DEGREE OF POLARIZATION MEASUREMENTS USING DIGITAL PROCESSING

OPTIMIZATION OF DEGREE OF POLARIZATION OF PHOTOLUMINESENCE MEASUREMENTS USING DIGITAL SIGNAL PROCESSING

By

STEVEN JAMES LASCOS, B.Eng

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree

Master of Applied Science

McMaster University

MASTER OF APPLIED SCIENCE (2007) McMaster University
(Engineering Physics) Hamilton, Ontario

TITLE: Optimization of Degree of Polarization of Photoluminescence Measurements Using Digital Signal Processing

AUTHOR: Steven James Lascos, B. Eng (McMaster University)

SUPERVISOR: Professor D. T. Cassidy

NUMBER OF PAGES: ix, 120

Abstract

The experimental technique of measuring the degree of polarization of photoluminescence is studied. Digital signal processing techniques are applied to characterize the spectrum of the experiment and optimize its configuration for accuracy and precision. Digital phase sensitive detection is analyzed and the design of a digital lock-in amplifier is presented. Mechanical frequency modulation of the photoluminescence is reported and a physical model is presented. Procedures for enhancing measurements are stated with a reduction in noise levels of approximately 50% reported.

Acknowledgements

I would like to thank my thesis supervisor Dr. Daniel Cassidy whose enthusiasm for combining engineering development with academic research made this work possible. I would also like to acknowledge the previous work done by researchers at McMaster University in establishing the spatially-resolved degree of polarization technique.

In addition to those mentioned above, I would like to thank my parents, James and Terrill for their support over the years.

Lastly I would like to thank Brooke for providing motivation, inspiration and understanding whenever it was needed.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Introduction

This section gives a brief overview of the degree of polarization (DOP) experiment and provides context for my results. Detailed information on the fundamental principles of DOP measurements as well as the experimental apparatus and method are presented in Chapter 2. DOP is a technique for creating two-dimensional maps of the difference of the normal components of the strain and of the shear strain along the facet of a photoluminescent III-V semiconductor crystal by measuring the degree of polarization of the emitted light. Measuring the stress or strain has many industrial applications including but not limited to: device lifetime, reliability, manufacturing process effects and process control.

The motivation for this work was to design, implement and characterize a digital signal processing platform for enhancing and expanding the capabilities of the DOP technique as well as to provide a reusable platform for other researchers. The goal of this thesis is twofold. Firstly, to use the platform to understand the interaction between the DOP and the apparatus as well as to characterize experimental parameters for optimal performance. To my knowledge, no optimization methodology or parameter analysis has yet been published for the DOP apparatus. Secondly, to analyze and report novel techniques for digital lock-in amplifiers which will provide a new means for measuring polarization-resolved photoluminescence. The fundamental principles of lock-in amplifiers and phase sensitive detection are presented in Chapter 3.

When performing digital signal processing, the effects of quantization both in time (sampling) and finite precision must be understood and accounted for. The effect of quantization on digital phase sensitive detection is presented in Chapter 4. The design of a fully modular and upgradeable, platform-based digital lock-in amplifier is described in Chapter 5. The characterization of the digital system with respect to optimization for DOP experiments is reported in Chapter 6.

Reflection HeNe maps that are capable of micron resolution in resolving non-luminescent device structure and providing alignment information are presented in Chapter 7.

A complete spectral analysis of the DOP experiment and apparatus is offered in Chapter 8. A form of mechanical frequency modulation of the terahertz light wave discovered in the DOP photoluminescence is reported and a physical model presented. Finally, the noise and error of the system is explained in terms of optical component configuration and recommendations for optimal performance are provided. Conclusions are summarized in Chapter 9.

1.2 Background

It is known that mechanical strain in semiconductors changes the electrical and optical characteristics owing to a reduction of the crystal symmetry [1]. Operational device parameters that are very sensitive to crystal structure can be significantly affected when under mechanical strain. Crystal deformations can also alter the band gap causing changes in spectral output [2,3], and cause reduced device reliability [4]. Strain can be

introduced through a variety of manufacturing and packaging processes. Examples include, but are not limited to; defects in the original semiconductor wafer, epitaxial [5] and metallization [6] layers, dislocations [7], and die bonding [8]. Measuring strain can provide insight for improving device operating characteristics and reliability. Many techniques have been utilized to measure semiconductor strain including spectroscopy [9,10], Raman spectroscopy [11], x-ray diffraction [12], cathodoluminesence [13,14], electron-beam-induced current [15], and the degree of polarization (DOP) of photoluminescence [16]. DOP offers a simple, non-destructive method of measuring strain using low-power visible and infrared radiation and low-cost equipment.

