both of the mentioned lidars have limited system life span due to the use of Q-switched lasers. The life span can be greatly increased by using lower peak power and incorporating pulse compression to compensate the range resolution.

Table 2.2 – MOLA parameters.

MOLA Parameter	*Value*
Wavelength	1.06 μm
Pulse Energy	45 mJ
Pulse Duration	20 to 540 ns
Peak Pulse Power	83 W to 2.25 kW
Pulse Repetition Frequency	10 Hz
Operating Altitude	~100 km
Range Accuracy	2 m
Range Resolution	3 m to 81 m

University of Kansas 1319 nm Hybrid RF/Laser Radar

To overcome the limitations of lidars based on high peak power Q-switched lasers, KU has developed hybrid RF/ laser radar using linear FM pulse compression [(14)]. The system uses linear FM pulse compression and the pulse compression operation is done in RF hardware after the photo detector.

Table 2.3 – KU lidar parameters.

KU Lidar Parameter	*Value*
Wavelength	1319 nm
Pulse Duration	200 μs
Transmitted Bandwidth	260 MHz
Pulse Repetition Frequency	1000 Hz
Range Accuracy	10 cm

The proposed self chirped lidar we have recently developed uses some of the same components as the KU hybrid RF lidar and has similar characteristics as shown in **Table 4**. The self chirped lidar system requires fewer parts and has demonstrated better sensitivity under comparable conditions.

Table 2.4 – Comparison of KU lidars.

Parameter	***KU Hybrid RF Lidar***	***KU Self Chirped Lidar***
Wavelength	1319 nm	1319 nm
Pulse Duration	200 μs	40 μs
Transmitted Bandwidth	260 MHz	400 MHz
Pulse Repetition Frequency	1000 Hz	9400 Hz
Range Accuracy	10 cm	10 cm

2.2 RADAR AND LIDAR PRINCIPLES

Radar and lidar principles encompass a broad range of topics such as Physics, Quantum and Electromagnetic Theory and Information Theory. This introduction will give an overview of the relevant principles encountered in designing the proposed lidar system.

Range Resolution

A pulsed-carrier radar transmits on-off pulses of a single center frequency carrier. In free space this pulse occupies some distance related to the pulse duration τ. The minimum radial distance between two targets such that they can be distinguished by the radar is referred to as the range resolution, which for such radars is [15]

$$\Delta R = c\tau / 2 \quad \text{m} \quad [2.1.1]$$

This results directly from the requirement that the echoes from two targets be separated in time by greater than the pulse width*. This implies that very short pulse duration allows fine range resolution while having a large signal bandwidth. On the other hand, infinitely long pulse duration allows no resolution of targets at all while its spectrum has infinitely small bandwidth. The general principle is that increasing the bandwidth improves the range resolution, a phenomenon of the uncertainty between time and frequency domains.

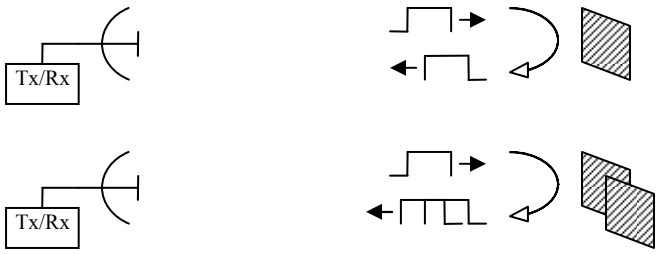


Figure 2.1 – Illustration of simple pulse overlap from multiple targets.

Pulse Repetition Frequency

A monostatic system such as GLAS requires balancing the amount of time used transmitting and receiving with its single telescope. Any signal that returns during ongoing transmission cannot be received because the transmit power will cause high background noise and greatly exceed the dynamic range of the receiver. The pulse repetition frequency (PRF) and pulse duration τ must be small enough to accommodate listening time over the range of possible target echoes, yet high enough to provide the minimum average transmitted power for the required SNR. Considering that GLAS' orbit altitude is 600 km, the utilization of available transmit time is very low. A high percentage of time (99.99998 %) is spent in receiving mode even though the range window is relatively small, and much of that time could otherwise be used for transmitting and improving SNR. A pulse compression system can utilize more time by transmitting longer pulses.

---

*A good illustration is found at http://www.radartutorial.eu/01.basics/rb18.en.html

Uncertainty Principle for Time and Frequency

The conjugate relationship between the time-domain and frequency-domain properties of a signal is an essential part of understanding radar and lidar operation. For any signal, the exactness of its position in time is inversely proportional to the exactness of its frequency. In other words, a signal with a very narrow bandwidth must have a very long time span and a signal with a very small time span must have a very large bandwidth. This is represented by the Fourier relation between time and frequency domains. For stochastic signals the relationship between the variances of the distributions in time (σ_t^2) and frequency (σ_F^2) obeys

$$\sigma_t^2 \sigma_F^2 \geq 1 \quad [2.2.1]$$

where the limiting case of unity is given by a signal with a Gaussian distribution in both time and frequency*. This is an important basis for understanding constraints that arise in the design of radar systems, e.g. Range Resolution, Doppler Resolution and Pulse Duration.

Range Accuracy

Range accuracy δR describes the ability to correctly measure the position of a solitary target. Range accuracy equations are approximate because the returned signal forms a statistical distribution depending on the signal bandwidth and SNR. **(15)**

$$\delta R \cong \frac{c}{2B\sqrt{2 \times SNR}} \quad \text{m} \qquad [2.2.2]$$

The range accuracy for the proposed lidar system as well as GLAS is 10 cm. This is usually calculated given a minimum 10 dB SNR.

Radar Cross Section

Calculating the amount of energy getting into the receiver first requires knowledge of the target's radar cross section (RCS). Target surfaces can range from specular reflectors (e.g. mirrors or water) to diffuse reflectors (e.g. paper) depending on the surface size and roughness compared to the wavelength. If the target is snow, the reflection profile can be approximated as Lambertian. A Lambertian surface appears diffuse and reradiates incident power from any direction equally in all directions like ordinary paper under a lamp. Assuming that a relatively small laser spot is illuminated on the snow, it can be thought of as an isotropic radiating point source and this allows application of the spherical propagation loss model.

Propagation Loss and Photon Limited Range

Even under ideal circumstances, the energy recovered by the receiver is extremely small due to substantial spherical spreading loss over large distances. By energy conservation, after traveling back to the receiver at distance R the power density is decreased in proportion to the surface area of the spherical wave front $4\pi R^2$. The received energy is often so small that it is on the order of the photon energy $h\nu$ and

*This relationship between complementary variables (known as Fourier Uncertainty Principle) also applies to other applications of conjugate variables. For instance it is essentially restated in the famous Heisenberg Uncertainty Principle. A mathematical explanation of the connection between the two concepts can be found at http://www.mathpages.com/home/kmath488/kmath488.htm.

thus can be counted in terms of the number of photons received. Even in the absence of all noise, the fact that energy is recovered in discrete amounts (photons) bars the possibility of achieving limitless detection. The signal energy recovered by the receiver can be calculated as

$$E_s = \frac{\rho P_T A_r \tau}{\pi R^2} \quad \text{J} \qquad [2.2.3]$$

and the average number of received photons is **(16-p.18)**

$$\overline{n_s} = \frac{E_s}{h\upsilon} \qquad [2.2.4]$$

where

E_s = received signal energy,

ρ = reflection coefficient,

P_T = transmitted power,

A_r = receiving area,

τ = pulse width or resolution time of measurement,

R = target distance.

h = Planck's constant,

ν = operating frequency

Assuming that the photo detector in the receiver has a quantum efficiency η the number of emitted photoelectrons is $\eta\overline{n_s}$, and to be detected this quantity must exceed some minimum number K within the observation time. The emission of

photoelectrons follows a Poisson arrival process resulting in probability bounds for the probability, or confidence of detection. As shown in **Table 2.5**, observing one photoelectron emission (K=1) within τ gives 63% confidence, >99% confidence is reached with K=5, and with K>13 the probability of error becomes vanishingly small ($P_e < 2x10^{-6}$) **(16-p.20)**.

Table 2.5 – Probabilities of photon limited detection.

$\bar{n}\eta\tau = (\bar{n}_s\eta)_{\min}$	$\sum_{m_\tau=1}^{\infty} P(m_\tau)$
1	0.632131
2	0.864665
3	0.950213
4	0.981684
5	0.993262
6	0.997521
7	0.999088
8	0.999665
9	0.999877
10	0.999955
11	0.999983
12	0.999994
13	0.999998

These detection probabilities merely account for the quantum nature of the received energy, and we must later include additional noise and loss processes in the analysis of receiver sensitivity. A summary of radar range equations can be found in the literature **(16-p.33)**.

Atmospheric Loss and Distortion

When the air in the path of the laser is turbulent, the refractive index is not uniform across the aperture and the traveling optical wavefront is distorted as different components of the wave experience different propagation constants. The transverse profiles of the wavefront phase and power will become randomly distributed, causing interference when the light is focused back into the fiber. This effect has caused decreased coupling efficiency in our lab and field tests. The air path also attenuates power depending on the density and composition of the atmosphere. One way to account for these effects is to model the received power attenuation with higher orders on the distance, e.g. $\sim R^{-3}$ or $\sim R^{-4}$. Air turbulence is one of the special challenges for lidar compared to radar because the operating wavelength is small compared to the phase fluctuations caused by wind and temperature flux.

Doppler Shift and Resolution

Another effect on wave propagation is the phenomenon of Doppler shift. The Doppler shift of a return signal is proportional to the target's relative longitudinal velocity as

$$f_d = 2v_r / \lambda \quad \text{Hz} \quad [2.2.5]$$

At optical frequencies this can be very large even for small radial velocity v_r and this can be problematic for coherent detection. On the other hand, Doppler is very useful for measuring target velocity.

Returning to the uncertainty relation between time and frequency we find another interesting property. That is, although improved range resolution requires increased bandwidth, improved resolution of Doppler shift requires increased effective pulse duration. This is can be understood by considering that a target's position can be established with a quick glance but it takes longer observation time to determine its velocity [(15-p.19)]. In fact there is no theoretical upper limit to the combined range and Doppler resolution; it is only limited by the maximum achievable SNR [(17-p363) (15-p134)].

2.3 PULSE COMPRESSION

The purpose of pulse compression is to enhance range resolution well below the limit imposed by the time overlap of return pulses and to reduce the peak transmit power. Ordinary pulse radar requires very short duration pulses to achieve high range resolution, and short pulses require high peak transmit power to impart enough energy in the signal to achieve the needed SNR. Fortunately, it is not the duration of the pulse but the bandwidth that determines the ability to resolve range[*]. Pulse compression makes use of this principle by introducing high signal bandwidth in long pulses.

---

*For single frequency radar the only contributor to the signal's bandwidth is the pulse amplitude envelope. The principle that the bandwidth of the signal determines the range resolution is still true because the rectangular on-off pulsing of the sinusoidal carrier widens the signal bandwidth e.g. into a sinc function.

A Basic Pulse Compression Method

The first implementation of pulse compression used dispersive or nonlinear phase components. In a classic pulse compression system the transmitted RF pulse is swept upwards in frequency. The receiver contains a dispersive delay line which imparts higher delay to lower frequencies in linear proportion. Therefore the long pulses that are transmitted became very short when they return as their frequency content is 'bunched' together by the delay line at a narrow point in time. In this fashion the delay line acts as a matched filter, producing the cross correlation of the swept frequency pulse. However, the same effect can be achieved by other means, and various pulse compression methods have been devised.

Linear FM Chirp with Analog Mixing

The presented lidar system uses the Linear FM Chirp method with analog mixing. This method replaces the rigor and complexity of digitizing and cross correlating the received waveform by an equivalent process that mixes analog signals in the time domain. Linear FM pulse compression can be performed in CW or pulse mode with combinations of upward and downward frequency sweep. In this work a pulsed version was implemented using only upward frequency sweeping.

The linear FM pulse compression system begins with a swept-frequency RF (chirp) signal. The frequency of this chirp increases linearly from f_1 to f_2 over the pulse duration τ. This chirp signal is transmitted via modulation of a single frequency carrier, forming a modulated pulse also of duration τ. As this pulse propagates in free

space and reflects from a target back to the receiver, a round-trip propagation delay Δt is induced relative to the original reference chirp. At the receiver this echo signal is recovered from the carrier via the receiver and converted back to a baseband RF signal. The reference and echo chirps are then mixed together. **Figure 2.2** illustrates the constant offset of frequencies that occurs between the reference and echo signals. By taking the product of the reference and echo chirps, a steady beat-frequency, f_R is produced that is proportional to Δt. The target range, R is hence related to f_R as

$$f_R = (f_2 - f_1)\Delta t / \tau = \frac{2BR}{c\tau} \qquad [2.3.1]$$

In practice, R is determined by analyzing the PSD of the beat-frequency (dechirped) signal to locate the frequency of peak power corresponding to a target.

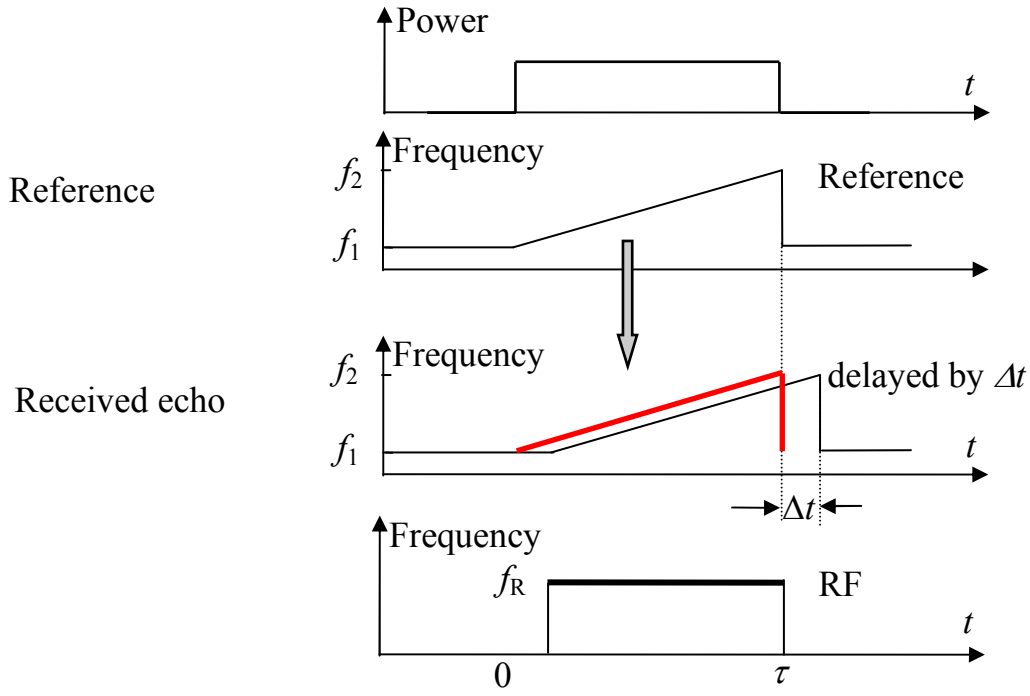


Figure 2.2 – Illustration of pulse compression using linear FM chirp.

Pseudo Random (PN) Coding Pulse Compression

Another common pulse compression method makes use of pseudo random code (PN code) modulation. This method does not use the beating frequency of two pulses. Instead, the range is found by correlation. First, a maximum-length (ML) code is generated using a shift register. The autocorrelation if this code is ideally 1 at the origin and zero at all other points. This ML code is used to modulate the carrier which is transmitted to the target and its echo received with delay Δt, as illustrated in **Figure 2.3**. Then, the cross correlation is found between the received pulse and the reference. The location of the peak of this cross correlation function indicates the amount of

relative delay between the reference and the echo. The range is then calculated from this delay.

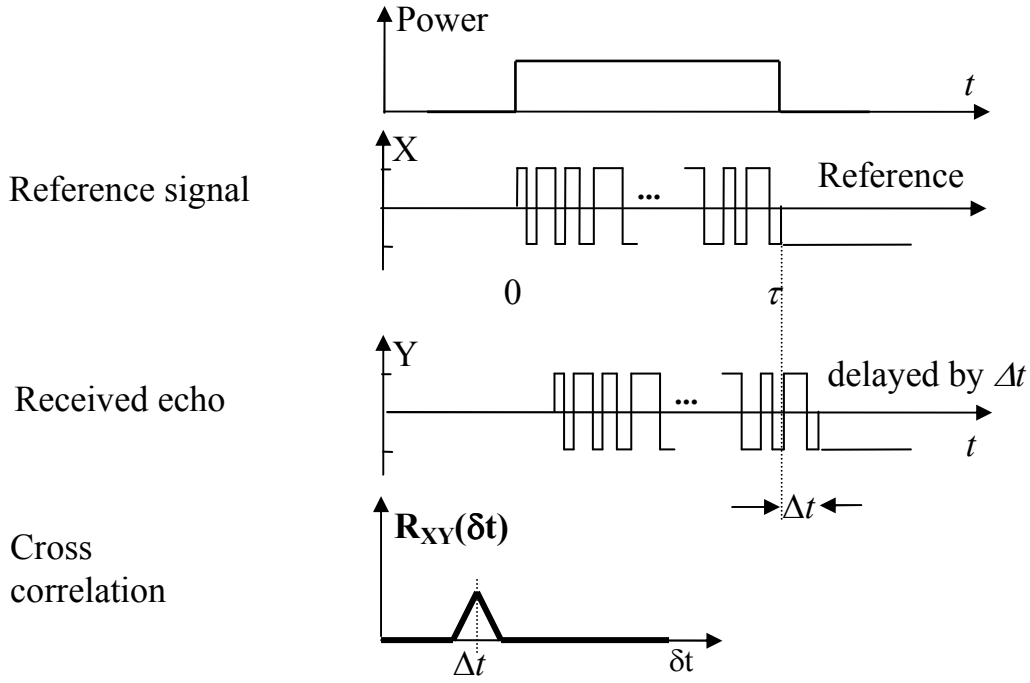


Figure 2.3 – PN code pulse compression.

The bandwidth of the pulse in this case corresponds with the symbol rate. The symbol rate in turn affects the width of the autocorrelation function, thereby determining the range resolution and accuracy.

2.4 Optical Components

The analysis of lidars generally requires modeling light under two conditions. The first case is the single mode transmission of light within an optical fiber, and the second is free space propagation and interaction with the environment. Starting with a steady state expression for laser light propagation in a single mode fiber under basic assumptions, the modulation, detection and noise processes will be described. Free space propagation is discussed later as it relates to the challenges with power coupling efficiency.

E Field Equation for Linear Polarized Plane Waves

The solution of Maxwell's equations for free space assumes the form of a uniform plane wave propagating in the direction orthogonal to the E and H fields. In optical fiber, the propagating light is bound by the cladding and cylindrical boundary conditions must be applied to solve the Maxwell equations. While this solution is rather complex, the signal carried by light in single mode conditions is sufficiently characterized by its amplitude and phase using phasor notation. As light travels along a fiber span, its power decays exponentially due to material absorption. However, this attenuation typically is less than 0.5 dB/km in the 1310 nm region so it can be neglected in the analysis. Given these generalizations, the electric and magnetic fields can be expressed in phasor notation as:

$$E_s(t,z) = E_0 e^{j(\omega t - \beta z)} \hat{a}_x \quad [2.4.1]$$

$$H_s(t,z) = \frac{E_0}{\eta} e^{j(\omega t - \beta z)} \hat{a}_y \quad [2.4.2]$$

where η is the intrinsic impedance given by $\eta = \sqrt{\mu/\varepsilon}$ and β is the propagation constant which is $\beta = \varpi\sqrt{\mu\varepsilon}$ for a lossless medium. The term E_0 is a real value representing the electric field amplitude in units of Volts/meter, and the measurable electric field is found by taking the real part of the complex expression. For the idealized case of lossless dielectric the E and H field amplitudes are proportional by the intrinsic impedance, therefore it is only necessary to keep track of the E fields in the analysis. For a comparison of real and phasor notation methods, see Appendix B.

Optical Fiber

The lidar system uses standard optical fibers to connect the optical components. The fiber used is cylindrical step-index single mode fiber (SMF) which comes in standard (isotropic) and polarization maintaining (PM) types.

An ordinary step index fiber is a single strand of glass that has a small central core within an outer cladding as shown in **Figure 2.4**.

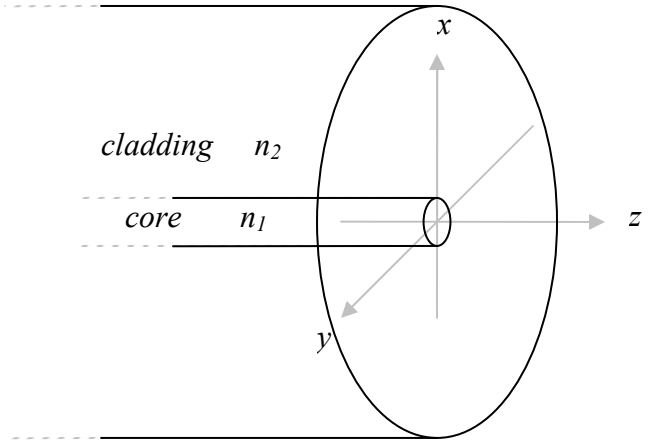


Figure 2.4 – Diagram of optical fiber.

The core and cladding have different chemical compositions which affect the materials' refractive index. The core has refractive index n_1 (typically =1.48) while the cladding has a slightly lower refractive index $n_2<n_1$. This arrangement allows for total internal reflection within the core and thus the potential for transmission with very low loss (although the explanation of guided mode propagation goes beyond ray optics). The reason that fiber communications extensively uses the 1310 and 1550 nm bands is because SiO_2-based glass exhibits extremely low attenuation regions at these wavelengths.

Using Snell's law and given the fiber refractive indices and core radius we can determine $\theta_{0,\max}$, the maximum acceptance angle for total internal reflection. This can then be used to define the numerical aperture *NA* which relates to the light acceptance capability of the fiber **(19-p.41)**

$$NA = n\sin\theta_{0,\max} = \left(n_1^2 - n_2^2\right)^{1/2} \approx n_1\sqrt{2\Delta} \qquad [2.4.3]$$

Typical values of *NA* for single mode fiber range from 0.19 to 0.30.

Solving Maxwell's equations for the cylindrical boundary conditions shows that the fiber can support guided modes as well as radiation modes and leaky modes. Radiation modes are those that are not guided because they are outside the acceptance angle of the fiber. Those modes radiate away from the core into the cladding; however they can be coupled back into the guided modes because both modes intersect slightly across the core-cladding boundary. This can cause a loss of power of the guided mode, so a lossy coating is usually applied to the outside of the cladding to absorb the radiated mode power. A third type of mode is the leaky mode which is partially confined to the core but whose power is dissipated out of the core as it travels along the fiber **(19-p.44, 45, 46)**.

The number of supported guided modes for a typical step index fiber* is approximately related to the *V* number (normalized frequency) which is defined as **(19-p.46)**

$$V = \frac{2\pi a}{\lambda} NA \qquad [2.4.4]$$

where *a* is the radius. The single mode cutoff for propagating modes requires that V<2.405 which is the first root of the Bessel function **(19-p.46)**. The mode order is equal to the number of zero crossings of the field across the waveguide, and TE_0 is the lowest order guided mode**. This is the one mode that is useful for communications.

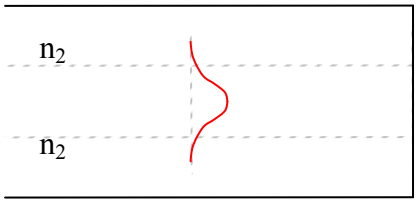


Figure 2.5 – TE_0 guided mode.

In principle, a perfectly symmetric single mode fiber does not cause polarization mode dispersion. However, in practice slight birefringence is randomly introduced along the fiber span by external stresses, temperature flux and imperfections (e.g. asymmetry) in manufacture. External bending stresses the fiber, creating birefringence along arbitrary axes due to the photoelastic, or Piezo-optical effect. The components of the traveling E field parallel and orthogonal to the axis of birefringence will experience different propagation constants, causing retardation of one component relative to the other. Since the birefringence is uncontrollable and varies randomly along the length of the fiber, a transmitted field can become significantly distorted and the final polarization will have a random distribution.