1.3 New Contributions

To date, the DOP experiments reported in the literature have used a custom designed analog lock-in amplifier to perform phase sensitive detection of the output of the photodetector. Analog circuits are difficult to maintain and upgrade since a high level of electronic circuit design expertise is required. Conversely, a programmable digital implementation such as a Field Programmable Gate Array (FPGA) requires less advanced knowledge and expertise to create custom processing solutions versus the analog counterpart. Before working with FPGAs it is necessary to be familiar with basic electrical circuits, Boolean logic and software compilers. Designing an analog signal processor may require advanced knowledge in circuit theory including transmission-line effects and impedance matching. This thesis includes the novel design of a platform based digital lock-in amplifier. The platform is designed to be modular such that

researchers can reuse the system by replacing the digital signal processing (DSP) core with another function while recycling the system interface peripherals. While the FPGA based lock-in amplifier was designed for the DOP experiment, it is flexible enough for general purpose use.

Traditionally, the DOP experiments have been configured with optical chopper and rotating polarizer frequencies of roughly 1000 Hz and 200 Hz respectively. Typically the polarizer frequency was fixed and the chopper frequency was varied to find a setting that produced as little noise as possible. Certain frequencies in the vicinity of 1000 Hz would provide an acute increase in noise (as high as 300% increase) resulting in degradation of the measurement precision. This thesis performs a thorough investigation of the spectrum of the detector's output. The source of this noise is identified and procedures to correctly configure the system are presented.

In addition to the empirical formulas for optimizing the system, frequency modulation of the terahertz light wave is discovered and a theoretical model for its creation is introduced.

1.4 Summary

This thesis presents a research oriented, reprogrammable digital signal processing platform which enables both in-circuit and offline (host computer) processing of experimental data. Collected digital data is analyzed in MATLAB (an industry standard mathematical software tool) to apply spectral characterization towards enhancing knowledge of the experimental processes. A fundamental mathematical analysis of the

effect of quantization on digital phase sensitive detection is provided to quantify induced error and promote educated design decisions. The digital and spectral analysis has resulted in procedures for optimizing the experimental configuration to obtain a reduction in noise levels of approximately 50% versus previously reported data.

Chapter 2. Degree of Polarization of Photoluminescence

2.1 Theory

In an optically isotropic material an emitted photon should have no preference for any particular polarization. Thus, unstrained isotropic semiconductors should yield a degree of polarization (DOP) of zero. However, if the crystal is strained then one would expect to see a change in the polarization of luminescence that is some function of the strain field. It has been previously shown [17] that the DOP of photoluminescence is proportional to the difference in strain of two orthogonal directions of the luminescent region in GaAs or InP based materials. The DOP for light propagating in the y direction, where y is the normal to the sample surface under study can be generally defined as [17]

$$p = \frac{L_x - L_z}{L_x + L_z}, \tag{2.1}$$

where p is the degree of polarization for light propagating in the y direction and L_x and L_z are the magnitudes of luminescence of light polarized along the x and z directions respectively. In Eq. (2.1), p can vary between -1.0 to $+1.0$, where a magnitude of one indicates that the light is completely polarized in a single direction. A more accurate relationship accounting for the spectral nature of the photoluminescence and spectral response of the detector (for measurement) is [18]

$$DOP_y = \frac{\int_0^\infty [L_x(E) - L_z(E)] R(E) dE}{\int_0^\infty [L_x(E) + L_z(E)] R(E) dE}, \tag{2.2}$$

where $L_x(E)$ and $L_z(E)$ are the magnitude of luminescence between energy E and $E+dE$ that is polarized in the x and z directions respectively, and $R(E)$ is the responsivity of the detector at energy E.

In previous work [17] it has been shown that for small stresses (below $2\text{x}10^9$ dyn/cm^2) a linear dependence between DOP and strain is valid:

$$DOP_y = -C_\varepsilon(\varepsilon_{xx} - \varepsilon_{zz}), \tag{2.3}$$

where ε_{xx} and ε_{zz} are normal components of the 2nd rank strain tensor in the x and z directions, respectively. C_ε represents an experimentally determined constant of proportionality for the linear relationship. The strain vectors and corresponding tensor are shown in Figure 2.1. Note that the DOP in Eq. (2.3) is only a function of direct strain, and not a function of shear strain. Furthermore, DOP is not affected by direct strain along

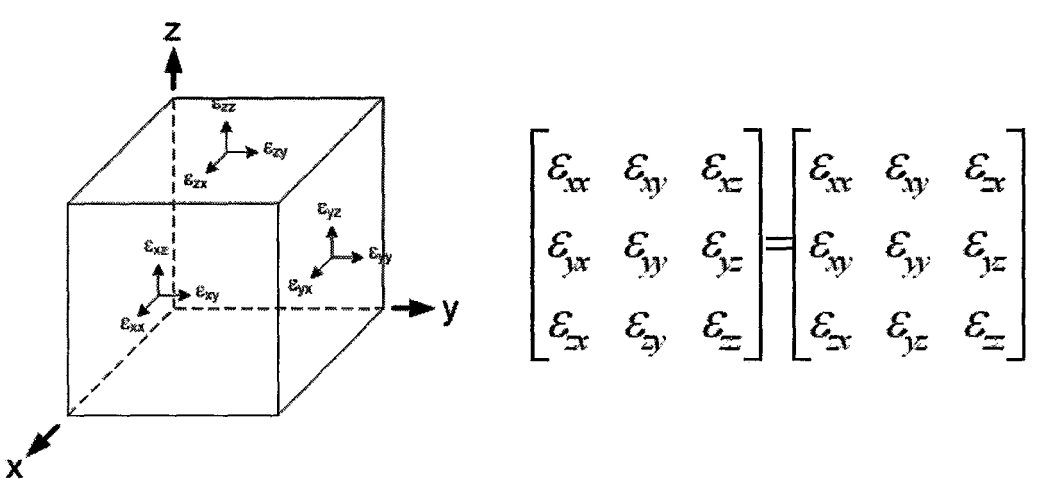


Figure 2.1 – (Left) Strain vectors on an infinitesimal cube. Vectors normal to a face of the cube are direct vectors; vectors parallel to a face are shear vectors. (Right) Strain tensor. The diagonal components are direct strains, the remainder are shear strains. The tensor is symmetric. I.e. ε_{xy} and $-\varepsilon_{yx}$ shear strains cause the same rotational distortion.

the direction of propagation, ε_{yy}.

It may be easier to experimentally measure or calculate stress rather than strain. Eq. (2.3) can be rewritten in terms of stress using the generalized three-dimensional version of Hooke's Law [19]. The general anisotropic form is shown below where the ε (stress) and σ (strain) vectors are reduced to six elements from nine since the 2nd rank tensors are symmetric.

$$\varepsilon_{ij} = s_{kl}\sigma_{ij}, \tag{2.4a}$$

$$\begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{yz} \\ \varepsilon_{zx} \\ \varepsilon_{xy} \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{21} & s_{22} & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & s_{32} & s_{33} & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & s_{43} & s_{44} & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_{55} & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_{66} \end{bmatrix} \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{zx} \\ \sigma_{xy} \end{bmatrix}, \tag{2.4b}$$

where s is the compliance matrix for anisotropic material. Cubic semiconductors like III-V materials are not isotropic in elasticity, but are orthotropic as a result of the crystal symmetry. This symmetry simplifies the anisotropic compliance tensor to the following form:

$$\begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{yz} \\ \varepsilon_{zx} \\ \varepsilon_{xy} \end{bmatrix} = \begin{bmatrix} \frac{1}{E_x} & -\frac{v_{yx}}{E_y} & -\frac{v_{zx}}{E_z} & 0 & 0 & 0 \\ -\frac{v_{xy}}{E_x} & \frac{1}{E_y} & -\frac{v_{zy}}{E_z} & 0 & 0 & 0 \\ -\frac{v_{xz}}{E_x} & -\frac{v_{yz}}{E_y} & \frac{1}{E_z} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2G_{yz}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2G_{zx}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2G_{xy}} \end{bmatrix} \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{zx} \\ \sigma_{xy} \end{bmatrix}, \tag{2.4c}$$

where $\frac{v_{xy}}{E_x} = \frac{v_{yx}}{E_y}$, $\frac{v_{zx}}{E_z} = \frac{v_{xz}}{E_x}$, $\frac{v_{yz}}{E_y} = \frac{v_{zy}}{E_z}$ due to tensor symmetry, $E_{x[y][z]}$ are Young's modulii, $v_{xy[zx][yz]}$ are Poisson ratios, and $G_{xy[zx][yz]}$ are shear (or rigidity) modulii. The factor of *2* in the shear modulii is due to the difference between shear strain and engineering shear strain γ, where $\gamma_{xy} = \varepsilon_{xy} + \varepsilon_{yx} = 2\varepsilon_{xy}$, etc. Since III-V materials are orthotropic, the elastic constants in Eq. (2.4c) are dependent on crystal orientation and choice of coordinate system. Therefore it is important to know which crystallographic plane corresponds to the sample cleavage plane so correct substitutions for elastic constants can be made for a given sample orientation.