---

* it must be assumed that the index difference is small.

**An illustration of waveguide modes can be found at http://www.falstad.com/embox/guide.html.

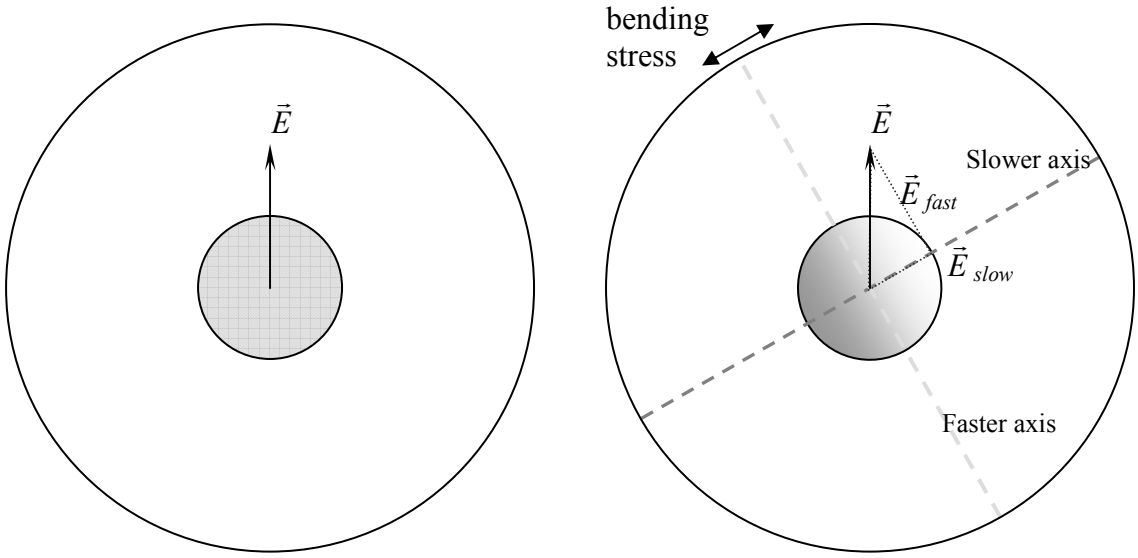


Figure 2.6 – Ideal fiber (left) and with stress-induced birefringence (right).

Polarization maintaining (PM) fiber eliminates this problem by introducing strong birefringence along one axis of the fiber. This is done by including tensioning rods alongside the core that create stress across one dimension of the fiber. Two common varieties are panda and bowtie named after the shape of the tension rods.

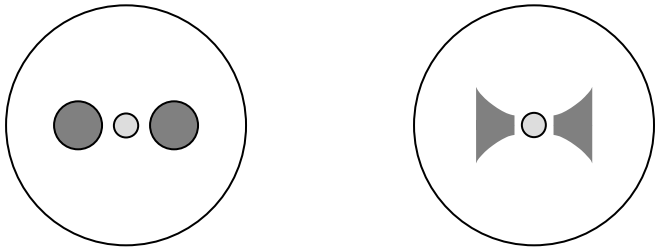


Figure 2.7 – PM fiber, panda (left) and bow-tie (right).

PM fiber strongly differentiates the x and y axis refractive indices, virtually eliminating coupling between them. This can be understood by considering that the wavelength of the same input field in the two axes is different, and so there is a loss of coherence between the E fields along the two axes. Another explanation is that the beat length between the two wavelengths is small enough that the net exchange of power is effectively canceled out.

Optical Field Distribution and Power Density

When making performance calculations it is sometimes necessary to know the spatial distribution of the electric field within the optical fiber. One way to model the radial distribution of the electric field is the Gaussian function given as

$$E_s(r) = E_0 e^{\left(-r^2/W_0^2\right)} \hat{a}_x \qquad [2.4.5]$$

where r is the radial distance from the core axis, E_0 is the field at zero and W_0 is the width of the electric field distribution **(19-p.63)**. This expression underscores the fact that some of the light propagates outside of the fiber core itself. However, for most purposes we can simply refer to the total E field, whose center is along the fiber's central axis. Therefore we can sidestep this high level of detail by allowing the fiber to have an effective area that will give the approximate relationship between power density and electric field **(19-p.491)**. For a standard single mode fiber the effective area

may be approximately 80 μm^2. Assuming an effective area A_{eff}, we can relate the measured power P_{meas} and the time averaged power density $P_{z,\,avg}$ as

$$\mathrm{P}_{meas} = \mathrm{P}_{z,avg} \cdot A_{eff} \quad \mathrm{W} \qquad [2.4.6]$$

The optical power is given by the Poynting vector, which for the lossless case results in

$$\mathrm{P}_{z,avg} = \frac{1}{2}\mathrm{Re}\left\{E_s \times H_s^*\right\} = \frac{1}{2}\frac{E_0^2}{\eta}\hat{a}_z \quad \mathrm{W/m^2} \quad \text{(lossless)} \quad [2.4.7]$$

$P_{z,avg}$ is the time-averaged optical power density in the z direction. Since lab bench measurements are always given in terms of total measured power, we need a relationship between the instantaneous electric field and the optical power read from a power meter. Rearranging the above expressions yields the following relationships:

$$E_s(t,z) = \sqrt{2\eta \cdot \frac{\mathrm{P}_{meas}}{A_{eff}}}\, e^{j(\omega t - \beta z)}\hat{a}_x \ \mathrm{V/m} \quad \Leftrightarrow \quad \mathrm{P}_{meas} = A_{eff}\frac{\left|E_s(t,z)\right|^2}{2\eta}\hat{a}_z \quad \mathrm{W}$$

[2.4.8]

Electro-Optic Modulation

Up to this point we have defined the steady state electric field to serve as a model of the optical power propagating from the laser source. The next step is to apply intensity modulation to this CW carrier, something that can be performed using an Electro-Optic Modulator (EOM). An EOM uses an arrangement of optical waveguide paths in the configuration of a Mach-Zehnder interferometer as shown in **Figure 2.8**.

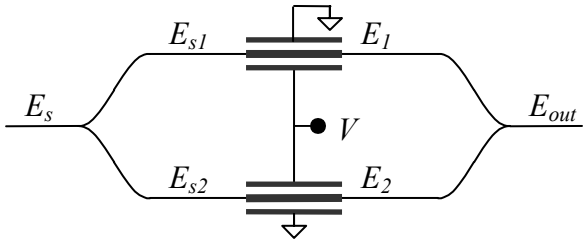


Figure 2.8 – EOM diagram.

Within the EOM, the input optical signal is first split into two paths. Assuming that an ideal 3dB coupler with zero insertion loss and excess loss is used, the electric fields in the two paths are

$$E_{s1}(t,z) = \frac{E_0}{\sqrt{2}} e^{j(\omega t - \beta z)} \hat{a}_x \qquad [2.4.9\text{ a}]$$

$$E_{s2}(t,z) = \frac{E_0}{\sqrt{2}} e^{j(\omega t - \beta z)} \hat{a}_x \qquad [2.4.9\text{ b}]$$

where E_0 is the magnitude of the electric field at the input. The EOM accepts a DC bias voltage and an RF modulation voltage signal on two separate inputs, and these input voltages are applied across the active region as illustrated in **Figure 2.8** so as to alter the refractive index of the material. This works on the principle of the linear electro-optic effect, which means that an externally applied E field produces a phase change $\Delta\phi$ of the transmitted electric fields as

$$\Delta\phi = \frac{\pi n_0^3 r V}{\lambda} \cdot \frac{l}{d} \quad \text{rad} \qquad [2.4.10]$$

where n_0 is the unperturbed index of refraction, V is the applied voltage, l is the interaction length, d is the electrode separation and λ is the free space wavelength [20]. The term r is an appropriately chosen element from a third-rank tensor that describes the effect of an applied electric field on the index of refraction within the material [20]. In other words, r varies depending on the spatial orientation of three factors: the crystal lattice, the occupying E field and the applied voltage. The change induced in the refractive index of the material alters the optical wavelength and thus varies the phase at the output relative to the input. For our particular EOM, the same input voltage is applied with opposite polarity across both paths, so the net phase difference is doubled. Denoting the phase changes as ϕ_1 and ϕ_2, the two E fields after the active region can be written as

$$E_1(t,z) = \frac{E_0}{\sqrt{2}} e^{j(\omega t - \beta z - \phi_1)} \hat{a}_x \quad \text{V/m} \qquad [2.4.11\text{ a}]$$

$$E_2(t,z) = \frac{E_0}{\sqrt{2}} e^{j(\omega t - \beta z - \phi_2)} \hat{a}_x \quad \text{V/m} \qquad [2.4.11\text{ b}]$$

And we can separate the two phases into mean and difference components as

$$\phi_0 = \tfrac{1}{2}(\phi_1 + \phi_2) \quad \text{rad} \qquad [2.4.12\text{ a}]$$

$$\Delta\phi = \tfrac{1}{2}(\phi_1 - \phi_2) \quad \text{rad} \qquad [2.4.12\text{ b}]$$

so the fields can be written as

$$E_1(t,z) = \frac{E_0}{\sqrt{2}} e^{j(\omega t - \beta z - (\phi_0 + \Delta\phi))} \hat{a}_x \qquad [2.4.13\text{ a}]$$

$$E_2(t,z) = \frac{E_0}{\sqrt{2}} e^{j\left(\omega t - \beta z - (\phi_0 - \Delta\phi)\right)} \hat{a}_x \qquad [2.4.13\ b]$$

Finally these two signals are merged to produce the output field E_{out}. Assuming that the combiner is another ideal 3dB coupler, the sum of the two electric fields will be

$$E_{out}(t,z) = \frac{1}{\sqrt{2}} E_1(t,z) + \frac{1}{\sqrt{2}} E_2(t,z) \quad \hat{a}_x \qquad [2.4.14\ a]$$

$$E_{out}(t,z) = \frac{E_0}{2} e^{j\left(\omega t - \beta z - (\phi_0 + \Delta\phi)\right)} + \frac{E_0}{2} e^{j\left(\omega t - \beta z - (\phi_0 - \Delta\phi)\right)} \quad \hat{a}_x \qquad [2.4.14\ b]$$

$$E_{out}(t,z) = \frac{E_0}{2} e^{j\left(\omega t - \beta z - \phi_0\right)} \left[e^{j\Delta\phi} + e^{-j\Delta\phi} \right] \quad \hat{a}_x \qquad [2.4.14\ c]$$

Therefore the output of the EOM is

$$E_{out}(t,z) = E_0 \cos(\Delta\phi) e^{j\left(\omega t - \beta z - \phi_0\right)} \quad \hat{a}_x \qquad \text{V/m} \qquad [2.4.15\ a]$$

$$P_{out}(t,z) = \frac{E_0^2}{2\eta} \cos^2(\Delta\phi) \quad \hat{a}_z \qquad \text{W/m}^2 \qquad [2.4.15\ b]$$

and we can note that altering the phase difference $\Delta\phi$ allows for intensity modulation of the signal. At some input voltage $V = V_\pi$ the phase difference $\Delta\phi$ is equal to $\pi/2$ and the output power becomes zero. Thus the EOM response is usually written as

$$E_{out}(t,z) = E_0 \cos\left(\frac{V}{V_\pi}\frac{\pi}{2}\right) e^{j(\omega t - \beta z - \phi_0)} \quad \hat{a}_x \qquad \text{V/m} \qquad [2.4.16\text{ a}]$$

$$P_{out}(t,z) = \frac{E_0^2}{2\eta}\cos^2\left(\frac{V}{V_\pi}\frac{\pi}{2}\right) \quad \hat{a}_z \qquad \text{W/m}^2 \qquad [2.4.16\text{ b}]$$

the transfer characteristic of the EOM is shown in **Figure 2.9**. Note that it is nonlinear and periodic.

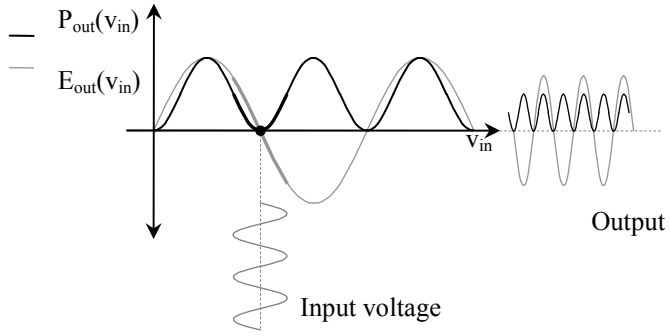


Figure 2.9 – EOM transfer characteristic.

We wish to obtain an approximately linear output response, and this can be achieved in two ways. First, if the output signal swing is kept within the linear region of the output power as in **Figure 2.10 (left)**, then the output is approximately linear. On the other hand, if the bias is changed so that the output signal swing encompasses the zero crossing and surrounding linear region of the output E field as in **Figure 2.10 (right)**, then the output frequency is doubled. However, this 2nd order output can be a useful frequency doubling modulation mode. This technique has been applied in

communications in the duobinary modulation scheme and is also used in the self chirped lidar system.

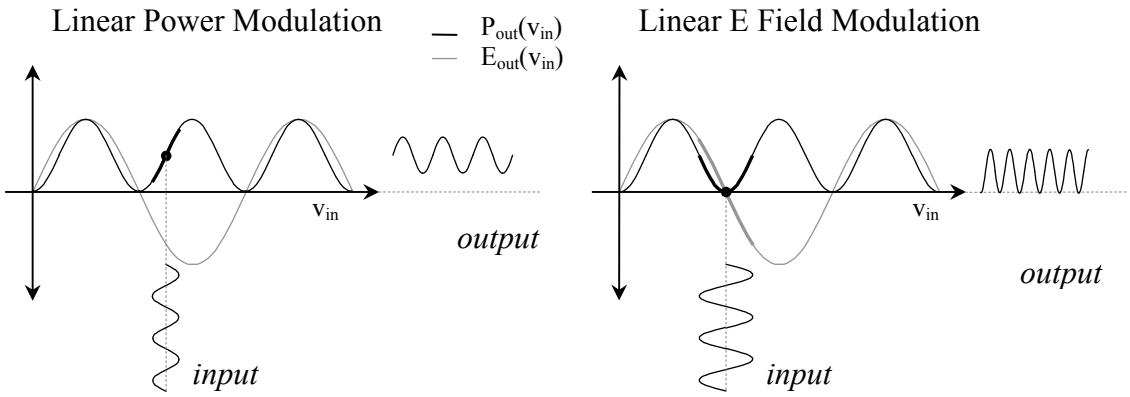


Figure 2.10 – EOM transfer function for linear power and E field modulation.

Optical Amplifiers

The light output from the EOM needs to be amplified before it can be transmitted from the lidar. The power can be increased using an optical amplifier, which can be classified as semiconductor optical amplifiers (SOA) or doped-fiber based amplifiers (DFA). Compared to SOAs, DFAs are insensitive to polarization and have very stable gain due to a slow response time, and thus low crosstalk in multi-wavelength applications. For the 1550 nm band, SiO_2 based fibers are used and for 1310 nm operation fluoride-based fibers are used. For this work an Erbium Doped Fiber Amplifier (EDFA) was used.

All optical amplifiers exhibit Amplified Spontaneous Emission noise (ASE). This noise stems from charge carriers that are spontaneously emitted within the energized

fiber, some of which fall within the fiber's numerical aperture and are thus incorporated into the output. The spontaneously generated light is also subject to the same gain process as the signal, therefore a low input signal will have high relative noise at the output. This makes optical amplifiers unsuitable for amplifying the low received power.

PHOTODETECTORS

After the lidar telescope captures the received light, a photodetector is used to convert the light into an amplified electrical signal. While several varieties of photodetectors exist, the lidar system uses a photodiode because of its small size, high performance, durability and low supply voltage, as well as its widespread availability from the communications industry.

Physics of the Photodiode

The photodiode is a semiconductor device that converts incident light into electrical current. The injected photons excite the transmission of charges across the photodiode's active region, producing a linear response of output electrical current to input optical power.

The most common optical receiver is the PIN diode, so named because it is layered with p-type, lightly n-doped intrinsic and n-type regions.

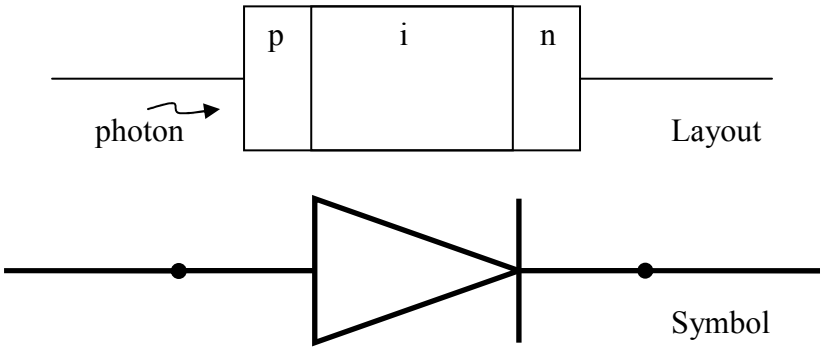


Figure 2.11 – PIN photodiode.

While an ordinary diode is a p-n junction, the addition of an intrinsic layer between the p and n layers allows incident photons to excite electron-hole pairs which are then accelerated across the region by an applied reverse bias voltage. The quantum efficiency η of the pin photodiode represents the percentage of electron-hole pairs generated per incident photon and depends on the material band gap, operating wavelength and the dimensions and doping of the p, i, n regions. Typical values of η are 30% to 95%. The responsivity $\Re$ gives the amount of photocurrent generated per incident optical power as

$$\Re = \frac{I_p}{P_0} = \frac{\eta q}{h\nu} \qquad \text{A/W} \qquad [2.4.17]$$

The fact that the number of stimulated charges is proportional to the incident number of photons is what gives photodiodes their useful square-law conversion property between optical input power and electrical output power.

Another common detector photodiode is the avalanche photodiode (APD). In an APD the stimulated electrons or holes are subjected to a high electric field, causing them to bombard and ionize bound valence electrons within the region. The ionized valence electrons in turn are also accelerated and can continue to ionize others. This multiplication mechanism gives rise to the APD current gain factor *M*, which is included in calculating the APD responsivity as

$$\Re_{APD} = M\,\Re = \frac{\eta q}{h\nu} M \quad \text{A/W} \qquad [2.4.18]$$

The APD current gain mechanism also gives rise to an additional noise factor $F(M) \approx M^x$ where x depends on the device parameters with $0 \le x \le 1$.

Response Time vs. Responsivity: Bandwidth - Efficiency Tradeoff

There is a tradeoff between bandwidth and quantum efficiency in photodiodes, which is why reducing the needed photodiode bandwidth in the new lidar concept is a key improvement over other systems. Bandwidth is characterized by the response time, or the time it takes for the output current to swing in response to a step input to within a given fraction of its final value. This characteristic is fundamentally limited by the transit time t_d required for stimulated charges to traverse the depletion region,

$$t_d = \frac{w}{v_d} \qquad [2.26]$$

Where w is the depletion layer width and v_d is the carrier drift velocity. Unfortunately, high bandwidth typically comes at the cost of decreased quantum efficiency. This is because increasing the depletion layer thickness to absorb more light also increases the transit time and thus reduces the bandwidth. This limits the amount of bandwidth that can be used in typical lidar, but fortunately the proposed lidar system's low RF bandwidth requirement eliminates this problem.

2.5 Optical Receiver

The receiver SNR depends on the signal gain and the various noise sources added to the signal. The photodetector circuit is shown below in **Figure 2.12** and the noise sources are listed in **Table 2.6**. In addition to various noise sources within the

photodetector, the incoming light itself can contain background radiation that adds noise to the system.

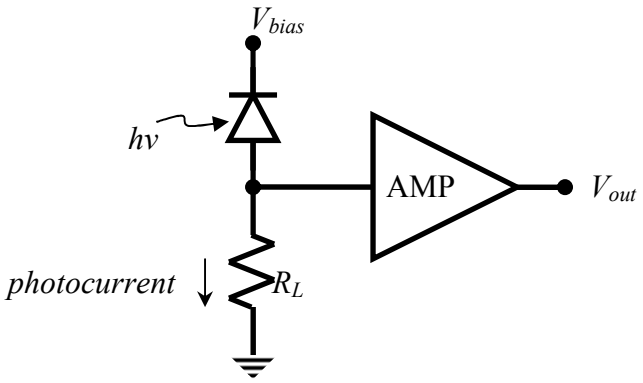


Figure 2.12 –Example photodetector circuit.

Table 2.6 – Receiver noise sources.

Input Light	**Photodiode**	$\boldsymbol{R_L}$	**AMP**
background radiation	quantum noise APD excess noise dark current	thermal noise (Noise temp.)	thermal noise (Noise figure)

Thermal Noise

Thermal (Johnson-Nyquist) noise refers to the current caused by the random motion of electrons due to their thermal energy. While every resistive element contributes a thermal noise current, the amplifier input impedance is typically much higher than the load resistance R_L*. The mean-square thermal noise for the PIN-based photodetector is

$$\left\langle i_T^2 \right\rangle = \sigma_T^2 = \frac{4kTB}{R_L} \quad \text{A}^2 \qquad [2.5.1]$$

Where k is Boltzmann's constant and B is the bandwidth. This noise can be reduced by increasing R_L, however doing so will also reduce the receiver bandwidth.

* See the appendix for analysis of amplifier noise given the noise figure.

The thermal excitation of electron-hole pairs in the p-n junction also causes a bulk dark current I_D that occurs regardless of incident light. The mean-square value of the bulk dark current noise for a PIN photodiode is

$$\left\langle i_{DB}^2 \right\rangle = \sigma_{DB}^2 = 2qI_D B \quad \text{A}^2 \qquad [2.5.2]$$

where q is the electron charge. Another dark current is the surface leakage current I_L which arises from surface defects and the photodiode design. The mean-square value of the surface dark current is given by

$$\left\langle i_{DS}^2 \right\rangle = \sigma_{DS}^2 = 2qI_L B \quad \text{A}^2 \qquad [2.5.3]$$

Photodetectors are typically designed so that the dark currents are negligible compared to other dominant noise sources such as thermal and shot noise.

Quantum (Shot) Noise

We already saw that the detection of received light follows a Poisson arrival process which places bounds on the minimum received energy for detection. The same random arrival process of electrons is also responsible for quantum noise in any current due to the movement of electrons in discrete energy quanta. The mean-square

quantum noise current is proportional to the photocurrent *Ip* and for a PIN photodiode is given by

$$\left\langle i_Q^2 \right\rangle = \sigma_Q^2 = 2qI_p B \quad \text{A}^2 \qquad [2.5.4]$$

For APD's the avalanche gain process results in excess noise, resulting in a total shot noise of

$$\left\langle i_Q^2 \right\rangle = \sigma_Q^2 = 2qI_p BM^2 F(M) \quad \text{A}^2 \qquad [2.5.5]$$

Where M is the gain and $F(M)$a noise figure that depends on the material and $F(M) \approx M^x \quad 0 \le x \le 1$. Given low background radiation and thermal noise levels, devices such as electron counters and coherent receivers can achieve shot noise limited operation. Because shot noise is a property of the signal current itself, it cannot be eliminated.

Additive White Gaussian Noise (AWGN) Channel Model

The analysis of SNR requires consideration of a channel model describing the statistical noise parameters. This is necessary for deriving the optimal transmitter and receiver design. In this lidar system, all of the thermal noise sources can be treated as white Gaussian random variables. While the shot noise is a Poisson process, it can also be approximated to high accuracy as white Gaussian noise. Thus, the receiver system can be modeled as an additive white Gaussian noise (AWGN) channel.

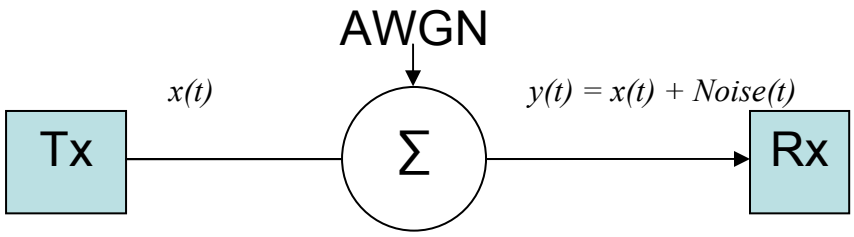


Figure 2.13 – AWGN channel.