From Eq. (2.4c) we can find that

$$\varepsilon_{xx} = \frac{1}{E_x}\sigma_{xx} - \frac{v_{yx}}{E_y}\sigma_{yy} - \frac{v_{zx}}{E_z}\sigma_{zz}, \qquad \varepsilon_{zz} = -\frac{v_{xz}}{E_x}\sigma_{xx} - \frac{v_{yz}}{E_y}\sigma_{yy} + \frac{1}{E_z}\sigma_{zz}, \tag{2.5}$$

Subbing Eq. (2.5) into Eq. (2.3) yields

$$DOP_y = -C_\varepsilon \left[\left(\frac{1}{E_x} + \frac{v_{xz}}{E_x} \right) \sigma_{xx} - \left(\frac{v_{yx}}{E_y} - \frac{v_{yz}}{E_y} \right) \sigma_{yy} - \left(\frac{v_{zx}}{E_z} + \frac{1}{E_z} \right) \sigma_{zz} \right], \tag{2.6a}$$

$$DOP_y = -C_\varepsilon \left[\left(\frac{1+v_{xz}}{E_x} \right) \sigma_{xx} - \left(\frac{1+v_{zx}}{E_z} \right) \sigma_{zz} \right] \quad for\ \sigma_{yy} = 0, \tag{2.6b}$$

which defines DOP for light propagating in the y direction in terms of stress components, the material elastic constants for when y is normal to the cleavage plane, and a proportionality constant. Since the facet under study is normal to the y-axis and exposed to free-space, no external force will be acting on the facet so σ_{yy} will be zero. Note that DOP in Eq. (2.6) is a function only of direct stress and not shear stress.

When dealing with stress analysis, if high precision is not required some accuracy can be sacrificed in simplifying the calculations by assuming the material is isotropic in elasticity. If we assume that $\frac{1}{E_x} = \frac{1}{E_y} = \frac{1}{E_z} = \frac{1}{E}$ and $v_{xy} = v_{zx} = v_{yz} = v$, then Eq. (2.6b) simplifies to

$$DOP_y = -C_\varepsilon \left[\frac{1+v}{E} \sigma_{xx} - \frac{1+v}{E} \sigma_{zz} \right] = -C_\varepsilon \frac{1+v}{E} (\sigma_{xx} - \sigma_{zz}), \tag{2.7}$$

where E is the average Young's modulus and v is the average Poisson's ratio for the material. The total error incurred through this simplification is expected to be a function of the difference between the actual values and assumed values for the elastic constants.

If the DOP is measured for a two-dimensional spatial array of points on a facet then by using Equations (2.3) or (2.6b) the difference in direct strain or direct stress fields at that surface can be calculated. It is clear from Eq. (2.7) that the DOP by itself provides

no information about the shear strain or shear stress fields, however. To obtain this information a slightly different measurement is required.

If the sample is rotated by 45° around the *y*-axis we can define the rotated degree of polarization (ROP) as [20]:

$$ROP_y = \frac{\int_0^\infty [L_{x'}(E) - L_{z'}(E)]R(E)dE}{\int_0^\infty [L_{x'}(E) + L_{z'}(E)]R(E)dE}, \tag{2.8}$$

where $L_{x'}(E)$ and $L_{z'}(E)$ are the magnitudes of luminescence with energy between E and $E+dE$ polarized along the x' and z' axis which are rotated 45° clockwise around the *y*-axis. Rotation of 45° counterclockwise around the *y*-axis will yield a negative value for ROP of the same magnitude. It has been shown [20] that a linear relationship between ROP and shear strain for small stresses (below $2\text{x}10^9$ dyn/cm^2) is:

$$ROP_y = 2C_\varepsilon \varepsilon_{zx}, \tag{2.9}$$

where C_ε is a positive, experimentally determined proportionality constant, which is equal to the constant in Eq. (2.3). and ε_{zx} is the tensor shear strain. Note that the ROP in Eq. (2.9) is dependent only on the shear strain that tends to deform the material tangentially around the *y*-axis. The ROP provides no information for shear strains that deform the material tangentially around the *x* or *z* axes.