Matched Filtering & Cross Correlation

The principle of matched filtering is to statistically minimize the impact of additive noise. The statistical properties of the signal and noise arriving at the receiver are determined by the transmitted signal characteristics, the added background radiation and the other additive noise sources within the receiver. In order to achieve the highest SNR and lowest detection error probability, a matched filter receiver amplifies the components of the received waveform in proportion to the expected incoming signal to noise ratio, providing the lowest possible detection error probability. For AWGN channels matched filtering is equivalent to performing cross correlation between the received signal and the original reference signal.

Fundamentally, pulse compression is an implementation of matched filtering and is therefore statistically optimal. Pulse compression uses cross correlation of the transmitted pulse against a reference pulse. The cross correlation between two signals is given by

$$R_{xy}(\tau) = E\{x(t)\, y(t+\tau)\} \qquad [2.5.6]$$

where $E\{\}$ is the expected value operator. For an energy signal the cross correlation is

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)\, y(t+\tau)\, dt \qquad [2.5.7]$$

and the power spectral density $S_{xx}(f)$ and autocorrelation $R_{xx}(\tau)$ of a random signal are a Fourier pair.

$$\begin{aligned} S_{xx}(f) &= F\{R_{xx}(\tau)\} \\ R_{xx}(\tau) &= F^{-1}\{S_{xx}(f)\} \end{aligned} \qquad [2.5.8 \text{ a, b}]$$

The input to the matched filter is the signal plus noise, and the output represents the optimum decision variable for detection. In synchronous systems such as a binary modulated communication channel, the matched filtering operation is completed by sampling the output at the end of each symbol time. At each sampling point a decision is made as to which symbol was sent by the transmitter. In radar applications the timing of the channel's response to a transmitted signal is not known beforehand. At any one sampling point, the output may represent a target response or a noise spike. Because of this the receiver makes a binary decision between detection or no detection at every sampling point, and its matched filter design yields the lowest detection error probability.

Analog mixing of the linear FM chirp assists in cross correlating the received signal for matched filtering. When the received and reference signals are multiplied together, their beat frequency contains the desired range information. Since the information is contained in the frequency of the beat signal, incorporating a matched

filter would require cross correlating the beat signal with every possible beat frequency. Luckily, this is exactly accomplished by performing an FFT on the signal and looking at the resulting frequency domain waveform. Taking an FFT of length N amounts to a matched filter bank of N possible symbols. In this way the linear FM chirp system is a very efficient implementation of an optimal receiver.

2.6 Optical Detection Methods

The amount of signal energy recovered in the receiver optics is very small, and the purpose of the lidar detection scheme is to amplify this signal while providing the highest possible signal-to-noise ratio (SNR). The method by which received light is detected and converted into an electrical signal generally falls into two categories, namely direct and coherent detection. The choice of detection method depends on the *SNR* limiting noise process and the tradeoffs with other performance considerations. SNR is defined as the ratio between the dechirped beat signal power (power at f_R) and the noise power spectral density level in the output of the FFT or spectrum analysis process. The final detection SNR typically must be at least 10 dB. SNR depends on the dominance of internal and external noise sources such as background radiation, thermal noise and quantum (shot) noise. Given low background radiation, direct and coherent detection are typically limited by thermal and quantum noise sources, respectively.

Direct Detection

The direct detection scheme is relatively simple in that no optical mixing is used. As shown in **Figure 2.14**, a CW laser is modulated by an RF chirp signal via an Electro-Optic Modulator. This modulated light pulse is amplified and transmitted to a target through a telescope. After it reflects back into the receiver optics, the echo signal is converted into electrical current at the photo detector. The photo detector can be a PIN or an APD photodiode which generates an electrical signal current that is proportional to the incident optical power. This process is similar to envelope detection in that the optical frequency and phase are discarded. The resulting baseband chirp signal is then mixed with the original chirp, resulting in the final dechirped signal. This signal is then processed by FFT or a spectrum analyzer to obtain the range information.

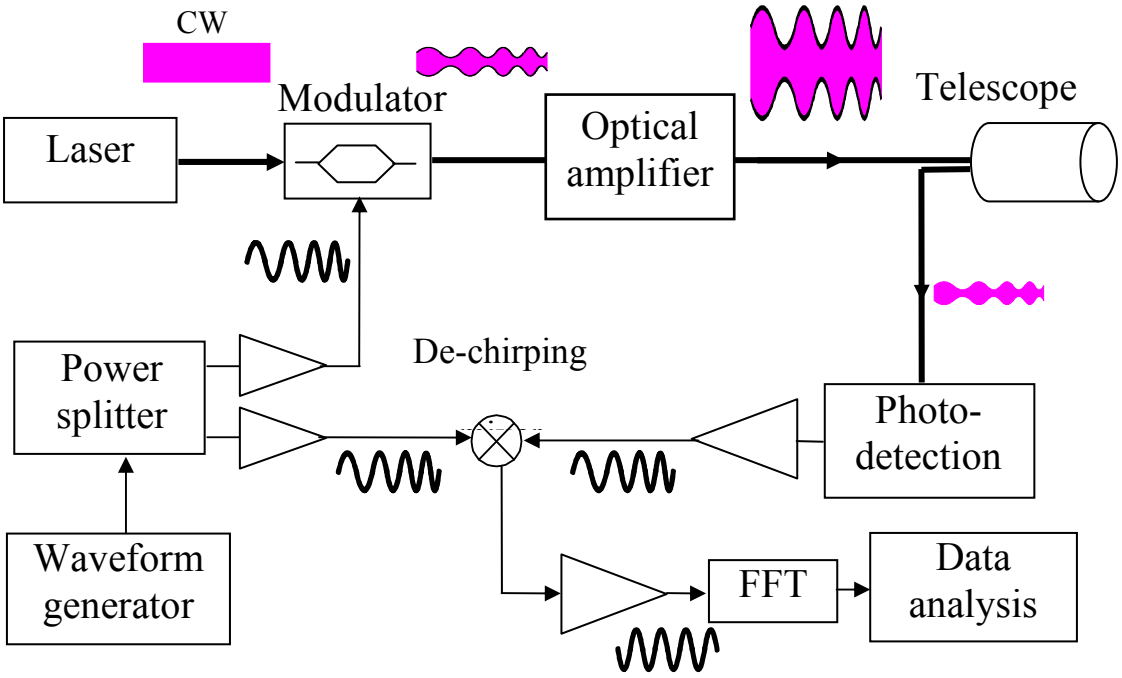


Figure 2.14 – Block diagram of lidar with direct detection.

Coherent Detection

The advantage of coherent detection is the ability to amplify the returned signal by multiplication with a strong local oscillator (LO) using optical mixing. Optical mixing occurs between two signals when their respective optical power is summed in an optical coupler and then squared due to the square-law detection property of the photo diode. The resulting electric current emanating from the diode contains the cross-product signal as well as other possible signal products. An optical heterodyne receiver can be implemented by applying a CW laser LO to one input of the receiver optical mixer. The wavelength of the LO is adjusted to provide a signal-LO cross

product at the chosen intermediate frequency (IF). By this method, coherent detection also allows optical phase information to be retained.

A lidar system with heterodyne detection is illustrated in **Figure 2.15**. The LO signal is generated by shifting the frequency of a part of the original signal through an Acousto-Optic Modulator, resulting in heterodyne operation at the intermediate frequency equal to f_m. Alternatively, a separate laser may be used to provide the proper wavelength.

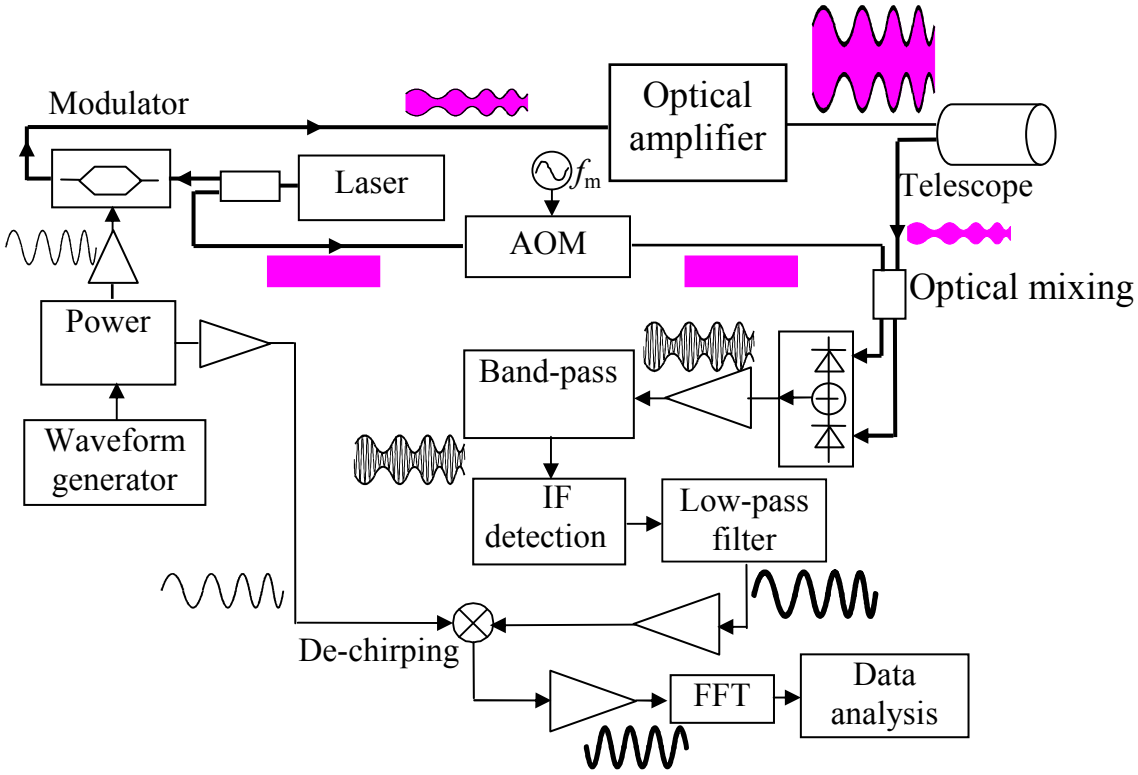


Figure 2.15 – Block diagram of lidar with coherent detection.

The mixing process provides gain of the input signal giving coherent detection the ability to overcome thermal noise. Coherent detection can thus achieve shot noise

limited SNR. **Figure 2.16** shows frequency domain illustrations of the mixing steps for each detection type.

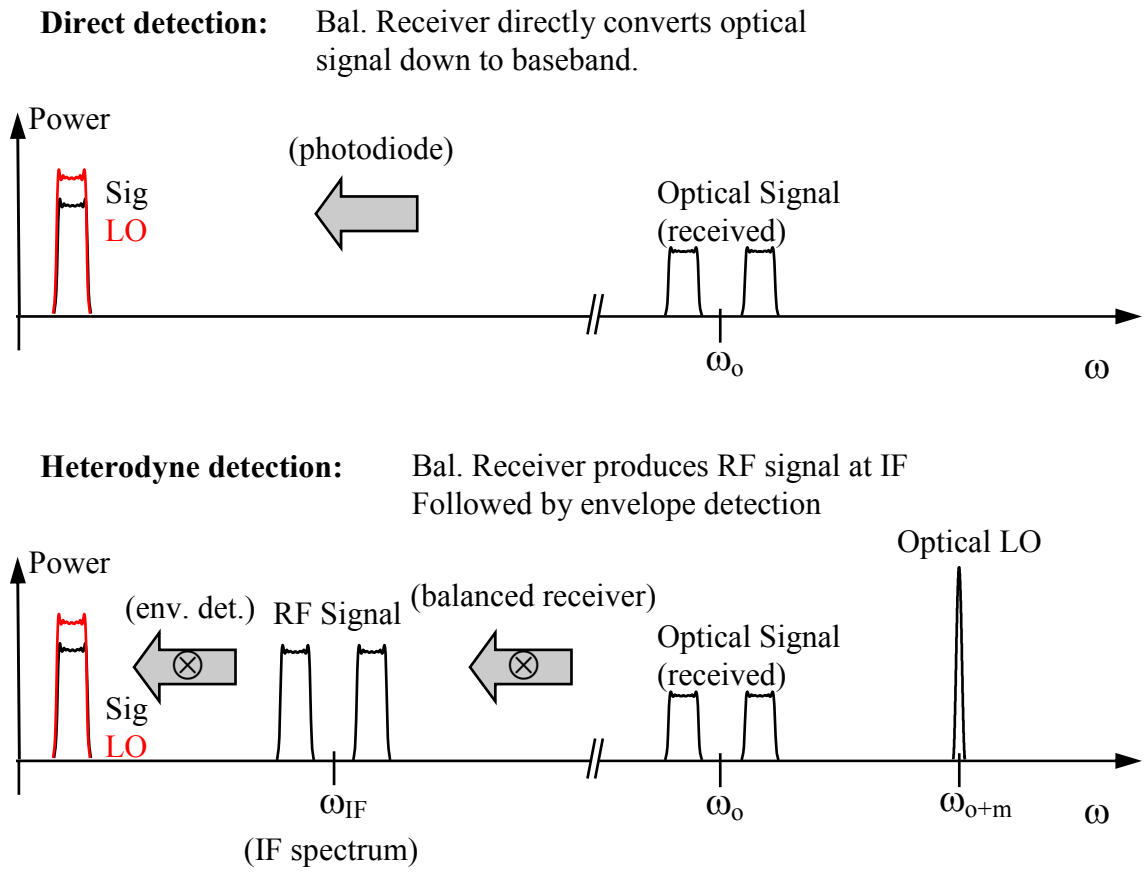


Figure 2.16 – Illustration of detection mixing processes.

SIGNAL TO NOISE RATIO COMPARISON

The limiting SNR for lidars depends on the detection method. Although many noise sources are present, a single noise source will dominate the total noise power. In direct detection the output of the photodiode has too little power to overcome the

constant thermal noise produced by the load resistor and amplifier. Given a total available input optical power P_{sig}, the limiting SNR for direct detection is

$$SNR_{dir} \approx \frac{2\Re^2 P_{sig}^2}{4kTB_e / R_L} \qquad [2.6.1]$$

For direct detection there is a 2 dB degradation of SNR for every 1 dB of signal power reduction. Coherent detection overcomes the thermal noise power by applying gain to the input light via the optical mixing process. The limiting SNR for coherent detection is

$$SNR_{coh} = \frac{\eta_q P_{sig}}{2h\nu B} \qquad [2.6.2]$$

In shot noise limited operation, for every dB reduction of signal power there is one dB degradation of SNR. This represents the best SNR that can be achieved, because the shot noise is a part of the signal current itself.

Calibrating the FFT Signal and Noise Levels to find SNR

The above equations give SNR based only on the received power and the noise process, not necessarily the SNR as observed on a spectrum analyzer. Spectrum analyzers can measure power spectral density by several methods, making a difference in comparing SNR across various sources. The HP 8565E is a superheterodyne analog spectrum analyzer. This type of SA uses a modulation scheme to mix the incoming RF signal down to an intermediate frequency (IF) where it is filtered with a narrow band pass filter. This results in the ability to measure the

power within a frequency bin specified by the IF band pass filter cutoff. Sweeping the local oscillator frequency allows the SA to steer this frequency bin to any point in the RF spectrum, producing a power spectrum trace.

The bandwidth of the IF filter determines the resolution bandwidth BW_{res}. When the signal bandwidth is smaller than the resolution bandwidth, then its absolute power is represented by the peak seen on the SA trace, as shown in **Figure 2.17** below. When it comes to noise, a different rule applies. Since noise has a broad and flat power spectrum, the amount of power captured in the IF filter is also affected by the bandwidth of the IF filter, as illustrated in **Figure 2.18**. To determine the noise power spectral density in standard units such as W/Hz, the noise level represented on the display must be first converted from dBm to a linear unit (e.g. Watts) and then divided by the resolution bandwidth.

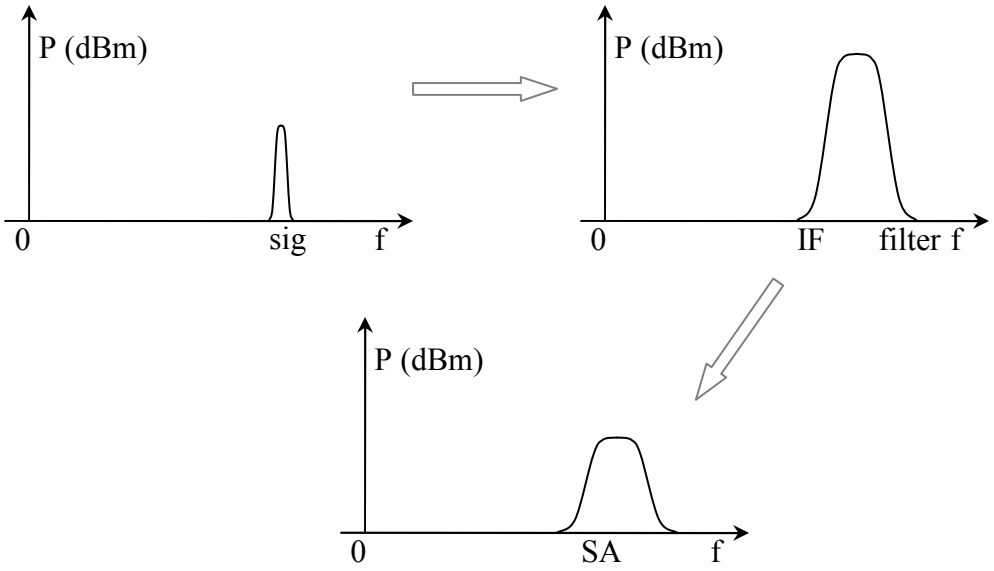


Figure 2.17 – Illustration of SA response to narrow band signal.

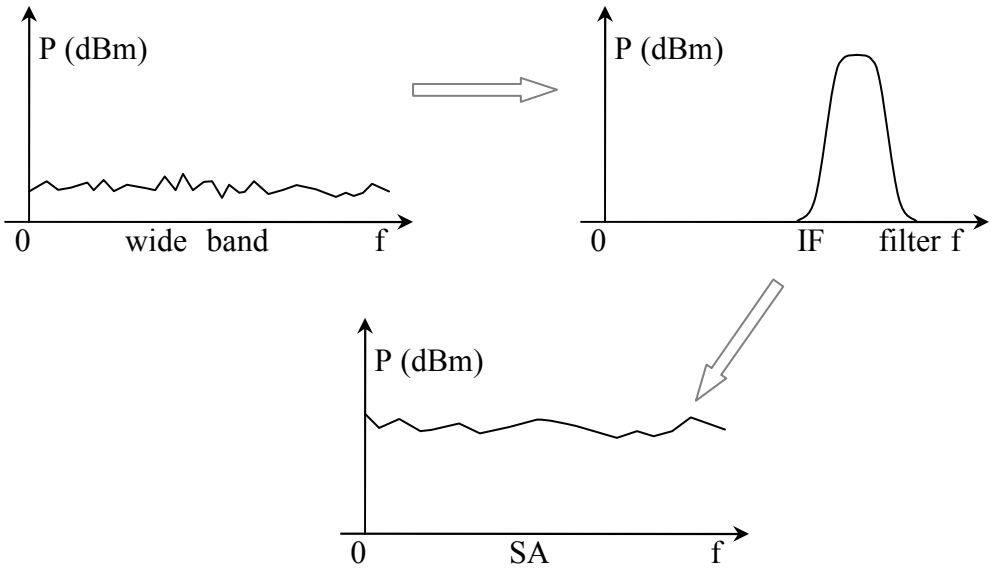


Figure 2.18 – Illustration of SA response to wide band noise.

Another widespread form of power spectrum measurement is the FFT, and it is necessary to account for the differences between the two. Like the superhet spectrum analyzer, the simulation FFT is limited in resolution. This stems from the property of the Fourier transform that the frequency resolution is the inverse of the time duration (and the time resolution is the inverse of the frequency span a la Nyquist). By normalizing the FFT by the number of points and the total bandwidth, the power spectral density in units of W/Hz can be found. Then, in the case of signals with bandwidth less than the frequency resolution, the actual signal power is found by multiplying the FFT point reading by its corresponding frequency resolution. In the case of wide band noise the frequency resolution can be disregarded and the noise power spectral density is read directly from the FFT. Adjusting this noise power spectral density, when expressed in W/Hz, to match the SA reading requires accounting for the SA resolution bandwidth.

The process for finding the SNR from FFT and matching it to the SA SNR reading requires several steps. First, the magnitude of the FFT of the received time domain voltage signal, V_{rx} is found and normalized by the square root of N, the number of sample points.

$$V_f = \frac{\left|FFT\left(V_{rx}\right)\right|}{\sqrt{N}} \quad \text{V} \qquad [2.6.3]$$

This voltage spectrum is converted into a power spectrum P_f by applying Ohm's Law with a known resistance R and normalizing by twice the one-sided bandwidth BW.

$$P_f = \frac{V_f^{\,2}}{R \cdot 2 \cdot BW} \qquad \text{W/Hz} \qquad [2.6.4]$$

This allows the noise floor to be accurately represented in W/Hz. However, it is important to note that the FFT doesn't actually *have* 1 Hz resolution in typical cases, but now has samples showing the correct W/Hz level at every sample point separated by the frequency resolution *df*, which is the inverse of the time duration *T*:

$$df = \frac{1}{T} \qquad \text{Hz} \qquad [2.6.5]$$

We can find the signal power *S* represented by a point Y in the power spectrum. To find signal power we first multiply *Y* by the frequency resolution *df*. Then the true signal power can be found as

$$S = 10\log_{10}\left(Y \cdot df\right) + 30 \qquad \text{dBm} \qquad [2.6.6]$$

As an extra step we can note that $2 \cdot BW$ is equal to the sample rate *Fs*, so we can write

$$\frac{1}{R \cdot 2 \cdot BW \cdot N} = \frac{1}{R \cdot F_s \cdot N} \qquad \text{s} \qquad [2.6.7]$$

We can regroup these several steps of normalization arriving at

$$\frac{df}{R \cdot F_s \cdot N} = \frac{1}{R \cdot T \cdot F_s \cdot N} = \frac{1}{R \cdot N^2} \qquad [2.6.8]$$

This reveals that the process of finding signal power could be done in one step, where instead of finding P_f we find another representation X_f directly as

$$X_f = \left| \frac{FFT(V_{rx})}{\sqrt{R} \cdot N} \right|^2 \quad \text{W} \quad [2.6.9]$$

from which the peak power S could be found as

$$S = 10\log_{10}(X_{peak}) + 30 \quad \text{dBm} \quad [2.6.10]$$

However, this way we would not express the power spectral density in W/Hz and therefore could not simultaneously determine the noise level.

Although the actual noise causes a random variation of the power spectrum, we are interested in finding the average noise level N_{avg}. Finding the average noise level requires averaging the spectrum over some valid (i.e. flat) range. Then, we can multiply this noise level by the spectrum analyzer resolution bandwidth $BWres$ and convert to dBm.

$$N_{avg} = 10\log_{10}(NoiseLevel \cdot BWres) + 30 \text{ dBm} \quad [2.6.11]$$

Finally, the SNR as it would be seen on the spectrum analyzer is given by

$$SNR = S - N_{avg} \quad \text{dB} \quad [2.6.12]$$

FREE-SPACE TRANSMISSION EFFECTS

Free space transmission introduces several modifications to the model of the returning light as a coherent linearly polarized plane wave. Coupling power from telescope optics into a single mode fiber involves the interaction of light with lenses, mirrors and apertures which requires us to examine diffraction. Interaction with the atmosphere degrades the coherence of the light field and causes random interference

patterns. These effects reduce the amount of usable signal power available to the receiver.

Near and Far Field Approximations

Consider an aperture in the *x-y* plane at *z=0* and a focal plane in the *x-y* plane at some distance *z=L*. Near and far field diffraction is differentiated by the Fresnel number,

$$F = \frac{a^2}{L\lambda} \qquad [2.6.13]$$

where *a* is the characteristic size (e.g. radius) of the aperture, *L* is the distance of the focal plane from the aperture and λ is the wavelength. The case $F \ll 1$ indicates Fraunhofer (far field) diffraction whereas $F \geq 1$ indicates Fresnel (near field) diffraction.

The Fresnel diffraction integral is

$$E(x,y,z) = -\frac{i}{\lambda}\int\int E(x',y',0)\frac{e^{ikr}}{r}\cos(\theta)dx'dy' \qquad [2.6.14]$$

Where *r* is the distance between *E(x,y,z)* and *E(x',y',0)*, k is the wavenumber $2\pi/\lambda$ and θ is the angle between the direction of *r* and the z axis. The Fresnel diffraction integral is generally difficult to compute and can be approximated. In the near field, varying *r* alters both the size and shape of the diffraction pattern; however as the distance from the source increases into the 'far field', the shape becomes constant and only the size changes.

For the far field case diffraction can be calculated from the more simple Fraunhofer diffraction integral

$$E(x,y) \propto \int E(x',y',0)\, e^{-ik(xx'+yy')/r} dx'dy' \qquad [2.6.15]$$

For cases when r is large relative to x and y the Fraunhofer diffraction integral simplifies conveniently so that the diffraction pattern as a function of the angle θ_x, θ_y is given by the 2D Fourier transform of the aperture intensity as a function of x, y.

$$E(\theta_x, \theta_y) \propto \int E(x', y')\, e^{-ik(\theta_x x' + \theta_y y')} dx'dy' \qquad [2.6.16]$$

or

$$E(\theta_x, \theta_y) \propto FFT\{E(x', y')\} \qquad [2.6.17\ a]$$

$$\theta_x \approx x/r \quad \theta_y \approx y/r \qquad [2.6.17\ b]$$

Fiber-to-Telescope Power Coupling Efficiency

The purpose of the telescope in the lidar is to focus optical power from a fiber end onto the target in the far field. The telescope then must couple the reflected light back into the same fiber. The transverse dimension of the light beam in the aperture plane has an effect on the smallest possible focused spot size in the focal plane. Diffraction through the aperture plane causes the focused spot to have an Airy disk profile. The first null of the Airy disk is

$$\sin \delta\theta = \frac{1.22\lambda}{d} \qquad [2.6.18]$$

where d is the telescope aperture diameter. Using the small angle approximation the smallest possible focused spot diameter is

$$D = \frac{1.22\lambda}{d} f \qquad [2.6.19]$$

where f is the focal length.

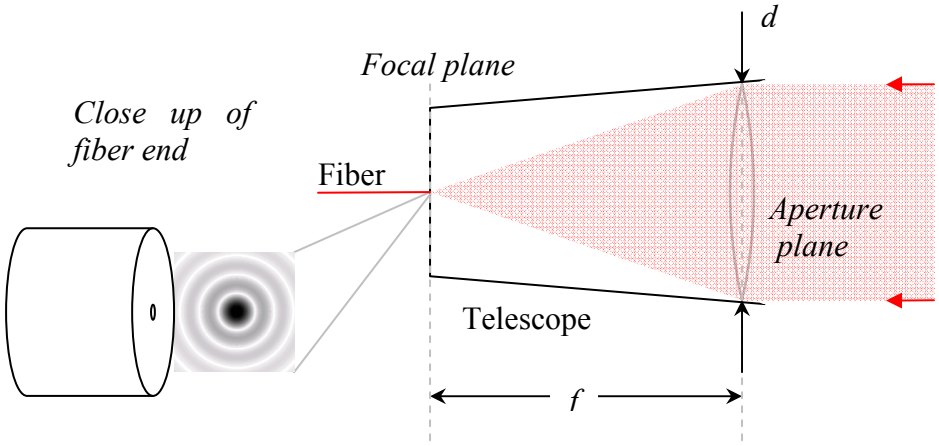


Figure 2.19 – Airy disk in telescope focal plane.

The spot size should be small enough that most of the power can be coupled into an 8 μm fiber, however the ratio d/f should be small enough to ensure meeting the acceptance angle of the fiber of about ± 6 degrees. For the lidar system these two criteria are satisfied with $0.18 \le d/f \le 0.20$ [21].

The optical power is uniform across the aperture plane given the assumption that the incoming light is a plane wave (due to the far field approximation). However, the light focused and captured into the fiber takes on a Gaussian radial profile. The mismatch between the focused Airy disk and the fiber's Gaussian profile reduces the

coupling efficiency into the fiber. The maximum theoretical coupling efficiency in this situation is approximately 42% [22]. Therefore, even under ideal conditions the coupling from air to fiber introduces 3.8 dB of loss.

Furthermore, the telescope usually includes a secondary mirror used to extend the effective size of the telescope body. This central obstruction blocks a portion of the aperture plane and affects the diffraction profile in the focal plane. This results in additional coupling loss which can be measured experimentally.

CHAPTER 3 – ANALYSIS OF SELF-CHIRPED HOMODYNE DETECTION

The self-chirped homodyne detection scheme relieves a significant performance limitation of typical lidars, namely that fine range resolution requires photo detectors with high sensitivity and wide output RF bandwidth. Unfortunately, high sensitivity and high bandwidth are not compatible in the design of the photodiode and the amplifier circuitry. Within the photodiode, better quantum efficiency can be achieved by increasing the active region volume for greater light absorption, while higher electrical bandwidth requires a smaller active region for shorter transit times. Likewise, the choice of load resistance and other amplifier parameters causes a tradeoff between noise level and bandwidth. The widest bandwidth is needed for heterodyne detection, where the RF signal produced by the photodetector is placed at the IF. For homodyne detection, the bandwidth must still be large enough to accommodate the baseband chirp bandwidth.

With the self-chirped homodyne technique, the photodiode directly produces the dechirped beat signal of the linear FM chirp. The major advantages of this system are:

- Eliminate RF dechirping and its associated loss.
- Allow use of large area PD and the possibility of using focal plane arrays.

The RF bandwidth requirement is no longer set by the chirped pulse bandwidth but by the range of dechirped beat frequencies at baseband. This allows for use of higher quantum efficiency and thus better receiver sensitivity as well as greatly increased chirp bandwidth. The new lidar system has demonstrated improved receiver

sensitivity in experiment and offers the potential to couple the returned light more efficiently using free space optics.

3.1 System Configuration

The system architecture is shown in **Figure 3.1**. An arbitrary waveform generator produces a linearly ramped chirp which modulates the laser output power via the EOM. The EOM is biased in the minimum power transmission point as discussed previously, so that the frequency of the chirp is doubled. This modulated optical signal is then split into two paths, one becoming the transmitted signal and the other becoming the local oscillator (LO). This is a significant simplification over heterodyne systems where the LO source must be frequency shifted from the main operating wavelength. After splitting, the signal portion is amplified through a fiber amplifier and fed to the telescope. The telescope is focused towards the target and returned power is reciprocally focused back into the fiber.

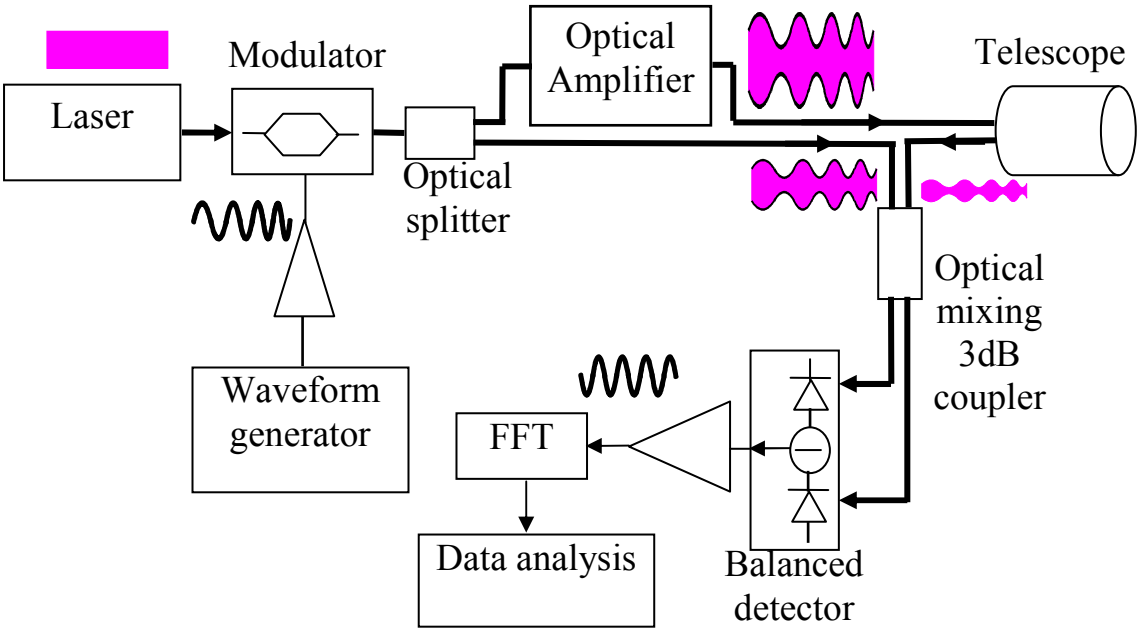


Figure 3.1 – Block diagram for self-chirped homodyne detection.

The receiver is composed of a 3dB coupler and balanced photodetector. In this setup the return signal and reference LO signal are combined and their optical beat frequency is found. The output of the balanced photodetector is the dechirped beat frequency corresponding to the target distance as described for linear FM pulse compression.

3.2 Signal Analysis of Self-Chirped Homodyne Detection

The analysis of the lidar detection scheme starts with the signal and LO E fields at the input to the photodetector,

$$E_{sig}(t) = \sqrt{2\eta P_{sig}}\cos\left(m\left(t-\Delta t\right)\right)e^{j\left(\omega t+\theta_{sig}(t)\right)} \qquad [3.2.1]$$

$$E_{LO}(t) = \sqrt{2\eta P_{LO}}\cos(m(t))e^{j(\omega t + \theta_{LO}(t))} \qquad [3.2.2]$$

P_{sig} and P_{LO} are the optical power, *m(t)* is the modulation term that is proportional to the input modulation signal voltage, and the terms θ_{sig} and θ_{LO} are randomly varying phase components. The time delay Δt is the round trip propagation delay corresponding to target distance. The modulation signal is driven by an arbitrary waveform generator to create the linear FM chirp,

$$m(t) = \cos\left[2\pi\left(f_1 + (f_2 - f_1)\frac{t}{\tau}\right)t\right] \qquad [3.2.3]$$

Note that when the EOM is biased for linear transfer that $\cos(m(t)) \approx m(t)$ in equation above. The cosine term is left in place for a rigorous analysis.

3.3 SELF-CHIRPED BALANCED DETECTION

The optical signals are directed into the 3dB optical coupler. The coupler and balanced photodiodes together will mix the two input signals and output the desired signal.

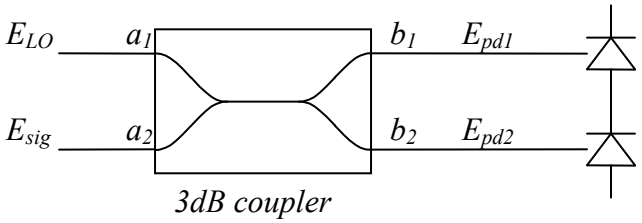


Figure 3.2 – Balanced receiver.

The ideal 3dB coupler scattering matrix is

$$S = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \qquad [3.3.1]$$

The outputs are thus found to be

$$\begin{bmatrix} E_{pd1} \\ E_{pd2} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \cdot \begin{bmatrix} E_{LO} \\ E_{sig} \end{bmatrix} \qquad [3.3.2]$$

$$\begin{aligned} E_{pd1} &= \frac{1}{\sqrt{2}}\left(E_{LO} + jE_{sig}\right) \\ E_{pd2} &= \frac{1}{\sqrt{2}}\left(jE_{LO} + E_{sig}\right) \end{aligned} \qquad [3.3.3]$$

Substituting the expressions for E_{sig} and E_{LO},

$$\begin{aligned} E_{pd1} &= \frac{1}{\sqrt{2}}\left[\sqrt{2\eta P_{LO}}\cos(m(t))e^{j(\omega t + \theta_{LO}(t))} + j\sqrt{2\eta P_{sig}}\cos(m(t-\Delta t))e^{j(\omega t + \theta_{sig}(t))}\right] \\ E_{pd2} &= \frac{1}{\sqrt{2}}\left[j\sqrt{2\eta P_{LO}}\cos(m(t))e^{j(\omega t + \theta_{LO}(t))} + \sqrt{2\eta P_{sig}}\cos(m(t-\Delta t))e^{j(\omega t + \theta_{sig}(t))}\right] \end{aligned} \qquad [3.3.4]$$

The photodiode current will be proportional to the incident optical power. In general the average optical power P_{pd} can be found using the Poynting vector,

$$\mathrm{P}_{pd} = \frac{1}{2}\mathrm{Re}\left\{E_{pd} \times H_{pd}^{*}\right\}\hat{a}_z = \frac{\left|E_{pd}\right|^2}{2\eta}\hat{a}_z \quad \mathrm{W/m^2} \qquad [3.3.5]$$

Putting in the received E field expressions,

$$P_{pd1} = \frac{1}{2\eta\sqrt{2}}\left|\sqrt{2\eta P_{LO}}\cos(m(t))e^{j(\omega t+\theta_{LO}(t))} + \sqrt{2\eta P_{sig}}\cos(m(t-\Delta t))e^{j\left(\omega t+\theta_{sig}(t)+\frac{\pi}{2}\right)}\right|^2 \qquad [3.3.6]$$
$$P_{pd2} = \frac{1}{2\eta\sqrt{2}}\left|\sqrt{2\eta P_{LO}}\cos(m(t))e^{j\left(\omega t+\theta_{LO}(t)+\frac{\pi}{2}\right)} + \sqrt{2\eta P_{sig}}\cos(m(t-\Delta t))e^{j(\omega t+\theta_{sig}(t))}\right|^2$$

After evaluating the complex conjugate product, we see the following optical power densities incident on the photodiodes:

$$P_{pd1} = \frac{1}{\sqrt{2}}\left(P_{LO}\cos^2(m(t)) + P_{sig}\cos^2(m(t-\Delta t)) + 2\sqrt{P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)+\tfrac{\pi}{2}\right)\right)$$
$$P_{pd2} = \frac{1}{\sqrt{2}}\left(P_{LO}\cos^2(m(t)) + P_{sig}\cos^2(m(t-\Delta t)) + 2\sqrt{P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{\pi}{2}\right)\right)$$

[3.3.7]

In general the photodiode output current will be (ignoring noise)

$$I_{sig} = \Re\, A_{eff} P_{pd} \qquad [3.3.8]$$

where $\Re$ is the responsivity, given by

$$\Re = \frac{\eta \cdot q}{h \cdot \nu} \quad \text{A/W} \qquad [3.3.9]$$

where η is the quantum efficiency, q is the electron charge, h is Planck's constant and ν is the optical frequency. There are two signal currents, one originating from each photodiode in the balanced photodetector.

$$I_{sig1} = \frac{\Re}{\sqrt{2}}\left(P_{LO}\cos^2(m(t)) + P_{sig}\cos^2(m(t-\Delta t)) + 2\sqrt{P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)+\tfrac{\pi}{2}\right)\right) \quad [3$$
$$I_{sig2} = \frac{\Re}{\sqrt{2}}\left(P_{LO}\cos^2(m(t)) + P_{sig}\cos^2(m(t-\Delta t)) + 2\sqrt{P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{\pi}{2}\right)\right)$$

.3.10]

Within each signal current there are two unwanted signal components, which are the directly detected LO and signal powers. To a large degree these components are eliminated by amplifying the difference between the balanced photodiode output currents. However, each optical power is effectively modulated by twice the chirp frequency due to the specific bias point chosen in the modulator, while the signal of interest is the difference between two chirp frequencies. Thus, the direct detected components will always be found at either DC or higher frequencies than the dechirped frequency range. This is an important factor that allows the simplified homodyne detection scheme to work by avoiding interfering frequency terms.

Finally, the signal output is found by amplifying the difference between the two currents, which results in

$$I_{sig1} - I_{sig2} = \Re\sqrt{2P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\left[\cos\left(\theta_{sig}(t)-\theta_{LO}(t)+\tfrac{\pi}{2}\right)-\cos\left(\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{\pi}{2}\right)\right] \quad [3.3.11]$$

which can be simplified as

$$I_{sig} = I_{sig1} - I_{sig2} = 2\Re\sqrt{2P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\sin\left(\theta_{sig}(t)-\theta_{LO}(t)\right) \quad [3.3.12]$$

The last term is a homodyne fading term that will cause fading of the output signal when $\theta_{sig}(t)-\theta_{LO}(t)$ is an integer multiple of π.

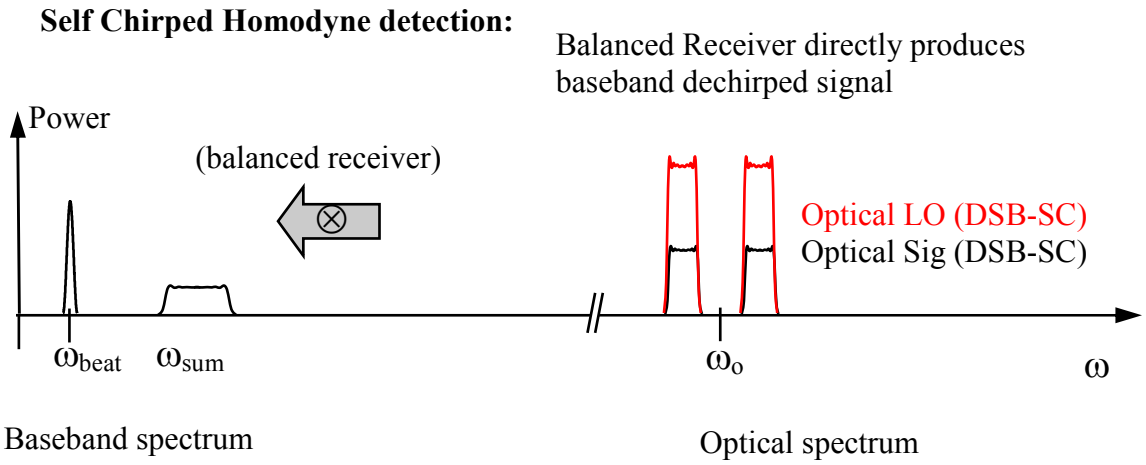


Figure 3.3 – Illustration of detection mixing processes.

CHAPTER 4 – NUMERICAL SIMULATION

The goal of simulation is to compare the performance of direct, heterodyne and homodyne detection systems. To validate our simulation, we compared the SNR results from experiment and simulation. Care was taken to provide accurate translation of SNR levels given by simulation, analysis and experiment. Thermal and shot noise sources were modeled as Gaussian random variables and for laser phase noise a Lorentzian distribution was generated.

METHOD OF OPERATION

The simulation generates time-domain signals stored as Matlab variables. Signals are represented by 1xN matrices corresponding to sample points in time at the rate Fs (Hz) over duration T (s). The signal variable names are documented in the block diagrams in the following figures.

The simulation process first read in a configuration file describing physical attributes and simulation setup, then returns data to the user, whether as variables or as plots. Given the complexity and memory limitations, it is not feasible to return every possible signal or variable. The following aspects make the program flexible:

- The program can run in one of two modes: single run or parameter sweep.
- In sweep mode the output plot can be specified through the configuration file.
- Parameters are organized by physical device in the 'params' struct variable.
- The FFT peak search is automatically plotted with each run.
- Signals can be 'probed' by inserting plot commands within the functions.

MATLAB FUNCTIONS

LIDAR_start.m

This is the start of the simulation. This function first reads in the configuration file *cfg.xls* to obtain the variables and parameters. Then, whether in single run or test sweep mode, it calls the appropriate function to initiate **LIDAR_main.m**. Since LIDAR_start is the parent function, it is the only one that can return variables to the workspace.

Figure 4.1 – LIDAR_start.m – Block diagram

The following parameters are associated with this function:

Table 4.1 – Simulation Parameters: Globals

Globals:				
sims	T	64E-6	s	duration
sims	Fs	3E+9	Hz	sampling rate
sims	upxn	1	integer	resampling factor to increase delay precision

T

This sets the time duration of signals in the simulation. The duration determines the resolution of the FFT and thus may affect the range resolution or accuracy.

Fs

Fs is the simulation sampling rate. The number of points in the signal variables will be **N** = **T** * **Fs**. The baseband signal bandwidth is roughly equal to the highest chirp frequency. Therefore, **Fs** should be greater than 2***chrp.F2**.

upxn

Depending on the sample rate, the implementation of propagation delay may have high error due to rounding to the nearest sample point. This can be overcome by increasing **upxn**, which specifies the factor by which the signal is temporarily resampled in order to increase this precision.

Table 4.2 – Simulation Parameters: System Selection

System Selection:				
sys	rsel	COH	string	receiver selection: APD, COH or PIN
sys	CxMode	hom	string	for COH: het=heterodyne hom=homodyne
sys	MxMode	FM	string	FM=FM Chirp PN=Pseudo Random
sys	smoothx	1	integer	FFT smoothing factor
sys	avg	1	integer	number of received chirps (Vrx) averaged
sys	fmaxopt	1E+9	Hz	optimal de-chirp frequency
sys	f_highpass	1E+6	Hz	Peak search minimum frequency
sys	BWres	30E+3	Hz	Spectrum analyzer resolution bandwidth

rsel

This string selects which receiver is modeled. Enter **APD** or **PIN** for the direct receiver models or **COH** for the coherent receiver.

CxMode

Selects between heterodyne and simplified homodyne detection.

MxMode

A pseudo random code scheme can also be used. To use, set MxMode to PN.

smoothx

Depending on the simulation duration **sims.T**, the FFT will most likely not have high enough resolution to produce a sample point at the true peak location. Instead, the data points will be on either side of the apparent peak, an effect referred to as 'scalloping' in radar terminology. Smoothing the FFT by the factor set in **smoothx** results in the occurrence of samples closer to the true peak and therefore reduced range error.

avg

To improve SNR, several chirps can be coherently averaged. This parameter only makes sense if each simulation contains one chirp, i.e. when **Tchp** > **T**/2.

fmaxopt

Use this to put a ceiling on the received frequency. The reference chirp will then also be delayed, forcing the dechirped frequency to this value. This helps maximize the amount of overlap of the received signals for dechirping, as determined by the formula

$f_r = \Delta t * (F2-F1) / \tau$. This will alter the value of **outputs.d_offset**.

LIDAR_main.m

This function is the top-level description of the FM Chirp lidar system. Based on input from **LIDAR_start**, this function runs the **recalc_params** function to calculate the derived parameters. It then generates a global time vector and calls the function blocks shown below. After each run it passes the results back to **LIDAR_start.**

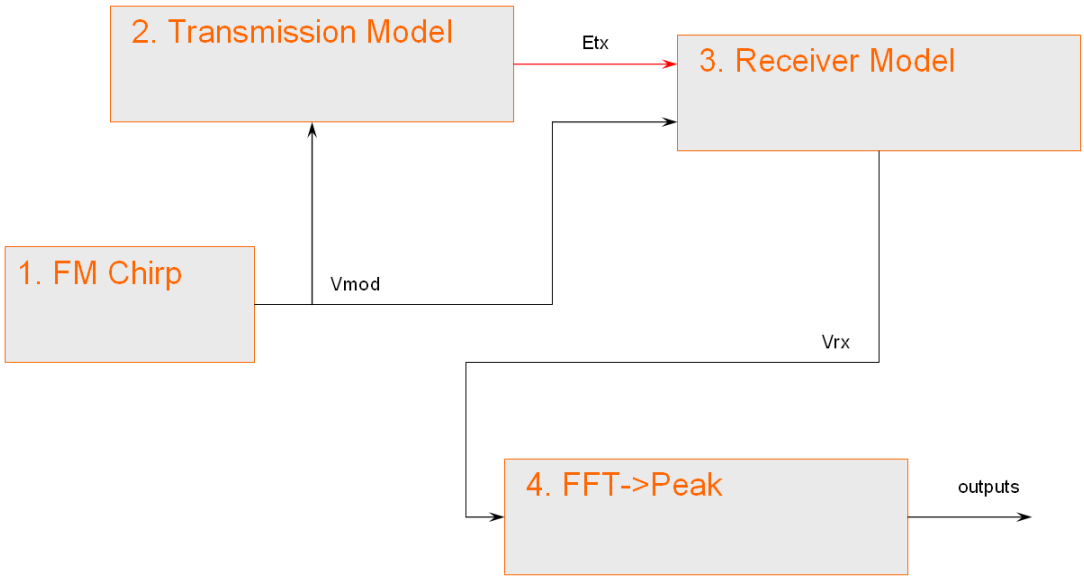


Figure 4.2 – LIDAR_main.m - Main block diagram of FM Chirp lidar

The individual blocks of this diagram are described next.