Solving for ε_{zx} in Eq. (2.4c) gives:

$$\varepsilon_{zx} = \frac{\sigma_{zx}}{2G_{zx}}, \tag{2.10}$$

We can find ROP in terms of stress by substituting Eq. (2.10) into Eq. (2.9)

$$ROP_y = 2C_\varepsilon \varepsilon_{zx} = 2C_\varepsilon \left(\frac{\sigma_{zx}}{2G_{zx}} \right) = \frac{C_\varepsilon}{G_{zx}} \sigma_{zx}, \tag{2.11}$$

As mentioned above, III-V materials are orthotropic with respect to elasticity. If the assumption of an isotropic compliance tensor is made, then G_{zx} is replaced with G, where G is the average shear (rigidity) modulus for the material. The expected error in this simplification is proportional to the difference between the true shear modulus for the '*zx*' component and the assumed value.

The calibration constant C_ε will vary with composition for III-V alloys and with different crystallographic planes. It is important to note that p in Eq. (2.1) defines the degree of polarization without regard to angle between the polarization vector and the crystal orientation. To illustrate this, we see that maximum DOP or ROP will both give a p magnitude of unity, but are related to different strain components (direct and shear respectively). To resolve this, a measurement system must be calibrated such that a DOP_y of $+1$ is aligned with the proper orientation to the crystal such that Eq.(2.3) is valid. Figure 2.2 shows the relationship between DOP_y and ROP_y for the {110} plane. In [18] it was experimentally determined that $C_\varepsilon = 65$ +/- 10 for InP substrates cleaved in the {110} plane. All DOP/ROP experimental data presented in this thesis were obtained from {110} InP facets with the system calibrated as shown in Figure 2.2.

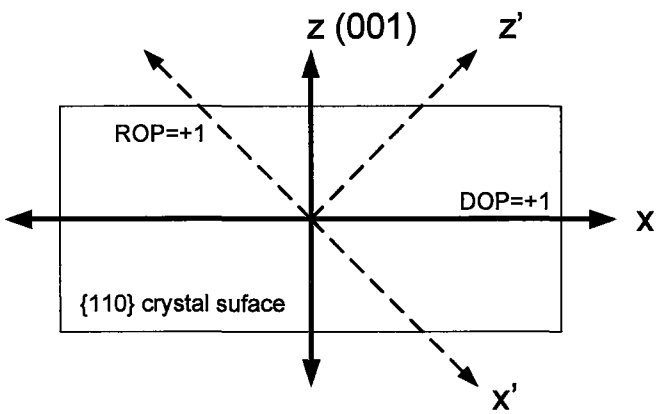


Figure 2.2 – DOP/ROP orientation diagram for InP {110} crystal. For Eq.(2.3) to be valid for InP, the system must be calibrated such that light that is 100% polarized along the x-axis yields a DOP of +1.0 for the crystal orientation indicated. The z-axis is defined to be the (001) direction when the y-axis (out of the page) is normal to the {110} plane. Recreated from [18].

Measuring DOP and ROP can map out the difference in strain components but not the actual component values themselves. However, the common factors inducing strain can produce characteristic 2D maps of DOP and ROP. DOP researchers [21] have performed finite element method simulations of a variety of strain inducing effects such that the simulated maps can be compared with measured data to obtain information about the strain fields and their cause.

2.2 Experimental Technique

To measure DOP and ROP, the apparatus shown in Figure 2.3 is used to excite photoluminescence from the sample under study. A HeNe laser is used as a pump source and focused onto the sample. The luminescence is collected by a 40× microscope objective (Newport M-40×) in a confocal arrangement and captured by a silicon p-i-n photodiode. The sample is mounted on a 3-axis translation, 3-axis rotation stage. During