FM_chirp.m

The first stage in the lidar system is the generation of an RF frequency-modulated chirp signal.

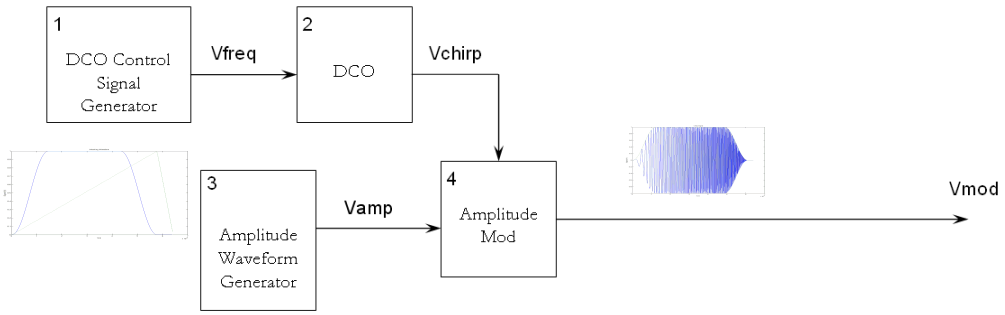


Figure 4.3 – FM_Chirp.m

Table 4.3 – Simulation Parameters: Chirp Waveform

Chirp Waveform:				
chrp	Tchp	64E-6	s	chirp period
chrp	F1	100E+6	Hz	minimum modulation frequency
chrp	F2	300E+6	Hz	maximum modulation frequency
chrp	F0	000E+0	Hz	frequency offset for F1 and F2 (F1=F1+F0…)
chrp	rdcyc	0.6250	%	ramp duty cycle (rising vs falling)
chrp	adcyc	0.9000	%	amplitude duty cycle (on vs transition)
chrp	qnt	000E+0	s	time quantization of inst. Frequency
chrp	Vpp	2.00	V	peak to peak voltage
chrp	Vdc	0.00	V	dc offset voltage
chrp	Ro	50	Ohms	output impedance

Tchp (see next page)

This specifies the period of each chirp signal. A new chirp is started at every multiple of this value within simulation duration **T**. Above that, the chirp signal will be zero-padded (no partial chirps generated). Regardless of the number of chirps, the FFT is

performed on the entire duration of the simulation. Because of the behavior of the FFT, it is optimal that only a single chirp exists, e.g. make **Tchp** equal to **sims.T**.

rdcyc

The ratio of chirp frequency rise time to fall time is specified by this duty cycle value.

adcyc

This adjusts the windowing function that reduces FFT side lobes. During the time that the chirp is ascending in frequency, the amplitude will rise for some time, hold for some time, and then fall back to zero for some time. This specifies the ratio that it is holding in the on state vs. rising or falling.

qnt

To experiment with a slowly updating frequency generator, this variable can be used to enter the amount of time that the frequency is held constant before incrementing to the next value, via a sample-and-hold function on the instantaneous frequency signal **Vfreq**.

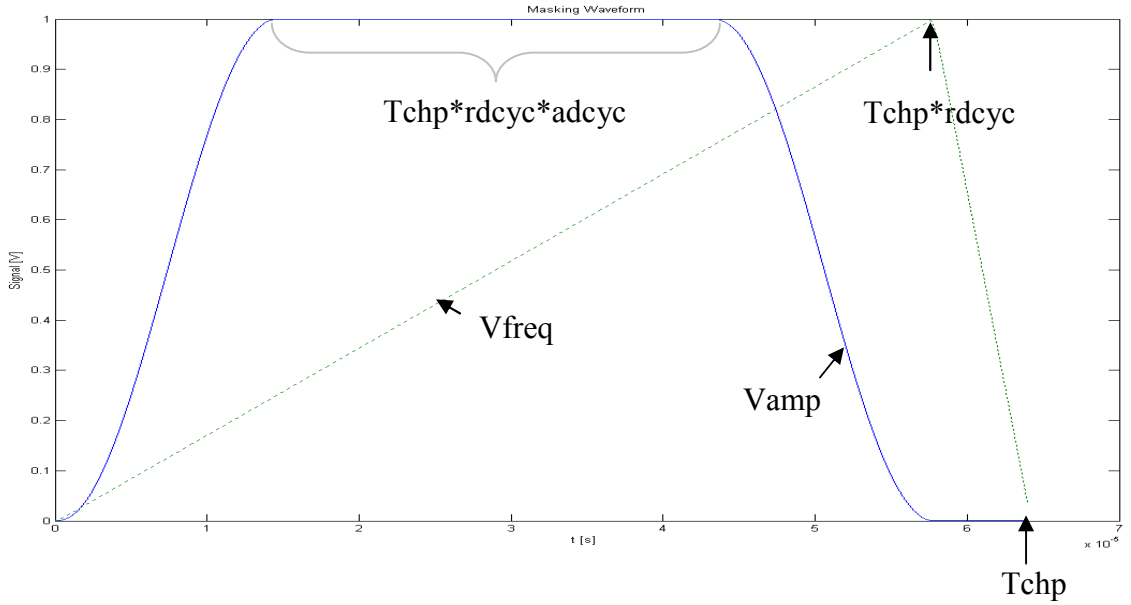


Figure 4.4 – Chirp signal dimensions

The figure above shows the dimensions of the RF chirp pulse as shown below:

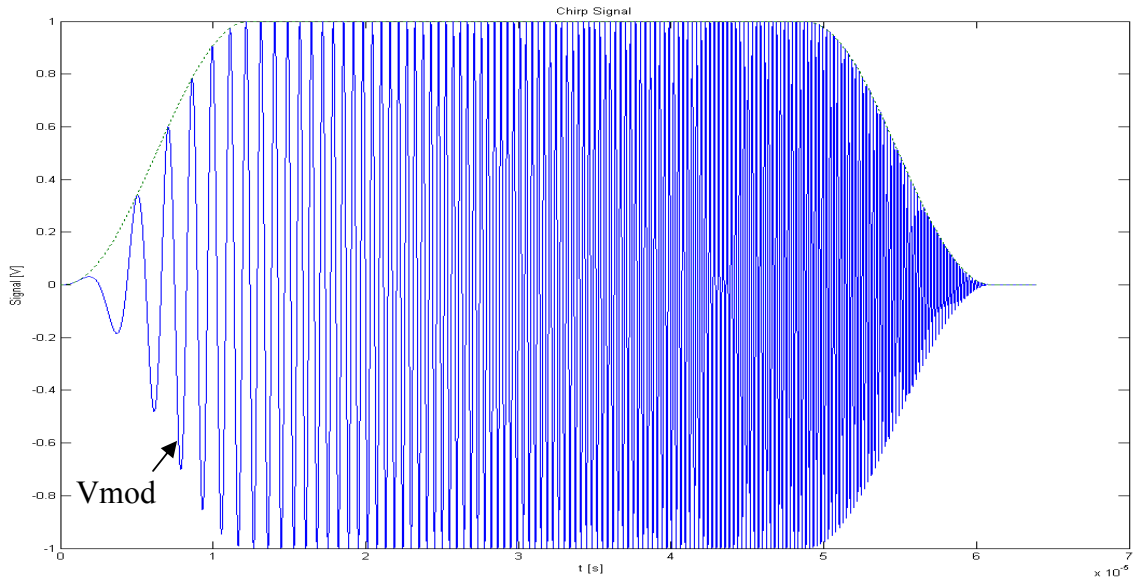


Figure 4.5 – Final RF Chirp signal

TX_model.m

The transmission model covers the laser source, modulator and propagation effects.

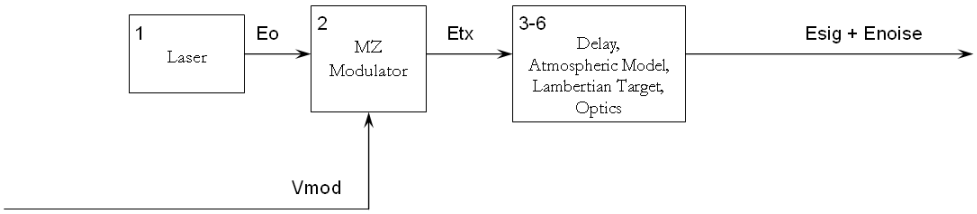


Figure 4.6 – TX_Model.m

Table 4.4 – Simulation Parameters: Laser Source

Laser Source:				
lsr	Lo	1.310E-6	m	laser source wavelength
lsr	Po	8.58E-03	W	laser source power
lsr	Lw	000E+0	Hz	single sided laser line width

Table 4.5 – Simulation Parameters: Mach-Zehnder Modulator

Mach-Zehnder Modulator:				
mzm	Vpi	2.00	V	characteristic switching voltage
mzm	Vdc	1.50	V	dc bias
mzm	Ri	50	Ohms	input impedance

Table 4.6 – Simulation Parameters: Transmission Channel

Atmospheric and Target properties:				
atm	Pattn	13	dB	Additional power attenuation
atm	state	off	string	noise & attn. on/off (Pattn is not affected)
atm	dst	1572.00	m	One way propagation distance to target
atm	Cn2	1.00E-14	m^-2/3	Atmosphere structure parameter
atm	lc	4.0E-3	m	Surface correlation length
atm	Wo	0.010	m	Beam Radius at exit aperture
atm	div	.4E-3		Half-angle beam divergence
atm	Wr	150.0E-3	m	Target Radius
atm	R_trg	0.460	W/W	target power reflectivity
atm	D_tel	0.2032	m	telescope diameter
atm	T_atm	0.995	V/V	atmospheric transmission
atm	T_opt	0.120	V/V	optical transmission

state

This parameter allows the atmospheric model to be disabled while the additional power attenuation remains in place. This is a quick way to match our lab-bench setup, where an optical fiber spool was put in place of the free-space optics.

RX_model.m

The receiver model contains three possible receiver choices. The receiver selection is made by setting **rsel** under the system parameters.

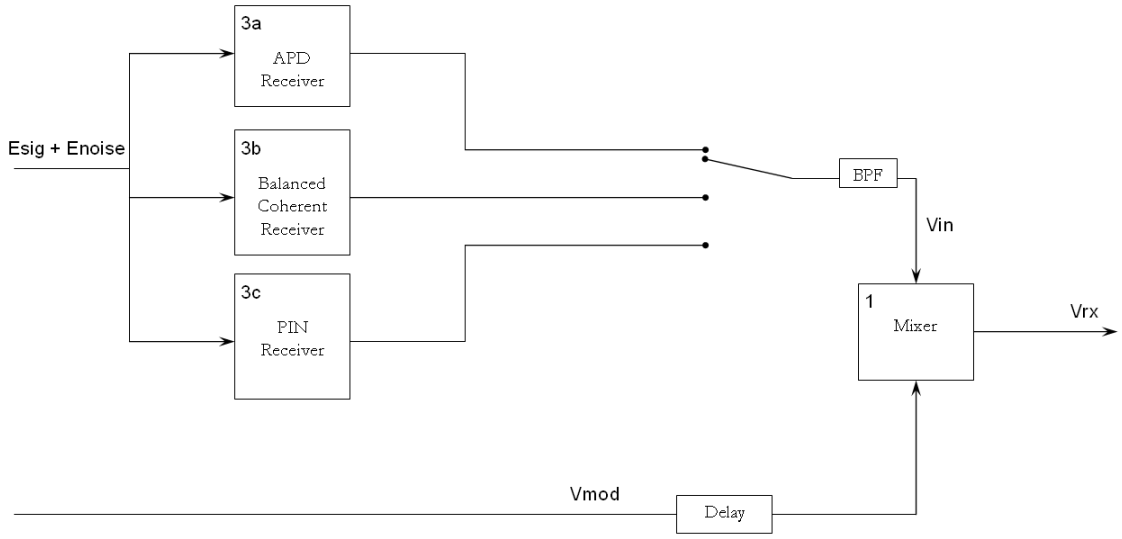


Figure 4.7 – RX_Model.m

Table 4.7 – Simulation Parameters: APD Photodetector

APD Photodetector:				
APD	Tfb	293.0	K	temp of feedback resistor
APD	Tdet	77.0	K	temp of detector
APD	Rfb	1E+3	Ohms	feedback resistor resistance
APD	Cfb	100E-15	F	feedback resistor capacitance
APD	INin	137E-15	A/rtHz	transimp. amp equivalent input current noise
APD	VNen	4E-9	V/rtHz	amplifier input voltage noise
APD	Ro	10E+9	Ohms	dynamic impedance
APD	Cdet	1E-12	F	detector capacitance
APD	M	1		receiver gain
APD	NF	1		excess noise factor
APD	Isol	000E+0	A/rtHz	solar current
APD	Ibgd	000E+0	A/rtHz	background current
APD	Ilkg	50E-12	A/rtHz	leakage current
APD	qe	0.80		quantum efficiency
APD	fo	3E+3	Hz	receiver center frequency

Table 4.8 – Simulation Parameters: Balanced Photodetector

Balanced Photodetector:				
COH	To	293.0	K	noise char temp
COH	Ro	50	Ohms	output impedance
COH	P_LO	2E-3	W	local oscillator power
COH	f_LO	1E+9	Hz	local oscillator center frequency
COH	lw	000E+0	Hz	single sided laser line width
COH	Isol	000E+0	A/rtHz	solar current
COH	Ibgd	000E+0	A/rtHz	background current
COH	Ilkg	50E-12	A/rtHz	leakage current
COH	Nep	18E-12	W/rtHz	Noise equivalent power
COH	qe1	0.240		quantum efficiency, arm1
COH	qe2	0.240		quantum efficiency, arm2
COH	Ztia	50.000	V/A	Transimpedance gain

Table 4.9 – Simulation Parameters: PIN Photodetector

PIN Photodetector:				
PIN	B	800E+6	Hz	receiver bandwidth
PIN	To	293.0	K	noise char temp
PIN	Ro	50	Ohms	output impedance
PIN	Isol	000E+0	A/rtHz	solar current
PIN	Ibgd	4E-6	A/rtHz	background current
PIN	Ilkg	50E-12	A/rtHz	leakage current
PIN	Nep	20E-12	W/rtHz	Noise equivalent power
PIN	qe	0.800		quantum efficiency
PIN	Ztia	700	V/A	Transimpedance gain

FFT_peak.m

The actual implementation of the receiver digital processing could vary. For the purposes of this simulation, the range is found by acquiring the FFT of the incoming de-chirped signal and then locating the peak value using Matlab commands.

Figure 4.8 – FFT_Peak.m

Table 4.10 – Simulation Parameters: Output Variables

Output Variables:				
outputs	Pchrp	0	dBm	Chirp Waveform Generator Output Power
outputs	Ptx	0	dBm	Transmitted Optical Power
outputs	Psig	0	dBm	Received Optical Power
outputs	Ne	0	dBm/Hz	Equivalent Amplifier Input Noise PSD
outputs	Ns	0	dBm/Hz	Photodiode Shot Noise PSD
outputs	Nt	0	dBm/Hz	Photodiode Thermal Noise PSD
outputs	d_offset	0	m	Distance offset due to local delay
outputs	d_rx	0	m	Detected target distance
outputs	d_lo	0	m	Lower bound on target 3dB uncertainty
outputs	d_hi	0	m	Upper bound on target 3dB uncertainty
outputs	d_error	0	m	Detected range error
outputs	d_3dB	0	m	3dB range resolution
outputs	d_res	0	m	Analytical range resolution
outputs	Crx	0	dBm	Received Signal Power read from FFT
outputs	Navg	0	dBm	Noise level read from FFT
outputs	SNR	0	dB	Signal to Noise Ratio

d_offset

The reference signal **Vmod** is delayed to maximize time-overlap of the two chirp signals. The distance that this delay represents is stored in **d_offset**. See **sys.fmaxopt**

d_rx

The range found by the lidar FFT is stored in this var.

d_lo & d_hi

The lower and upper bounds on the range are saved in d_lo and d_hi, respectively. They represent the distance reading found at the 3dB points on the FFT peak.

d_error

The difference between the actual distance and the found distance. **d_error** = **dst-d_rx**

d_3dB

The uncertainty of distance reading, calculated by subtracting **d_lo** from **d_hi**.

d_res

This is the analytical range resolution based on the bandwidth and **SNR_in**.

Testing

The following parameters determine the generation of a simulation test sweep. The idea is to vary one parameter to note the trends. Set the parameter **state** to 'on' in order to sweep any parameter over a range of values, for instance to obtain the SNR vs Received Power plot.

Table 4.11 – Simulation Parameters: Parametric Testing

Parametric Testing:				
test	state	on	string	test on or off
test	x_dep	outputs.Psig	string	name of x variable (Type.Name)
test	y_dep	outputs.CNR	string	name of y variable (Type.Name)
test	x_var	atm.Pattn	string	name of independent variable (Type.Name)
test	x_start	32		start value
test	x_stop	102		stop value
test	n	10	integer	number of points to compute
test	m	1	integer	number of times to repeat each point
test	i_scale	lin	string	scale lin=linear; log = logarithmic
test	x_scale	lin	string	scale lin=linear; log = logarithmic
test	y_scale	lin	string	scale lin=linear; log = logarithmic

CHAPTER 5 – EXPERIMENTAL VERIFICATION

To test the performance of the three detection methods, we assembled experimental lidar systems with direct, heterodyne and simplified homodyne detection. The systems used as many common components as possible. To avoid the uncertainties due to target reflectivity, the coupling efficiency of the telescope and turbulence of free space transmission, a 22.7-km standard single-mode optical fiber and optical attenuator were used to simulate an ideal channel. This allowed the three systems to be compared fairly without introducing the telescope and associated variability.

SYSTEM OVERVIEW

The optical pulse duration was 40 μs and the pulse repetition rate was 9.4 kHz yielding a duty cycle of approximately 38%. The modulation frequency was linearly chirped from 100 MHz to 300 MHz within each pulse, producing a 5 MHz/μs chirp rate. An RF spectrum analyzer was used to perform FFT and the resolution bandwidth was set to 30 kHz. A diode pumped 1319-nm Nd:YAG laser was used as the source for direct and homodyne detection, while two 1550-nm lasers were used for heterodyne detection to achieve a 15-GHz IF through optical mixing. A balanced photodiode with 800-MHz bandwidth was used as the detector for both direct and homodyne detection due to their relatively low receiver bandwidth requirements. Because heterodyne detection requires a much wider receiver bandwidth to accommodate the IF, a photodiode with a 20-GHz bandwidth was used whose

responsivity is much less ($\Re \approx 0.45$) than that of the 800 MHz balanced photodiode model ($\Re \approx 0.95$). In comparing the sensitivity of the three systems, the heterodyne sensitivity was reduced by an additional 3.25 dB because of this factor.

Self Chirped Lidar Equipment Specifications

Table 5.1 lists the equipment used in the field tested lidar and their parameters.

Table 5.1 – Equipment specifications for the self-chirped lidar.

Equipment	Make Model	Parameters	
Function Generator	Agilent 33205 A	Period Waveform Amplitude Duty Cycle	106μs square 4.0 Vpp 50%
Arbitrary Waveform Generator	Analogic 2045	Output Entry mode	B equation
LPF 1	Mini-Circuits SLP-450	Bandwidth	DC-450 MHz
LPF 2	Mini-Circuits SLP-400	Bandwidth	DC-400 MHz
Amplifier 1	Mini-Circuits ZHL-3010-SMA	Gain Bandwidth $P_{1dB\ (ref.\ to\ output)}$	30 dB ±1 50-1000 MHz 26 dBm
Voltage Source	Hewlett Packard E3630A	Voltage	5.50 V
Laser	Lightwave Electronics 125/126	Wavelength Power	1319 nm 8 dBm
MZ Modulator	JDS Uniphase OC 192 10020427	Wavelength Band Bandwidth	1310 nm > 10 GHz
Fiber Splitter 1	Newport F-CPL-B12351	Wavelength Bandwidth Max. IL, 1-2 Max. IL,1-3	1310/1550 nm ± 40 nm 0.65 dB 11 dB
Polarization Controller	n/a	Type	manual fiber loop
Fiber Spool	Corning Fiber	Length	22.7 km
PDFA	IPG Photonics FluoroAmp 1310	Gain	0 - 25 dB
Fiber Splitter 2	Newport F-CPL-B12355	Wavelength Bandwidth	1310/1550 nm ± 40 nm
Telescope	Celestron n/a	Optics Type Resolution Aperture Dia.	Newtonian diffraction limited ≈ 5 ”
3dB Coupler	Newport F-CPL-B22355	Wavelength Bandwidth Max. IL	1310/1550 nm ± 40 nm 3.6 dB
Balanced Photodetector	Newfocus 1617-AC	Wavelength Bandwidth Typ. Max. Resp. Transimpedance Gain Min. NEP Saturation Power	900-1700 nm 40 kHz – 800 MHz 1.0 A/W 700 A/V 20 pW/√Hz 0 dBm
Amplifier 2	SHF 105 P	Bandwidth Gain $P_{1dB\ (ref.\ to\ output)}$	50 kHz – 36 GHz 22 dB 10 dBm
Spectrum Analyzer	Hewlett Packard 8565E	BWres	30 kHz

The arbitrary waveform generator was set in equation mode, and the following equation was entered in the memory:

F42 = AT TRIG RPT 1(FOR 40u 0.4*SIN((100M+200M*t/40u)*t)) CLK = 1.25n

[5.1]

A photo of the lidar lab bench assembly is shown in **Figure 5.1** below.

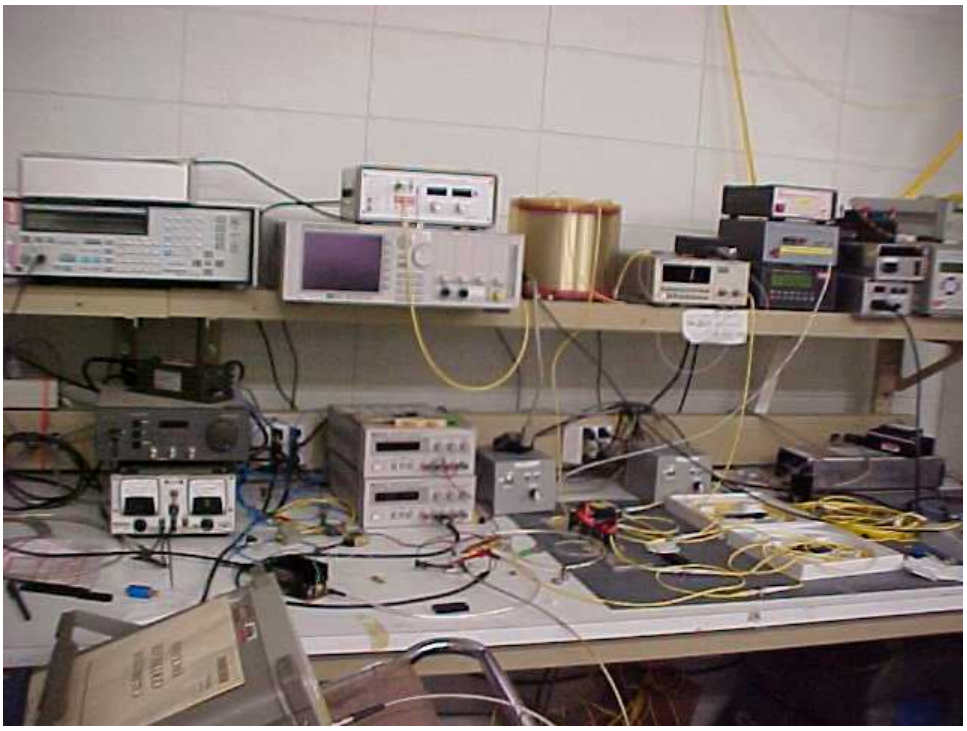


Figure 5.1 – Lab bench setup of the self chirped lidar.

RANGE VERIFICATION

Figure 5.2 shows an example of the dechirped beat frequency produced by the simplified homodyne system. The location of f_{RX} at approximately 22 MHz is determined by the difference between the propagation delay of 110.4 μs and the delay of 106 μs before the second pulse as

$$f_{RX} = \frac{200MHz}{40\mu s}(110.4\mu s - 106\mu s) = 22MHz \qquad [5.2]$$

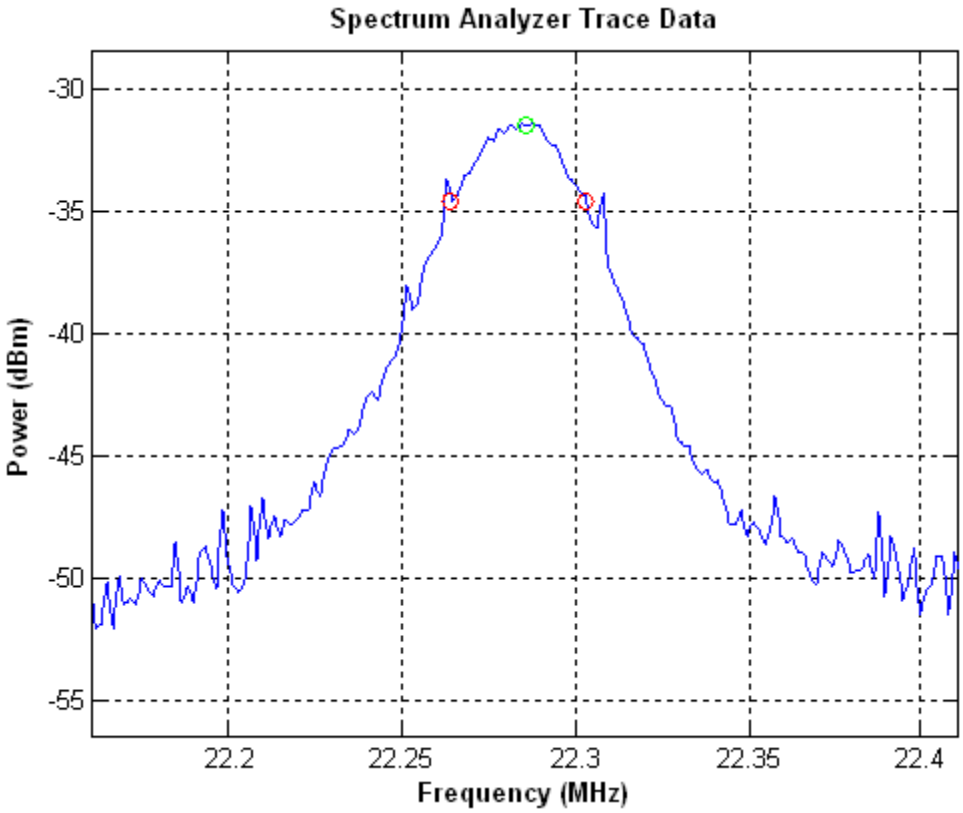


Figure 5.2 – Recorded trace of the dechirped beat signal.

(BW_{res}=30 kHz, τ=40 μs, PRT=106 μs, B=200 MHz, R≈22.7 km/2, f_{RX}≈22 MHz, n=1.46)

In order to validate the linear FM range finding method, various fiber lengths were measured and inserted in the transmission path of the lidar. The one-way propagation distance was increased by up to six meters. For each test, 20-dB CNR was maintained by adjusting the optical attenuator, and the polarization and modulation index were manually adjusted for highest CNR. The spectrum analyzer was used to zoom in on the recovered carrier signal, and the resulting trace data was recorded to text files.

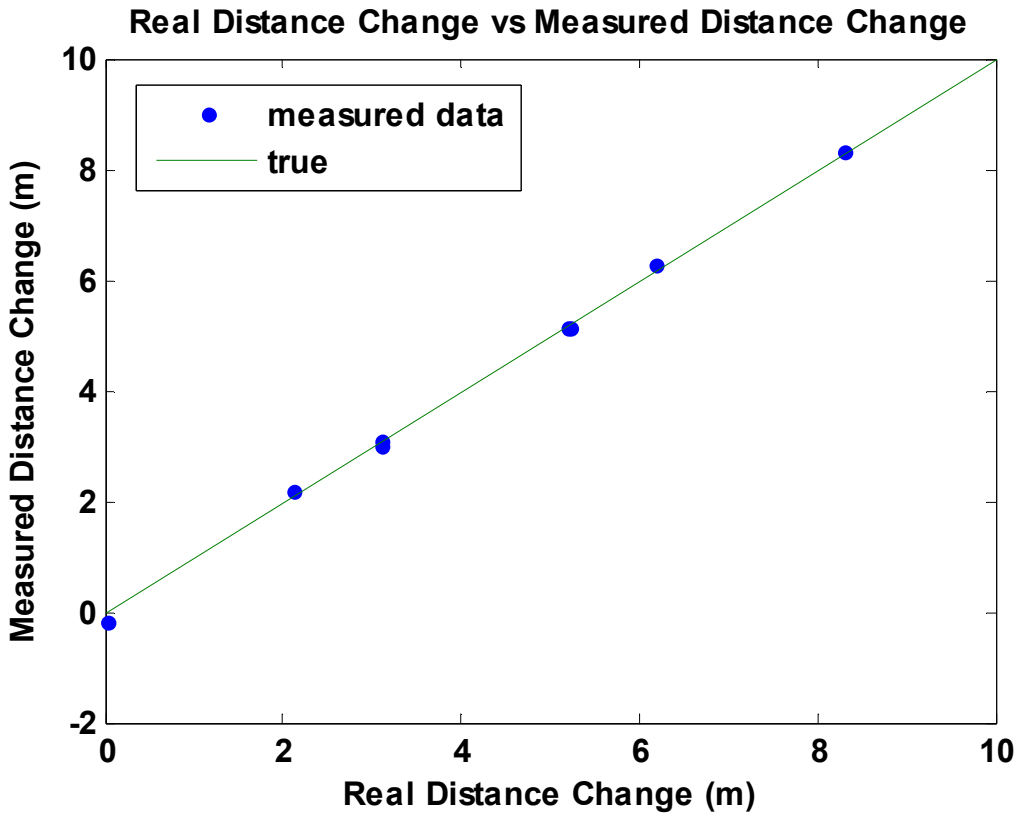


Figure 5.3 – Range finding experiment results.

(τ=40 μs, BWres=30 kHz, R≈22.7 km/2, f_{RX}≈23 MHz, SNR≈20dB, n=1.46, σ=8.5 cm)

A calibration value was used to remove the mean error from the distance measurements due to the unknown exact length of the 22.7-km delay spool. The standard deviation of error was 8.5-cm. The experimental value for the length of the delay line is 22.697-km.

BLOCK DIAGRAMS OF THE LIDAR SYSTEM TEST BED

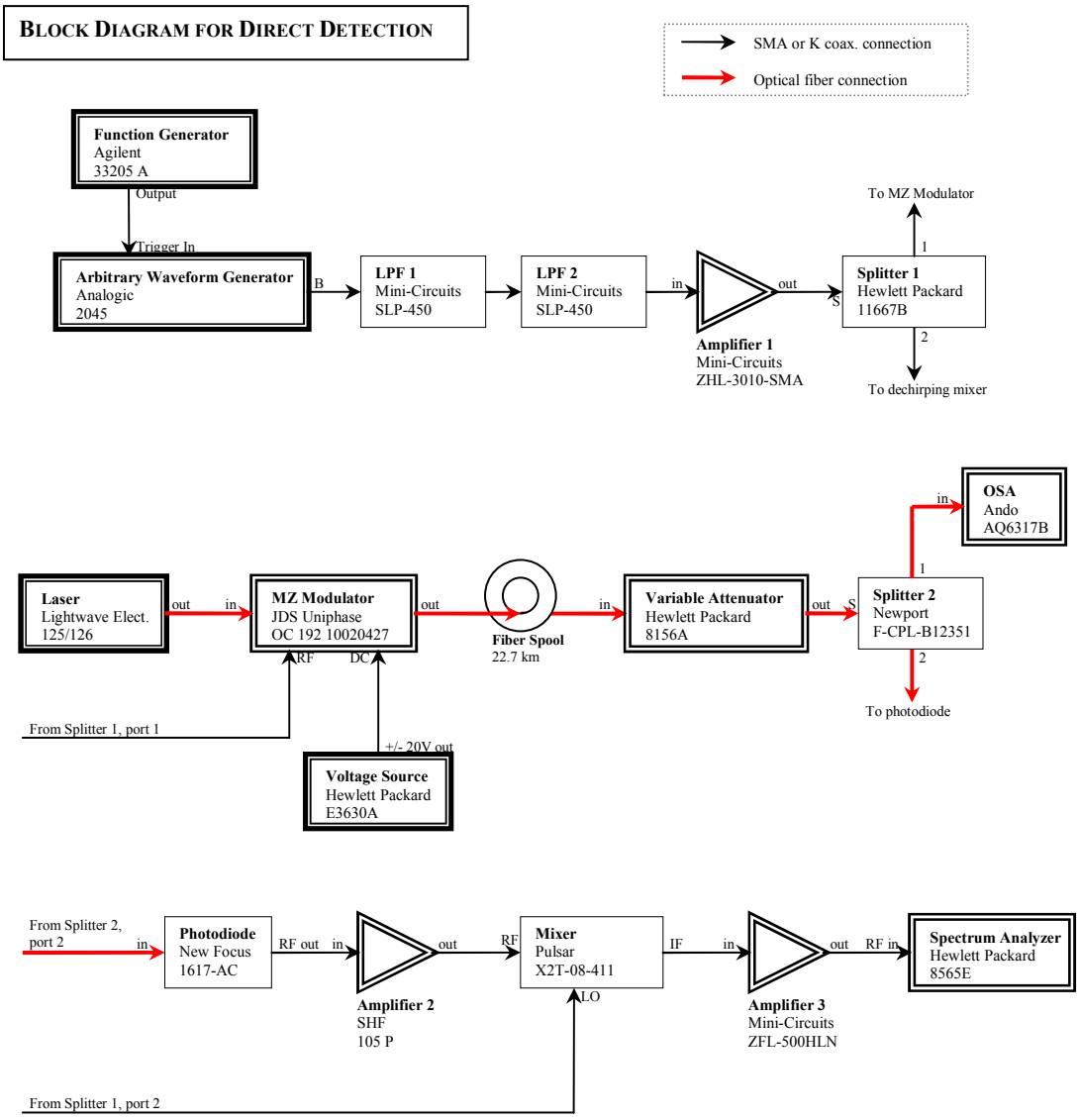


Figure 5.4 – Block diagram of direct detection prototype.

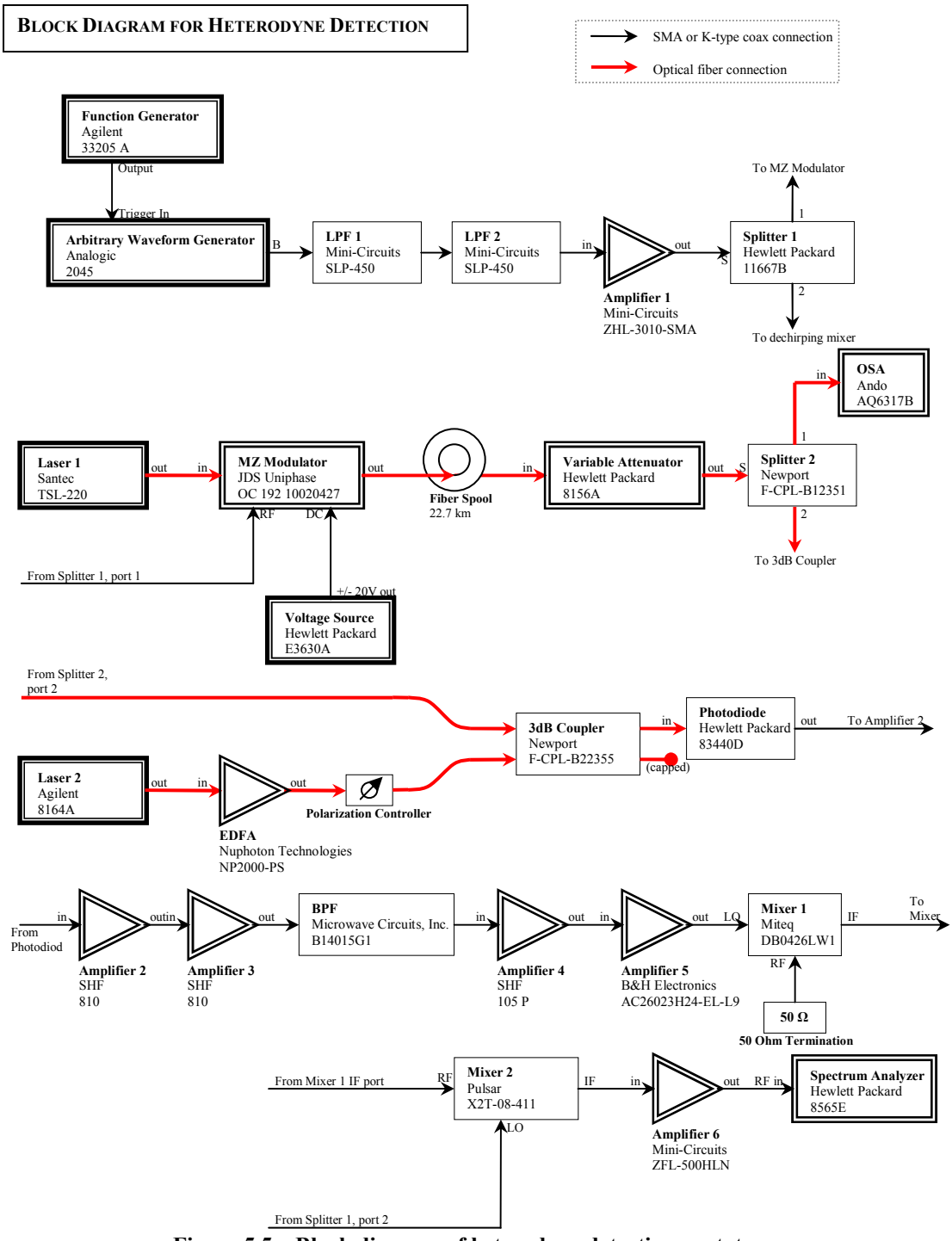


Figure 5.5 – Block diagram of heterodyne detection prototype.

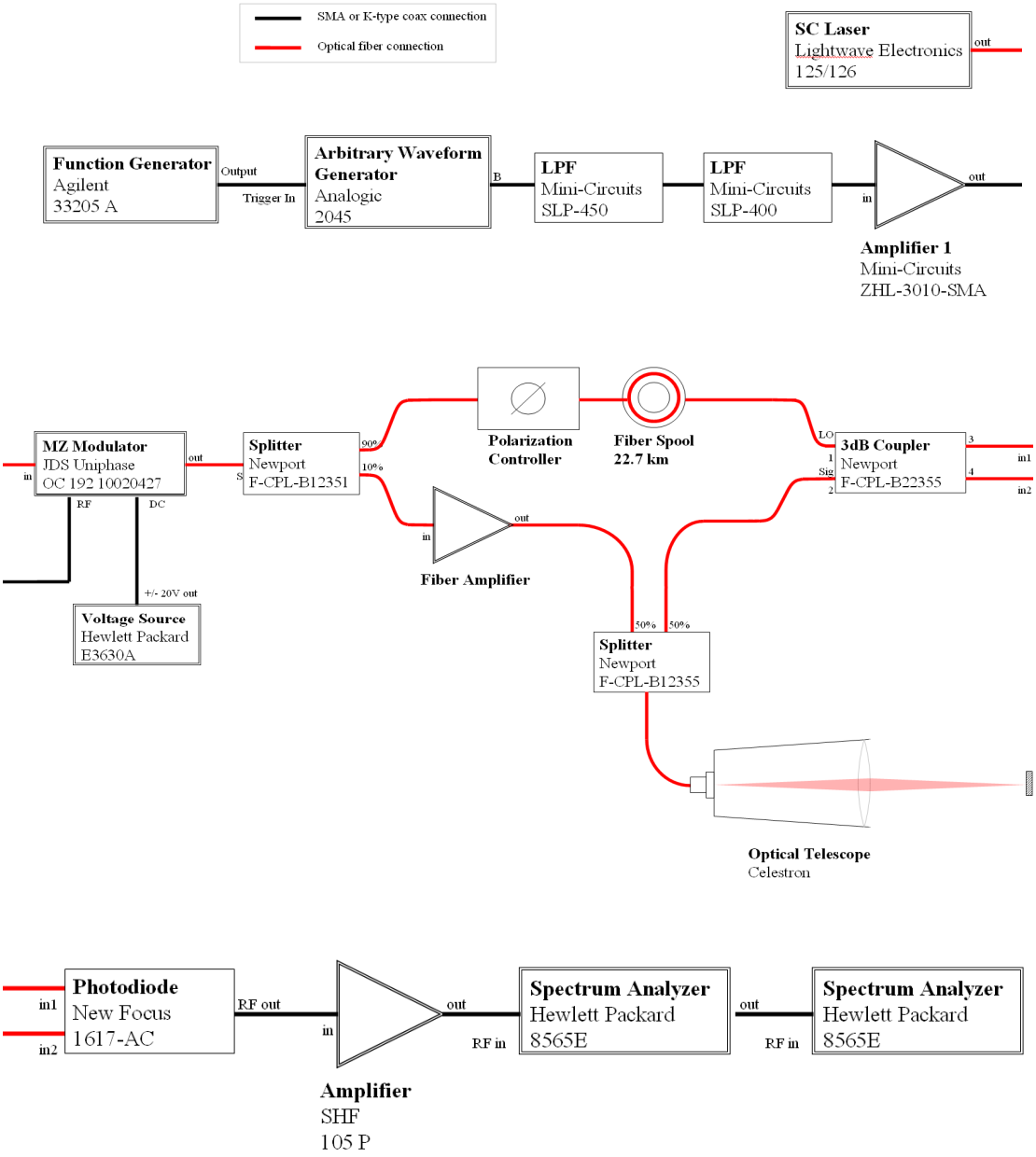


Figure 5.6 – Block diagram of self chirped homodyne lidar prototype.

RECEIVER SENSITIVITY

Experimental Results

For direct detection, the RF chirp is split into two parts, one for modulating the transmitted light and the other for dechirp mixing after photodetection as shown in the block diagram in **Figure 5.4**. For direct detection only one of the photodiodes was used. The photodiode performs envelope detection on the modulated light, reproducing the RF chirp signal which is then fed to the dechirping mixer. The sensitivity of the system was roughly -40 dBm for the 10 dB SNR requirement.

Table 5.2 – Direct detection sensitivity data.

Pin (dBm)	SNR (dB)
-25	38.00
-27	34.50
-29	31.00
-31	27.00
-33	24.00
-35	20.00
-37	16.00
-39	11.50
-41	8.00
-43	5.00

The measured sensitivity is lower than the theoretical equation due additional losses from the real system's duty cycle and nonlinearity of the MZM modulation.

The SNR results for heterodyne detection are shown in **Table 5.3**. The efficiency of the envelope detection and dechirping operations were poor and caused roughly 30 dB less than the expected shot noise level sensitivity. By extrapolating the data

slightly to 10 dB SNR, we obtain a sensitivity limit of -70 dBm for the heterodyne system.

Table 5.3 – Heterodyne detection sensitivity data.

Pin (dBm)	SNR (dB)
-36.30	43.70
-38.30	41.60
-40.30	39.70
-42.30	37.80
-44.30	35.70
-46.30	33.70
-48.30	31.70
-50.30	29.70
-52.30	27.90
-54.30	25.90
-56.30	24.00
-58.30	22.20
-60.30	20.20
-62.30	18.80
-64.30	18.00

The heterodyne detection system required the greatest complexity as well as the highest bandwidth components, and did not reach shot noise limited sensitivity. This motivated the development of the homodyne system that removes the need for RF processing. For comparison, the results for the self chirped homodyne system are shown in **Table 5.4**. The SNR data are also presented graphically for comparison in the next section.

Table 5.4 – Self chirped lidar sensitivity data.

Pin (dBm)	SNR (dB)
-54.92	41.21
-57.20	40.39
-59.47	38.99
-61.75	37.46
-64.03	35.72
-66.30	34.38
-68.58	31.43
-70.85	31.17
-73.13	30.59
-75.41	28.27
-77.68	26.70
-79.96	24.38
-82.23	23.08
-84.51	20.99
-86.79	19.15
-89.06	17.11
-91.34	13.83
-93.61	13.43
-95.89	10.62

Receiver Sensitivity Comparisons

The theoretical shot-noise limited SNR, assuming perfect mixing efficiency, is marked by the solid line in **Figure 5.7**, where the quantum efficiency of the photodiode was 0.85 and the resolution bandwidth was 30 kHz. As expected, comparing the SNR for direct and coherent detections we find that the slopes are roughly 2 dB/dB and 1 dB/dB, respectively.

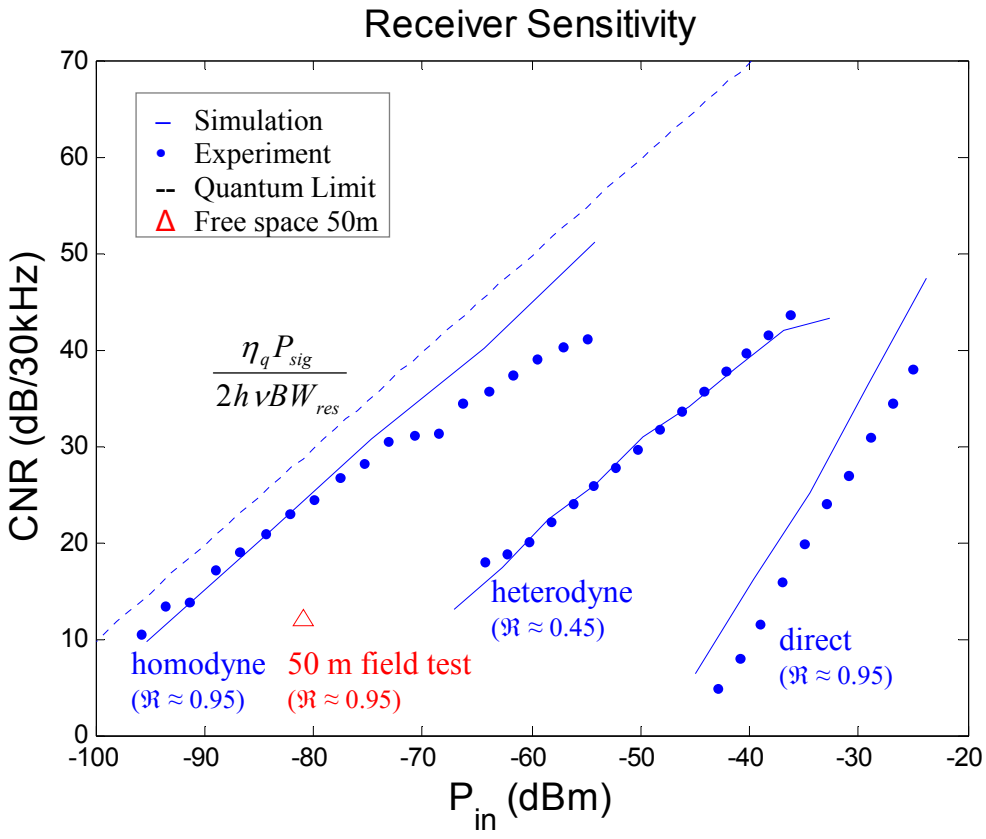


Figure 5.7 – Sensitivity of direct, heterodyne and self chirped detection.