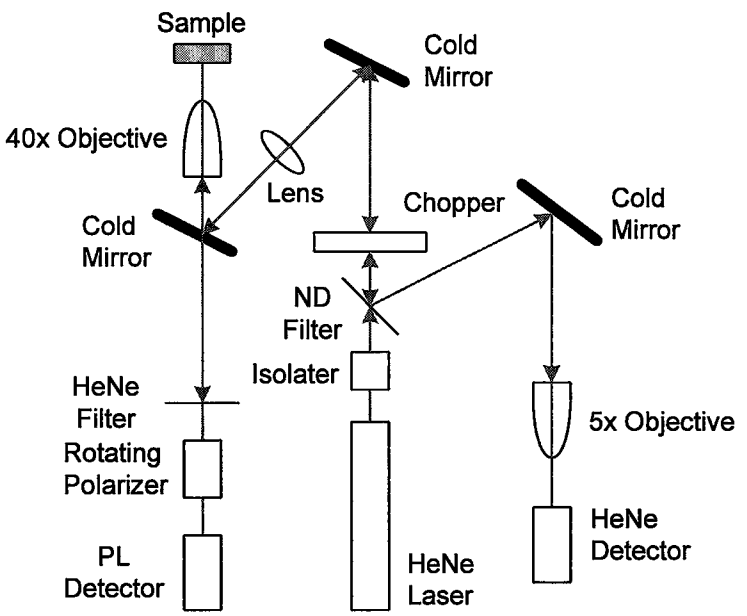


Figure 2.3 – DOP experimental apparatus. The HeNe pump laser excites the sample. Isolator prevents back reflected light from destabilizing the HeNe laser cavity. Chopper modulates pump light so total IR photoluminescence (PL) can by measured by phase sensitive detection. Cold mirrors reflect HeNe and transmit IR. Objective focuses pump light and collects PL. HeNe filter removes unwanted, reflected HeNe. Rotating polarizer modulates light so inphase and quadrature response yields DOP_y and ROP_y. A neutral density filter is placed between the chopper and laser in order to redirect HeNe light that has reflected from the sample. The reflected HeNe is collected by a second photodetector though a microscope objective.

a surface scan, the sample is translated in the *xz*-plane by Melles Griots nanomovers under computer control while the optical system remains fixed. The rotational sub-stage is used to align the crystal facet perpendicular to the optical beam. The emitted IR light has a broad spectrum whose energy must be greater than the bandgap of the material. For InP, the bandgap is 1.35 eV at 300K [22] which corresponds to a photon with a 920 nm wavelength and must be filtered to remove any reflected HeNe wavelengths (633 nm) before photodetection. The silicon photodetector aperture contains a 200 μm pinhole to

provide greater spatial resolution while the detector voltage response represents the time-averaged, squared amplitude of the optical signal. A filter (Melles Griot 03-FCG-111) is placed before the PL detector to remove the reflected HeNe from the IR signal of interest. Additionally, a neutral density filter (optical density = 0.1) is placed between the laser and chopper to guide some of the remaining reflected HeNe light to a second p-i-n photodetector with a 100 μm pinhole through a 5× objective. The reflection HeNe map allows us to see the device features (such as a ridge waveguide) that are independent of strain and PL and provides a more accurate method of checking sample alignment. When the sample facet is not perpendicular to the HeNe light beam, the PL and reflected HeNe will exhibit a linear ramp in magnitude as the beam scans in the x and z directions. This is caused by the increasing distance of the facet scanning point with respect to the focal plane of the microscope objective. In my research I have found that the reflected HeNe maps (see Chapter 7) are more sensitive to facet alignment, yet less sensitive to burns and minor scratches as compared to the corresponding PL maps. To achieve this precision, it is important to place the ND filter between the laser and chopper in order to spectrally isolate the modulated reflected light from the unmodulated (zero Hz) pump light in the detector output.

The voltage signal at the detector is the transmission of the rotating polarizer, amplitude modulated by the chopper which has a square wave modulating envelope:

$$V_s = \left(L_x \cos^2(\omega_m t) + L_z \sin^2(\omega_m t)\right) \times \left[\frac{1}{2} + \frac{2}{\pi}\sum_{n=0}^{\infty} \frac{\sin(2n+1)\omega_c t}{2n+1}\right], \tag{2.12}$$

where the first term on the left of Eq. (2.12) is the transmittance of the polarizer and the right side is the Fourier series of a square wave. The angle of the polarizer transmission axis is $\omega_m t$ where ω_m is the angular frequency of the rotating polarizer and ω_c is the angular frequency of the mechanical chopper. Note that the Eq. (2.12) assumes the chopper to be an ideal on/off modulator and the polarizer to be ideal with no material effects. The transfer functions of the chopper and polarizer are measured and analyzed in Chapter 8.

Lock-in amplification techniques can be used to measure the total photoluminescence $\int_0^\infty [L_x(E) + L_z(E)]R(E)dE$ by mechanically chopping the HeNe pump at approximately 1000 Hz and performing phase sensitive detection (PSD) at this frequency. Similarly, the photoluminescence is modulated by a linear polarizer (Polarcor optical glass polarizer) mounted in a rotating stage at approximately 200 Hz. Inphase detection at this frequency yields $\int_0^\infty [L_x(E) - L_z(E)]R(E)dE$, and the quadrature component yields $\int_0^\infty [L_{x'}(E) - L_{z'}(E)]R(E)dE$ in accordance with Figure 2.2. The ratio of inphase (or quadrature) response to total photoluminescence yields DOP (or ROP). In the past, a polarizing beam-splitter cube was used to separate L_x and L_z components, however precise alignment of two detectors is required. More recently, a rotating polarizer approach has been implemented requiring a single photodetector and permits simultaneous measurement of DOP and ROP which is generally independent of wavelength (850 – 1800 nm) [18].