(BW_{res} = 30 kHz, τ = 40 μs, PRT ≈ 100 μs, BW_{chirp} = 200 MHz, No averaging)

There is an apparent saturation effect for homodyne detection at high SNR. This effect was due to underestimating the SNR during tests, because the polarization and modulator biasing were more precisely tuned as the margin of SNR became smaller. For homodyne detection a maximum sensitivity of -96 dBm was obtained with 10 dB SNR (30-kHz resolution bandwidth, no averaging). Thus the homodyne receiver has an additional 30 dB of sensitivity compared to the heterodyne receiver in our experiment. What accounts for this significant difference?

By adjusting the parameters to match the hardware, we compared the simulation sensitivity with that of experiment as shown in the solid lines in **Figure 5.7**. The simulation of the heterodyne system confirms a similar sensitivity reduction to experiment.

One concern with coherent detection is how much local oscillator power is needed for shot noise limited operation. This power cannot exceed the saturation limit of the photodetector's RF amplifier, and this depends on the common mode rejection of the balanced receiver. The simulation allowed us to verify that the shot noise level exceeds the thermal noise by roughly 10 dB with 10 dBm input power, confirming that shot noise limited sensitivity can be achieved using this photodetector.

The sensitivity degradation of heterodyne detection must be attributed to losses arising from RF envelope detection and analog mixing. Compared to homodyne detection, heterodyne detection naturally suffers an extra 3dB loss due to IF down conversion and roughly 3 dB loss due to reduced quantum efficiency in the high bandwidth photodiode. There is also a 3dB loss by capping one arm of the 3dB coupler, because only a single high speed photodiode was available. Furthermore, there are other mixing products from second order harmonics that mix into the baseband waveform. This introduces some distortion and deteriorates the SNR. Compared to this, the self chirped setup alleviates all of these issues because the mixing process relies only on the square law response of the photodiode.

FIELD TRIALS OF THE SELF CHIRPED LIDAR SYSTEM

In the last section, an optical fiber was used as a placeholder for the free space optics in order to make comparisons between detection sensitivities. For the field test setup we incorporated a 5-inch diameter Newtonian telescope as shown in **Figure 5.9**. To make the system mobile all the hardware was migrated from the optics bench to a rolling cart as shown in the following figures.

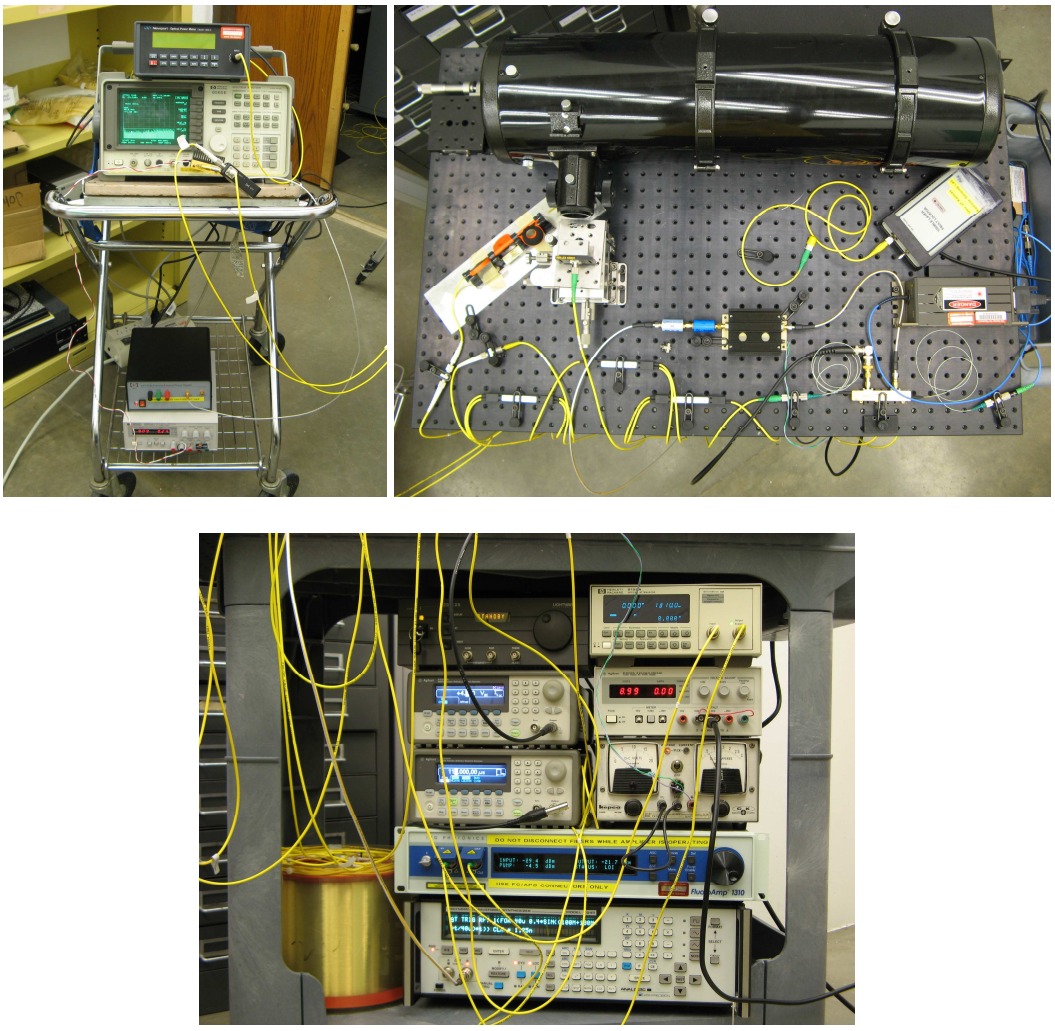


Figure 5.9 – Photos of the lidar cart.

For free space transmission the modulated optical signal is split into two parts by a 10-dB fiber splitter. One part (10%) goes through a praseodymium-doped fiber amplifier (PDFA) to boost the power to the telescope while the other part (90%) is used as the LO for detection. A polarization controller is used to adjust the polarization state of the LO and a 22.7-km single-mode fiber delay line is used to provide flexibility in adjusting the frequency of the beat signal, and to remain consistent with the lab bench setup. A 3-dB fiber coupler is used to separate the transmitted signal and the detected signal from the telescope and another 3-dB coupler is used to combine the optical signal with the LO. A balanced photodetector is used for homodyne detection and the dechirped RF signal is measured by an RF spectrum analyzer. An APC (angled physical-contact) fiber connector was mounted to the telescope for transmission and reception to minimize the Fresnel reflection from the fiber terminal. The position of the fiber connector was adjusted to focus the transmitted beam at the distance of the target and in this way the reflected power from the target was reciprocally focused back onto the open fiber end.

In the first trial, a sheet of white paper was used as the target which was placed 50 m away from the telescope. In this measurement, the PDFA was not used because the optical power from the transmitter was sufficient. The power reaching the target was found to be -14 dBm measured by a handheld power meter with the photodetector active area much larger than the beam size. Taking into account the telescope aperture r, spherical spreading of the reflected power at one way distance d and the further loss

L_{3dB} due to the 3-dB fiber coupler, the total power loss is about 67 dB according to the relation

$$\left(\frac{P_{returned}}{P_{transmitted}}\right)_{dB} = 10\log_{10}\left(\frac{\pi r^2}{4\pi d^2}\right) - L_{3dB} \qquad [5.3]$$

Using this, the signal power returned to the optical receiver was estimated to be approximately -81 dBm. The observed SNR at this power level was about 12 dB although it fluctuated over time. The power fluctuation of the dechirped signal is largely due to the well-known effect of carrier fading in homodyne detection, and a phase-diversity solution to this problem is presented in the next section. This measured -81-dBm receiver sensitivity (with 10-dB SNR) as shown in **Figure 5.7** is about 16 dB worse than the -97 dBm shot-noise limit. This discrepancy is mainly attributed to the coupling efficiency from the target to the single mode fiber, because misalignment and the secondary mirror obstruction reduce the power coupled back into the fiber.

A second trial was conducted using the concrete wall of a nearby building on campus as the target as shown in **Figure 5.10, 5.11**, where the one way distance is roughly 370 m. Due to the increased distance, the PDFA was used to boost the transmitted power to approximately 8 dBm. Unfortunately it was not feasible to determine the actual amount of power incident on the target under this setup. At this

distance it was still possible to achieve up to approximately 12 dB SNR, again with random fluctuation due to optical phase mismatch between the signal and the LO.

Figure 5.10 – Photo of the lidar aimed towards the target building.

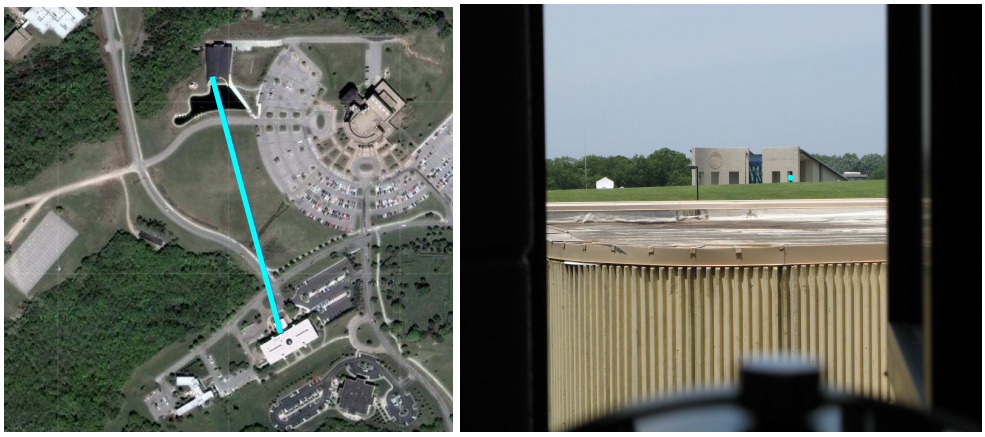


Figure 5.11 – Building target at 370 m range.

Image from Google Earth (left), photo from lidar position (right).

An aerial photo was measured to find the approximate distance of 370 m to the target, resulting in a predicted 72.3-MHz target beat frequency. The measured beat

frequency was 72.38 MHz indicating a true distance of 371.3 m. **Figure 5.12** shows the beat signal captured from the spectrum analyzer.

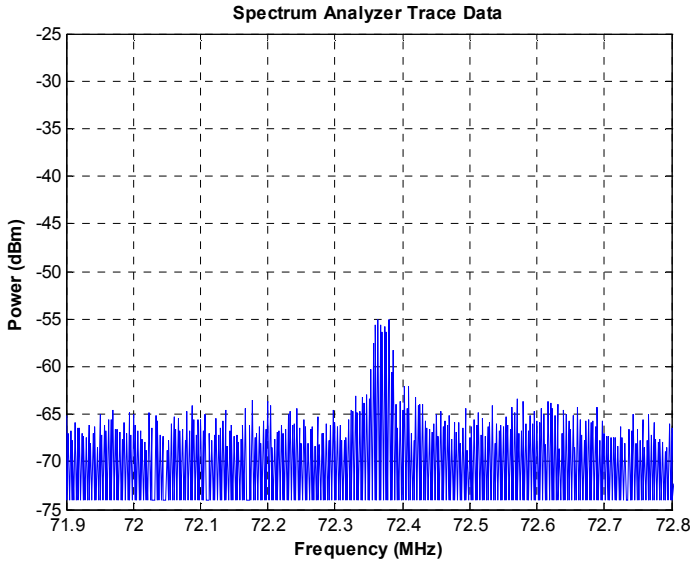


Figure 5.12 – Limestone target echo at 370 m.

(no averaging, 30-kHz res. BW, 500-ms sweep time over 2 MHz span).

PHASE DIVERSITY RECEIVER

While having many benefits, one potential weakness of the proposed homodyne scheme is the reliance on optical phase stability at the receiver. The optical carrier phase fluctuates naturally due to vibrations and turbulence, causing loss of signal product from the optical mixing process.

We observed significant random fading of the beat signal in our experiments due to optical phase fluctuations. To overcome this problem, we modified the receiver by inserting a 90º-hybrid optical coupler in place of the 3-dB coupler before the photodetector, and then measured the two photodiode output signals separately.

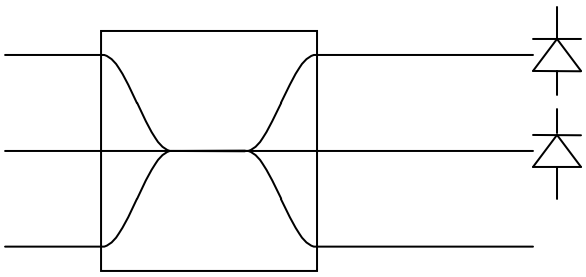


Figure 5.12 – Phase diversity receiver setup.

Whereas the scattering matrix of the ideal 3-dB coupler is

$$S_{3dB} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & j \\ j & 1 \end{bmatrix} \qquad [5.4]$$

the scattering matrix of the 90º coupler is

$$S_{hybrid} = \begin{pmatrix} \sqrt{0.2} & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) \\ \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.2} & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) \\ \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.2} \end{pmatrix} \qquad [5.5]$$

We connected two of the three ports, which gave the two photodiode signal currents

$$\begin{aligned} I_{arm1} &= \Re\sqrt{0.32P_{LO}P_{sig}}\,\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{LO}(t)-\theta_{sig}(t)-\tfrac{3\pi}{4}\right) \\ I_{arm2} &= \Re\sqrt{0.32P_{LO}P_{sig}}\,\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{LO}(t)-\theta_{sig}(t)+\tfrac{3\pi}{4}\right) \end{aligned} \qquad [5.6]$$

Thus the fading terms become complementary due to the ¾π phase offset (see the appendix for a derivation). Samples of these carrier signals were taken at roughly 1-s

intervals, producing the independently faded power measurements shown in **Figure 5.13.**

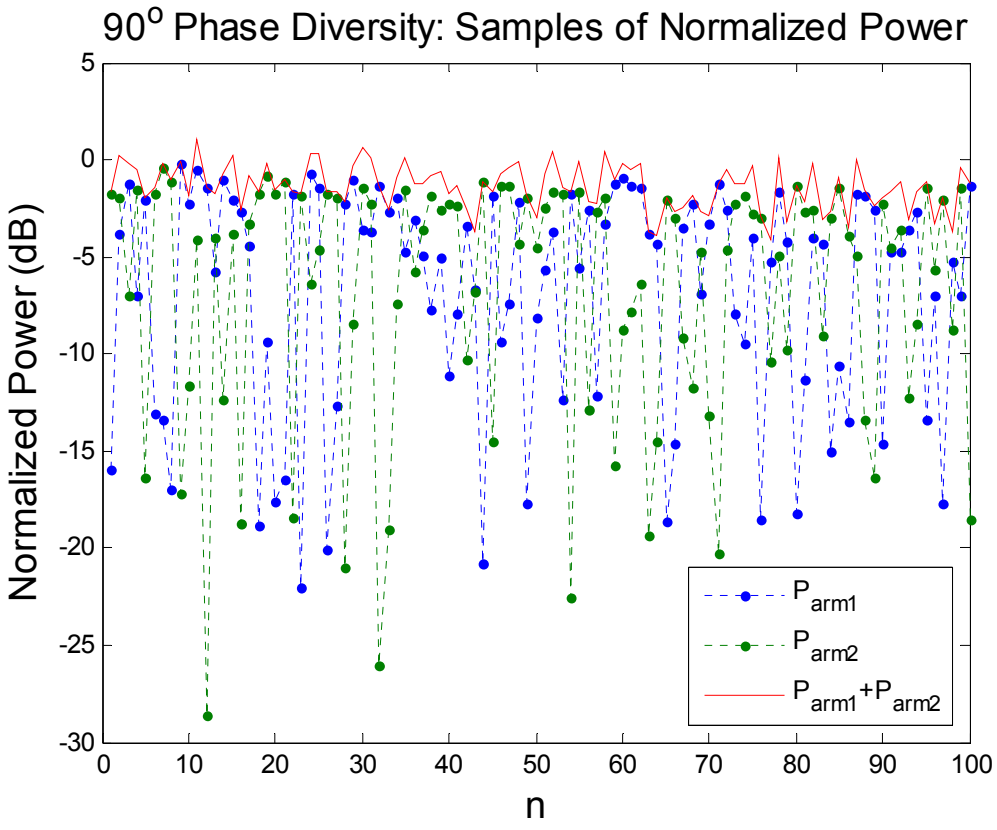


Figure 5.13 – Samples of normalized received power, individually and summed.

The fading of these powers P_{arm1} and P_{arm2} varied between the lower limit of the noise floor and the upper limit of maximum carrier power. The power samples were each normalized to a peak of 0 dB and summed, demonstrating in **Figure 5.13** that the combined power fading was reduced to approximately 5 dB.

Additionally, a minimum mean squared error (MMSE) weighting algorithm was used to minimize the variance of the summed powers, but no significant improvement

could be achieved. The remaining residual power fluctuation of 5 dB can be attributed to amplifier nonlinearity and random intensity noise (RIN). These experiments were done using a fiber optic channel instead of free space, and the properties of the phase variation would likely be different in free space conditions.

CHAPTER 6 – CONCLUSIONS

A new high performance lidar prototype using a simplified homodyne detection scheme has been developed. This system has demonstrated potential as a useful airborne altimeter for ice sheet measurement. Quantum noise limited sensitivity was achieved using off-the-shelf fiber optic components and a low peak power laser, reducing cost and increasing system lifetime. The detection method uses a chirped local oscillator to directly convert the received light into a range-indicating beat frequency signal. This significantly eases the bandwidth requirement of the photodetector, allowing for larger area photodiodes and higher photodiode responsivity. The chirped LO combined with homodyne detection eliminates IF down conversion and dechirping mixing, reducing complexity and avoiding the associated SNR degradation. The system concept was validated by assembling direct and coherent systems for comparison. The improved performance was proven in part by demonstrating a 30 dB sensitivity advantage compared to the comparable heterodyne system.

CHALLENGES

There are several challenges to overcome in achieving the best performance from this lidar system. Firstly, there are many factors that degrade the coherence of the returned light. Atmospheric turbulence, target speckle, target relative motion and laser line width contribute to randomization of the phase of the returned light and destructive interference of the signal. Air turbulence, target speckle as well as limited

telescope resolution work to spread the refocused spot distribution, resulting in lost signal power. There is a need for a method to deal with the distortion and spreading of the returning wavefront caused by all of these processes.

Another issue is Doppler shift, which may be significant due to small operating wavelength. The nature of the self chirped homodyne process is that Doppler shift causes rapid amplitude modulation on the beat signal. In this way, the carrier fading problem is closely linked to Doppler shift. In the event of relative motion between the lidar and a target, the Doppler shift overrides the slow carrier fading process, causing the beat signal to develop modulation sidebands or random spreading of the beat spectrum (this issue is discussed in the appendix). **Figure 6.1** shows a mapping of issues and possible solutions.

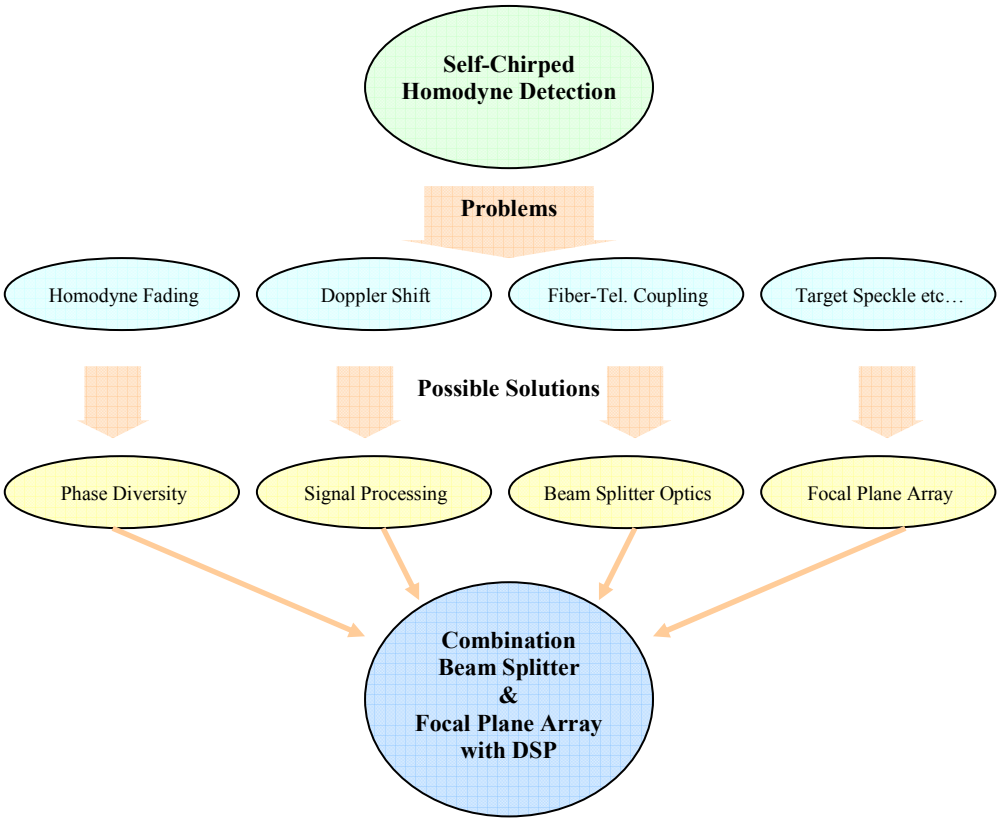


Figure 6.1 – Problems and possible solutions in the self chirped lidar.

The efficiency with which light can be coupled back into the telescope is limited by the very small aperture of single mode fiber. One promising alternative to fiber-telescope coupling is to instead use a beam splitter receiver as shown in **Figure 6.2**.

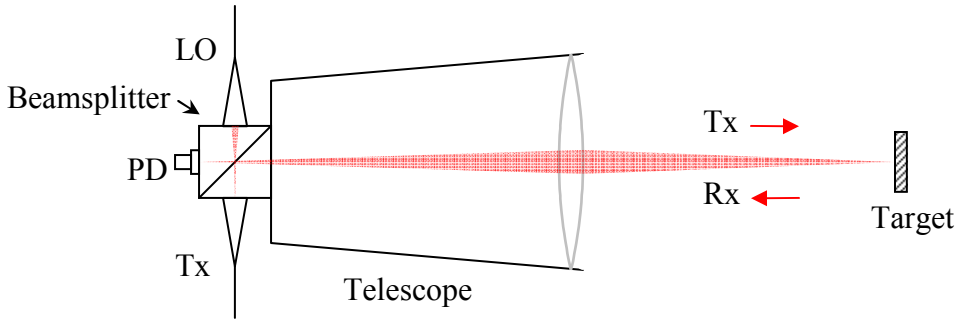


Figure 6.2 – Beam splitter optics concept.

Because the lidar has low RF bandwidth requirements, a large surface photodiode could be placed in the focused spot as shown. This would allow more optical power to be captured while performing the self chirped homodyne detection by injecting the local oscillator into the beam splitter cube. However, there are two limitations. Firstly, the design would need to incorporate a method for providing balanced detection. Balanced detection is necessary to eliminate the large unwanted DC component in the photodiode output. Secondly, the optical mixing process occurring on the face of a large photodiode may not produce more signal power. If the optical phase is not uniform across the photodiode surface, then destructive interference would occur as the response of the photodiode is integrated across its active area.

Instead of a large photodiode, a finely spaced focal plane array (FPA) combined with optimizing DSP could potentially solve or improve upon each of the mentioned difficulties. The FPA would contain multiple small photodiodes whose individual outputs could be combined optimally through DSP. The separate processing of pixels might allow recovery of information even if the optical wavefront is non coherent. A set of DSP hardware could be developed to optimize the combination of responses

from the pixels while simultaneously compensating for Doppler shift. This could be a promising direction as imaging arrays and DSP devices are currently seeing rapid advances in performance.

REFERENCES

CHAPTER 1

1. Climate Change 2007: The Physical Science Basis, Summary for Policymakers, Intergovernmental Panel on Climate Change (2007-02-05), retrieved from http://www.ipcc.ch/SPM2feb07.pdf, accessed 2007-06-19.