Before running a scan, a linear polarizer with the transmission axis aligned with the x-axis (consistent with Figure 2.2) is placed before the rotating polarizer and the phase of the PSD reference signal is set such that DOP is maximized and ROP is zero. An experimental calibration constant is used to normalize the reading such that DOP is 100%. This experimental constant is further studied in Chapter 6. The time to complete a scan depends on the number of averages per spatial location, as well as the number of points to be scanned. A DOP standard deviation of less than +/- 0.1% has been previously achieved with 0.70 seconds per location on the analog system (300 ms wait time after motors stop, 200 averages). Using this configuration, a typical 101×40 μm area can be scanned in approximately 47 minutes. The enhancements achieved in this thesis obtain the same quality measurements (+/- 0.1%) in 0.28 seconds per location. The same area can be scanned in approximately 19 minutes, a reduction in scan time of 60%.

Chapter 3. Phase Sensitive Detection and Lock-in Amplifiers

3.1 Introduction

In scientific research, measurements involving physical systems are often converted to electrical form through sensors or transducers. Frequently, the signal of interest is buried in noise which may be several orders of magnitude stronger. If one can control the excitation of the signal of interest harmonically at a frequency f_0, then a bandpass filter centred at f_0 will pass the signal and reject the unwanted frequencies. The Q-factor of a filter is defined to be [23]

$$Q = \frac{f_0}{\Delta f}, \tag{3.1}$$

where f_0 is the filter centre frequency and Δf is defined to be the bandwidth at which amplitude response is $1/\sqrt{2}$, or –3 dB. Qualitatively, the Q-factor is a measure of how narrow the magnitude response is near the resonance frequency, where a high Q-factor (steep roll off) is desirable in the filter. In the transition band, a lowpass or highpass filter's magnitude response curve will always become $6n$ dB/octave or $20n$ dB/decade where n is the number of poles in the filter [24]. In filter theory, each pole increases the slope however it also increases the complexity of the system. A bandpass filter must be at least second order, it can be formed by cascading a first order lowpass and highpass filter. Thus for bandpass filters, n above is not the total number of poles in the filter, but is instead the number poles associated with a transition band.

Analog systems commonly use resistor-capacitor (RC) circuits and an op-amp to create a two-pole (second order) system which can then be cascaded to obtain higher

orders. Discrete-time (digital) finite impulse response (FIR) systems use taps to increase filter order, where each additional tap requires a multiplier and an adder. Instead of using a complex bandpass filter to extract the signal of interest, phase sensitive detection can be used to obtain very high Q-factors when combined with a lowpass filter in a configuration known as a lock-in amplifier.

3.2 Theory

Using the trigonometric identity for cosine addition/subtraction we can determine the effect of modulating (multiplying) one sinusoid by another.

$$A\cos(x+y) = A\cos x\cos y - A\sin x\sin y \tag{3.2}$$

$$A\cos(x-y) = A\cos x\cos y + A\sin x\sin y \tag{3.3}$$

Equations (3.2) and (3.3) are common identities. Adding Equations (3.1) and (3.2) and dividing by 2 yields:

$$\frac{A}{2}\cos(x+y) + \frac{A}{2}\cos(x-y) = A\cos x\cos y \tag{3.4}$$

From Eq. (3.4) we see that multiplying two cosines of different angles yields two new harmonics with sum and difference angle generation and amplitude $A/2$ where A is the product of the amplitudes of the original cosines.

We can rewrite Eq. (3.4) where angles x and y vary harmonically with time. Let the first cosine have angle $(\omega_1 t+\phi_1)$ and amplitude A, and the second have angle $(\omega_2 t+\phi_2)$ with unit amplitude where ω and ϕ are the angular frequency and initial phase, respectively.

$$\frac{A}{2}\cos[(\omega_1+\omega_2)t+(\phi_1+\phi_2)]+\frac{A}{2}\cos[(\omega_1-\omega_2)t+(\phi_1-\phi_2)] = A\cos(\omega_1 t+\phi_1)\cos(\omega_2 t+\phi_2) \tag{3.5}$$

When $\omega=\omega_1=\omega_2$ the left side of Eq. (3.5) becomes,

$$\frac{A}{2}\cos[2\omega t+(\phi_1+\phi_2)]+\frac{A}{2}\cos(\phi_1-\phi_2) \tag{3.6}$$