2. Instrumental_Temperature_Record.png, retrieved from http://en.wikipedia.org/wiki/Image:Instrumental_Temperature_Record.png, accessed 2007-06-19.

3. Burton, M.L., Hicks, M.J., "Hurricane Katrina: Preliminary Estimates of Commercial and Public Sector Damages." Marshall University Center for Business and Economic Research, September 2005.

4. Titus, J.G., and M.S. Greene, 1989, "An Overview of the Nationwide Impacts of Sea Level Rise. In The Potential Effects of Global Climate Change on the United States Appendix B: Sea Level Rise," Washington, D.C., Environmental Protection Agency.

5. Rignot, E. and Kanagaratnam, P., "Changes in the velocity structure of the Greenland Ice Sheet," 2006, Science 311: 986-990.

6. Zwally, H.J., Giovinetto, M.B., Li, J., Cornejo, H.G., Beckley, M.A., Brenner, A.C., Saba, J.L. and Yi, D., "Mass changes of the Greenland and Antarctic ice sheets and shelves and contributions to sea-level rise: 1992-2002," 2005, Journal of Glaciology 51: 509-527.

7. Lohoefener, A., “Design and Development of a Multi-Channel Radar Depth Sounder,” M.S. Thesis, Electrical Engineering and Computer Science, The University of Kansas, 2005.

8. Kanagaratnam, P., “Airborne Radar for High Resolution Mapping of Internal Layers in Glacial Ice to Estimate Accumulation Rate,” Ph.D. Dissertation, Electrical Engineering and Computer Science, The University of Kansas, 2002.

9. Chiing, A.W.Y., “Design and Development of an Airborne Stretch Radar for Depth Sounding the Jakobshavn Outlet Glacier,” M.S. Thesis, Electrical Engineering and Computer Science, The University of Kansas, 1998.

CHAPTER 2

10. Stone W.C., Juberts M., Dagalakis N, Stone J., Gorman J., “Performance Analysis of Next-Generation LADAR for Manufacturing, Construction, and Mobility,” NISTIR 7117, NIST, May 2004. (http://www.fire.nist.gov/bfrlpubs/build04/PDF/b04032.pdf)

11. GLAS Characteristics, retrieved from http://www.csr.utexas.edu/glas/, accessed 2007-07-01.

12. MGS MOLA Characteristics, retrieved from http://marsprogram.jpl.nasa.gov/mgs/scsys/molatext.html, accessed 2007-07-01.

13. MGS MOLA Specifications, retrieved from http://ssed.gsfc.nasa.gov/tharsis/spec.html, accessed 2007-07-01.

14. Allen, C., Y. Cobanoglu, S. K. Chong, and S. Gogineni, “Performance of a 1319 nm laser radar using RF pulse compression,” Proceedings of the 2001 International

Geoscience and Remote Sensing Symposium (IGARSS '01), Sydney, session SS52, paper Th03-04, July 2001.

15. Kingsley S. and Quegan S., Understanding Radar Systems, McGraw-Hill, New York, 1999.

16. Goldstein B.S., “Radar Range Equations for Narrow-Band Optical Sources”, MIT Lincoln Laboratory Technical Report, 1963, found in Selected Papers on Laser Radar, SPIE Milestone Series, Vol. MS 133, SPIE, Washington, 1997.

17. Peebles P.Z., Radar Principles, John Wiley & Sons, New York, 1998.

CHAPTER 3

18. Yariv A, Optical Electronics, Holt Rinehart and Winston, Inc., 1985.

19. Gerd K., Optical Fiber Communications, 3rd Ed., McGraw Hill, New York, 2000.

20. “A Survey of Methods Using Balanced Photodetection”, Application Note 14, New Focus.

21. Dawood, M., Hui R., “Experimental Study on Efficiency of 1310-nm Laser Coupling Between Schmidt-Cassegrain Telescope and Single Mode Fiber”, The University of Kansas Information and Telecommunication Technology Center (ITTC).

22. Winzer P.J., Leeb W.R., "Fiber coupling efficiency for random light and its applications to lidar," Opt. Lett. 23, 986-988, 1998.

APPENDICES

APPENDIX A. AMPLIFIER NOISE FIGURE ANALYSIS

Although every amplifier adds thermal noise, the first amplifier in a series adds the most noise relative to the signal power. A low noise amplifier (LNA) is thus a necessary component of noise analysis. Given the dB values of noise figure *NF* and gain *G,* these parameters are first converted into their linear coefficients as

$$F = 10^{\left(\frac{NF}{10}\right)}$$

$$G_n = 10^{\left(\frac{G}{10}\right)}$$

Given some *F* and the ambient temperature *T*, the equivalent noise temperature is

$$T_e = (F-1)\cdot T \quad \text{(K)}$$

The noise power spectral density (noise equivalent power) depends on the room temperature plus the additional equivalent temperature due to amplifier noise,

$$N_e = k\cdot(T+T_e) \ \text{(W/Hz)}$$

where *k* is Boltzmann's constant. Multiplying this by the bandwidth of the noise as

$$N_1 = N_e \cdot 2 \cdot BW \ \text{(W)}$$

gives the thermal noise power in Watts. The noise current would have the form of a scaled random variable

$$I_n(t) = \sqrt{\frac{N_1}{R_{in}}} \times N(0,1) \quad \text{(A)}$$

where R_{in} is the input impedance and $N(0,1)$ represents a random signal with Gaussian (normal distribution) with zero mean, variance 1 and bandwidth BW. To generate a current with these noise characteristics in simulation, a random Gaussian sequence should be generated at sample rate of $2BW$ with the appropriate scaling value. The current can then be converted to voltage and summed with the signal voltage as

$$V_{out}(t) = \sqrt{G_n} \cdot (V_{in} + I_n \cdot R_{in}) \text{ (V)}$$

where V_{in} is the input signal. This equation sums the incoming voltage and the equivalent input noise voltage and amplifies by the voltage gain to produce the output voltage signal.

APPENDIX B. REAL VS PHASOR NOTATION

If the E field is a real sinusoid, why do we use complex exponentials? This can be a source of confusion in understanding the meaning of analyses found in texts and literature. The reason is that phasor notation produces less tedious math and hence less tendency for errors [(18-ch.1)]. We know that the instantaneous power is proportional to the square of the instantaneous E field, and we can take the time integral of this to find the time averaged power. Alternatively, phasor notation provides a more simple way to obtain the time averaged power directly from the instantaneous E field. Demonstrating the two approaches is a good way to introduce the concept of optical mixing. Suppose that two light beams of different frequency θ_1 and θ_2 and amplitudes A and B are superimposed on the surface of a detector. This results in an interference beat note between the two fields and we want to know the power. Using the first method (real notation), we can square the value of the total instantaneous E field and then integrate over time to find the time averaged power:

$$P_{z,avg} \propto \int_{-\infty}^{\infty} \left(A\cos(\theta_1) + B\cos(\theta_2)\right)^2 dt$$

$$= \int_{-\infty}^{\infty} \left(A^2\cos^2(\theta_1) + B^2\cos^2(\theta_2) + 2AB\cos(\theta_1)\cos(\theta_2)\right) dt$$

$$= A^2 + B^2 + 2AB\cos(\theta_1 - \theta_2)$$

Using the second method of phasor notation, we can take the magnitude squared of the summed E fields, which is equivalent to taking the complex conjugate product, and we will directly obtain the time averaged square value.

$$P_{z,avg} \propto \left| Ae^{j\theta_1} + Be^{j\theta_2} \right|^2$$

$$= \left(Ae^{j\theta_1} + Be^{j\theta_2} \right)\left(Ae^{-j\theta_1} + Be^{-j\theta_2} \right)$$

$$= A^2 e^{j\theta_1} e^{-j\theta_1} + B^2 e^{j\theta_2} e^{-j\theta_2} + ABe^{j\theta_1} e^{-j\theta_2} + ABe^{-j\theta_1} e^{j\theta_2}$$

$$= A^2 + B^2 + 2AB\cos(\theta_1 - \theta_2)$$

Both approaches yield the same result, demonstrating that the two mathematical expressions are in fact equivalent. Despite the use of complex expressions, the propagating E field is a real sinusoid and both ways of expressing the E field describe the same thing. The complex phasor notation is simply more convenient for analysis.

APPENDIX C. 90 DEGREE HYBRID ANALYSIS

The 90 degree coupler scattering matrix is

$$S = \begin{pmatrix} \sqrt{0.2} & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) \\ \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.2} & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) \\ \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.2} \end{pmatrix}$$

Using only two inputs and two outputs, we can obtain

$$\begin{bmatrix} E_{pd1} \\ E_{pd2} \end{bmatrix} = \begin{bmatrix} \sqrt{0.2} & \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) \\ \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right) & \sqrt{0.2} \end{bmatrix} \cdot \begin{bmatrix} E_{LO} \\ E_{sig} \end{bmatrix}$$

$$E_{pd1} = \sqrt{0.2}E_{LO} + \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right)E_{sig}$$

$$E_{pd2} = \sqrt{0.4}\exp\left(j\frac{3\pi}{4}\right)E_{LO} + \sqrt{0.2}E_{sig}$$

Substituting the expressions for E_{sig} and E_{LO},

$$E_{pd1} = \sqrt{0.4\eta P_{LO}}\cos(m(t))\cdot e^{j(\omega t + \theta_{LO}(t))} + \sqrt{0.8\eta P_{sig}}\cos(m(t-\Delta t))\cdot e^{j\left(\omega t + \theta_{sig}(t) + \frac{3\pi}{4}\right)}$$

$$E_{pd2} = \sqrt{0.8\eta P_{LO}}\cos(m(t))\cdot e^{j\left(\omega t + \theta_{LO}(t) + \frac{3\pi}{4}\right)} + \sqrt{0.4\eta P_{sig}}\cos(m(t-\Delta t))\cdot e^{j(\omega t + \theta_{sig}(t))}$$

Using the Poynting vector, and putting in the received E field expressions,

$$P_{pd1} = \frac{1}{2\eta}\left|\sqrt{0.4\eta P_{LO}}\cos(m(t))\cdot e^{j(\omega t+\theta_{LO}(t))} + \sqrt{0.8\eta P_{sig}}\cos(m(t-\Delta t))\cdot e^{j\left(\omega t+\theta_{sig}(t)+\frac{3\pi}{4}\right)}\right|^2$$

$$P_{pd2} = \frac{1}{2\eta}\left|\sqrt{0.8\eta P_{LO}}\cos(m(t))\cdot e^{j\left(\omega t+\theta_{LO}(t)+\frac{3\pi}{4}\right)} + \sqrt{0.4\eta P_{sig}}\cos(m(t-\Delta t))\cdot e^{j(\omega t+\theta_{sig}(t))}\right|^2$$

We will have the following optical power densities incident on the photodiodes:

$$P_{pd1} = 0.2P_{LO}\cos^2(m(t)) + 0.4P_{sig}\cos^2(m(t-\Delta t)) + \sqrt{0.32P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)+\tfrac{3\pi}{4}\right)$$

$$P_{pd2} = 0.4P_{LO}\cos^2(m(t)) + 0.2P_{sig}\cos^2(m(t-\Delta t)) + \sqrt{0.32P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{3\pi}{4}\right)$$

The photodiode output signal current will be

$$I_{sig} = \Re\, A_{eff} P_{pd}$$

$$I_{sig1} = \Re\left(0.2P_{LO}\cos^2(m(t)) + 0.4P_{sig}\cos^2(m(t-\Delta t)) + \sqrt{0.32P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)+\tfrac{3\pi}{4}\right)\right)$$

$$I_{sig2} = \Re\left(0.4P_{LO}\cos^2(m(t)) + 0.2P_{sig}\cos^2(m(t-\Delta t)) + \sqrt{0.32P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{3\pi}{4}\right)\right)$$

Using a bandpass filter to reject the direct detection components and simplifying, we obtain

$$I_{sig1} = \Re\sqrt{0.32P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{3\pi}{4}\right)$$

$$I_{sig2} = \Re\sqrt{0.32P_{LO}P_{sig}}\cos(m(t))\cos(m(t-\Delta t))\cos\left(\theta_{sig}(t)-\theta_{LO}(t)+\tfrac{3\pi}{4}\right)$$

The signal will never go to zero in both branches at the same time. To demonstrate this, suppose that signal one is faded, which occurs when

$$\cos\left(\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{3\pi}{4}\right) = 0$$

$$\theta_{sig}(t)-\theta_{LO}(t)-\tfrac{3\pi}{4} = k\pi - \tfrac{\pi}{2}$$

$$\theta_{sig}(t)-\theta_{LO}(t) = k\pi + \tfrac{\pi}{4}$$

where k is an integer. In that case, the fading term in the second current will have the following phase:

$$\begin{aligned}
&\theta_{sig}(t) - \theta_{LO}(t) + \tfrac{3\pi}{4} \\
&= k\pi + \tfrac{\pi}{4} + \tfrac{3\pi}{4} \\
&= (k+1)\pi \\
&\cos((k+1)\pi) = \pm 1
\end{aligned}$$

This demonstrates that when one term is minimized the other is maximized and they are never simultaneously zero.

APPENDIX D. DOPPLER CORRECTION ANALYSIS

While the self chirped homodyne detection scheme provides high sensitivity with a very simple receiver arrangement, the loss of phase coherence of the received light may pose a significant problem in practical applications.

For idealized homodyne detection, the signal and LO should be perfectly matched in frequency and phase, assuring the generation of a single pure beat note. In practice this ideal homodyne mixing cannot be realized because the propagation time of the returned light is constantly altered by target motion and varying refractive index of the atmosphere. This implies that the relative frequency and phase of signal and LO vary randomly in time.

When the relative velocity is very small or zero, the system will exhibit gradual destructive interference (fading) of the beat signal. This is due to the signal phase falling in and out of synchronization with the local oscillator due to minute variations in the travel time of the return signal. A phase diversity (quadrature) receiver alleviates this problem; however such a system requires a 3-port hybrid coupler which is less efficient (smaller signal-LO product) and has reduced common mode rejection. Furthermore, this method cannot deal with the problem of Doppler shift.

We have observed that unless we allow the system to settle the vibrations and relative motion of the system cause rapid fading of the beat note. The mechanisms of fading and Doppler shift are related but occur on different time scales. Doppler shift may be beneficial because it changes the phase so rapidly as to modulate the beat

signal instead of slowly attenuating it. How can we manage or possibly utilize the Doppler shift?

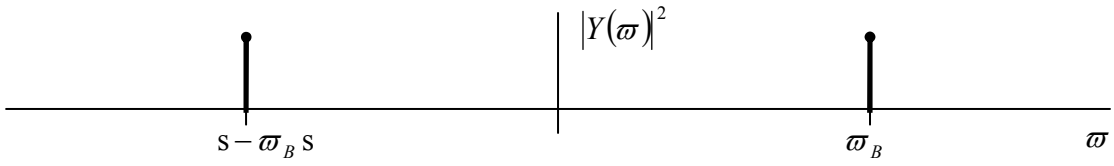
The Doppler shift is given by $\Delta\varpi = 2\pi v/\lambda$, where the wavelength is $\lambda = 1310\ nm$. First, the Doppler bandwidth must be constrained, because we require that $\Delta\varpi < \varpi_B$ to avoid ambiguity from negative frequency components. For example, we may assume a maximum bandwidth of 10 MHz, corresponding to a radial velocity of about 13 m/s. The receiver bandwidth of our lidar is 800 MHz so we can set a relatively high limit on the beat frequency ϖ_B, for instance let $\varpi_B > 2\pi(100MHz)$ so that we provide adequate headroom for the Doppler bandwidth.

The basic expression for the beat signal $y(t)$ is

$$y(t) = \cos(\theta_\Delta(t))\cos(\varpi_B t)$$

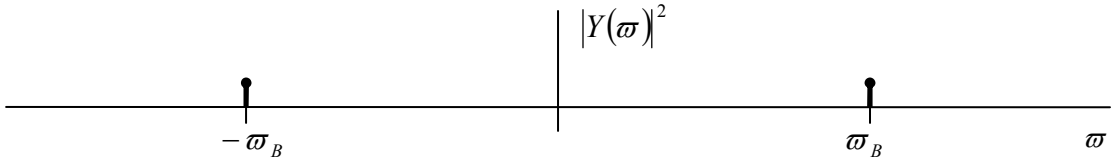
where $\theta_\Delta(t)$ is the relative phase between the signal and the local oscillator (LO). Firstly, under ideal conditions θ_Δ is a multiple of π so that the amplitude is maximized as shown below.

$$y(t) = \cos(\varpi_B t)$$



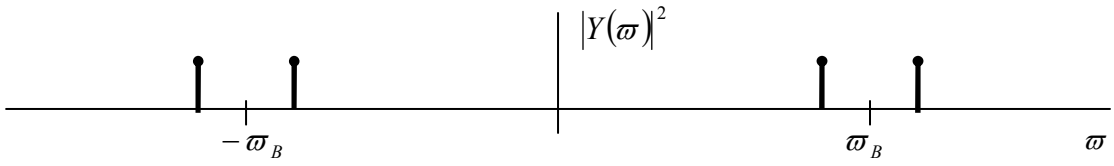
Now, when there is phase mismatch the fading term $\cos(\theta_\Delta(t))$ is equal to some value less than one, causing attenuation of the signal.

$$y(t) = \cos(\theta_\Delta(t))\cos(\varpi_B t)$$



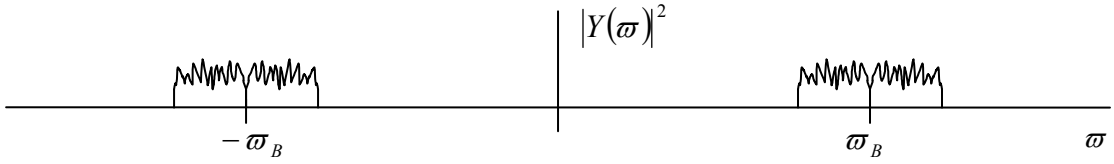
When there is a steady radial velocity the Doppler shift Δf will be induced and now the beat frequency will begin to change with time as well. The relative phase θ_Δ no longer causes fading because it is now summed with the phase change induced by Doppler shift.

$$y(t) = \cos(\Delta\varpi t + \theta_\Delta(t))\cos(\varpi_B(t)\, t)$$



However, when the Doppler shift is due to natural motion of the lidar system and the target, the Doppler shift will have a random distribution in time. For instance Doppler spreading can be a Gaussian random signal with a certain bandwidth *BW*.

$$y(t) = \cos(\Delta\varpi(t)t + \theta_\Delta(t))\cos(\varpi_B(t)\,t)$$



Because the Doppler changes rapidly, we cannot simply determine the exact Doppler shift and somehow compensate for it. Can we use the time-domain Doppler signal to somehow reverse the effect and recover the range information?

One approach might be to square the signal. This is analogous to double side band AM demodulation where the beat frequency ϖ_B acts like the carrier and the Doppler spreading acts like amplitude modulation. (We can simplify the expression of $y(t)$ for clarity. Although the frequencies are time dependent we can drop the formalism of adding *(t)* and we can ignore θ_Δ because Doppler shift is present.)

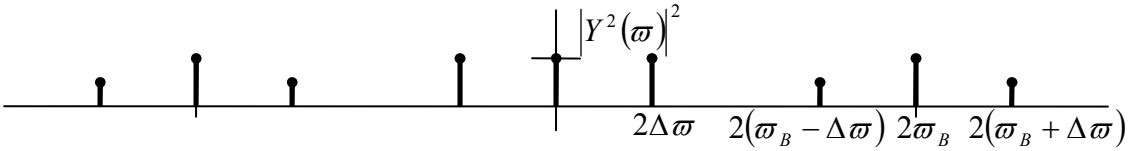
Given the signal

$$y = \cos(\Delta\varpi t)\cos(\varpi_B t)$$

If we square the signal we will obtain

$$y^2 = \cos^2(\Delta\varpi t)\cos^2(\varpi_B t)$$

$$= \tfrac{1}{2}\left[1 + \cos(2\Delta\varpi t) + \cos(2\varpi_B t) + \tfrac{1}{2}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{2}\cos(2(\varpi_B + \Delta\varpi)t)\right]$$



After squaring the signal we get one component that is independent of the Doppler shift,

$$\tfrac{1}{2}\cos(2\varpi_B t)$$

This should provide recovery of the beat frequency, however the noise level will be raised.

Furthermore, we now have the Doppler information converted into a baseband signal. Can we lowpass filter $y^2(t)$ and use the Doppler signal to improve detection? We can lowpass and bandpass filter the signal into two parts:

$$y_1 = \tfrac{1}{2}\left[\cos(2\varpi_B t) + \tfrac{1}{2}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{2}\cos(2(\varpi_B + \Delta\varpi)t)\right]$$

$$y_2 = \tfrac{1}{2}\left[1 + \cos(2\Delta\varpi t)\right]$$

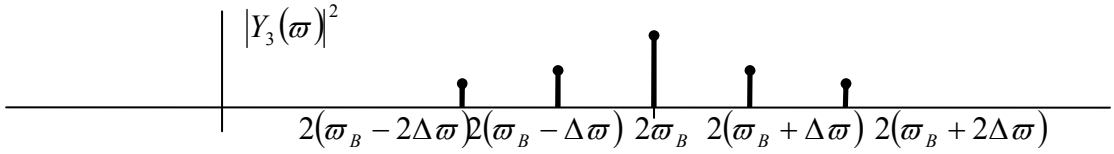
Multiplying these together we get:

$$y_3 = y_1 * y_2 = \tfrac{1}{4}\left[\cos(2\varpi_B t) + \tfrac{1}{2}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{2}\cos(2(\varpi_B + \Delta\varpi)t)\right] +$$
$$\tfrac{1}{4}\cos(2\Delta\varpi t)\left[\cos(2\varpi_B t) + \tfrac{1}{2}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{2}\cos(2(\varpi_B + \Delta\varpi)t)\right]$$

$$= \tfrac{1}{4}\cos(2\varpi_B t) + \tfrac{1}{8}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{8}\cos(2(\varpi_B + \Delta\varpi)t) +$$
$$\tfrac{1}{4}\cos(2\Delta\varpi t)\cos(2\varpi_B t) + \tfrac{1}{8}\cos(2\Delta\varpi t)\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{8}\cos(2\Delta\varpi t)\cos(2(\varpi_B + \Delta\varpi)t)$$

$$= \tfrac{1}{4}\cos(2\varpi_B t) + \tfrac{1}{8}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{8}\cos(2(\varpi_B + \Delta\varpi)t) +$$
$$\tfrac{1}{8}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{8}\cos(2(\varpi_B + \Delta\varpi)t) +$$
$$\tfrac{1}{16}\cos(2\varpi_B t) + \tfrac{1}{16}\cos((2\varpi_B - 4\Delta\varpi)t) +$$
$$\tfrac{1}{16}\cos((2\varpi_B + 4\Delta\varpi)t) + \tfrac{1}{16}\cos(2\varpi_B t)$$

$$= \tfrac{3}{8}\cos(2\varpi_B t) + \tfrac{1}{4}\cos(2(\varpi_B - \Delta\varpi)t) + \tfrac{1}{4}\cos(2(\varpi_B + \Delta\varpi)t) +$$
$$\tfrac{1}{16}\cos((2\varpi_B - 4\Delta\varpi)t) + \tfrac{1}{16}\cos((2\varpi_B + 4\Delta\varpi)t)$$



A portion of the $2\varpi_B$ carrier in this result originates from multiplication by the DC level of the signal. For this 1/4 fraction the noise contribution would be identical to the previous signal. However the multiplication also mixes a small portion (1/8) of the Doppler sidebands back into $2\varpi_B$ and this will introduce new signal with independent noise. If multiple targets were present then this process would introduce target ambiguity. However for our application only one target will be present.

There may also be a way to exploit the fact that the Doppler shift and the beat frequency itself are related. Integrating the frequency of the Doppler shifting should correspond with the changing of the beat frequency as the target moves relative to the lidar. It may be possible to compare the Doppler shift and the beat frequency to improve the range detection.

If a focal plane array were implemented in this lidar system, it might be beneficial to individually square the signals before summing. This could be an effective way to implement phase diversity while reducing the effects of target speckle and air turbulence. However, squaring the signal degrades the SNR on par with the envelope detection process required for heterodyne detection.