If Eq. (3.6) is low-pass filtered such that the 2ω term is removed we have,

$$\frac{A}{2}\cos(\phi_1-\phi_2) \tag{3.7}$$

Eq. (3.7) is a time-independent value that varies only with the phase difference of the original harmonics. When the phase difference is 90°, the result is zero and when the phase difference is zero, the result is maximized. Thus, if we modulate a narrowband signal of interest at frequency ω by a unit-amplitude reference-harmonic of the same frequency and ideally low-pass filter out the 2ω sum harmonic, we obtain a result that is dependent only on the amplitude of the signal of interest and the phase difference. Equations (3.6) and (3.7) are the mathematical description of a phase sensitive detector. When the detector is used as a narrow bandpass filter, it is called a lock-in amplifier. A schematic of a typical lock-in amplifier is shown in Figure 3.1.

Using a frequency domain analysis, we see from Eq. (3.5) that a lock-in amplifier will frequency-shift the components around a reference frequency ω_0 down to a centre frequency of zero, where the magnitude of those shifted components is dependent on the phase match condition, as well as the magnitude response of the low-pass filter following the amplifier. Replacing an n^{th} order bandpass filter with a phase sensitive detector and

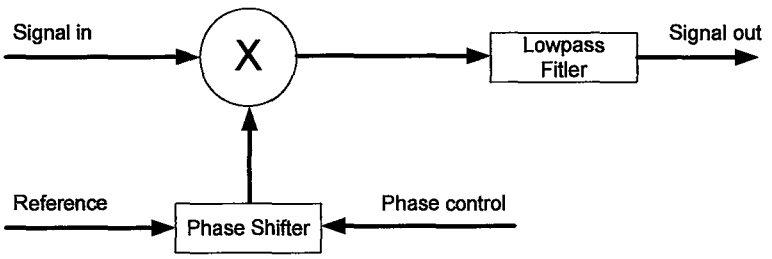


Figure 3.1 – Functional block diagram of a conceptual lock-in amplifier. The reference is a sinusoid at the same frequency as the signal in. The multiplier will create sum and difference frequencies. The phase control is used to maximize the signal out of the low-pass filter which extracts the zero Hertz component. When the signal out is maximized, the signal in and reference are phase matched. By monitoring the signal out and adjusting the phase control, phase sensitive detection is achieved.

an n^{th} order lowpass filter will significantly increase the effective Q-factor while providing a tunable architecture.

This can be illustrated by further considering Eq. (3.1). A second order (two pole) bandpass filter centred around 1,000 Hz will have a magnitude response of −6 dB at 2,000 Hz, a frequency difference of one octave. If a lock-in amplifier is employed with a reference signal of 1,000 Hz and a second order lowpass filter with a cutoff of 10 Hz, then a magnitude response of −12 dB at 20 Hz will be achieved. We can obtain a much narrower passband with the lock-in amplifier. Figure 3.2 illustrates the effective Q-factors and bandwidths of these two scenarios.

The result of a lock-in amplifier configuration is it effectively applies the narrow bandwidth of the lowpass filter up at the higher reference frequencies creating a large effective Q-factor. An important consideration when designing the lowpass filter for a lock-in is the inverse relationship between filter bandwidth and time constant. Abrupt step changes in the amplitude of the signal of interest will not be immediately available at

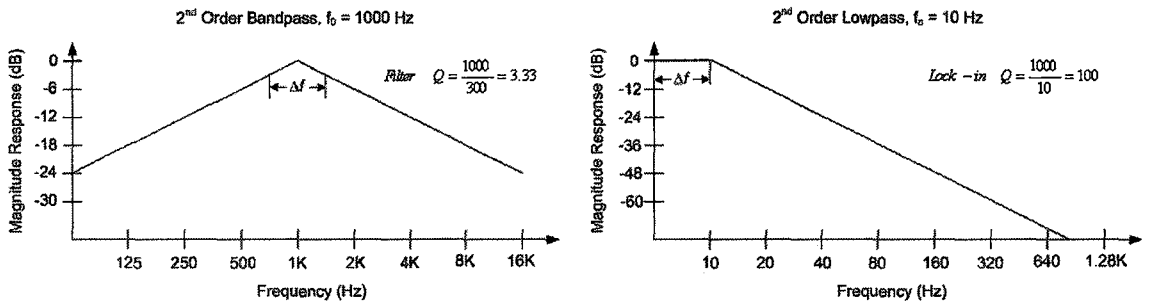


Figure 3.2 – Magnitude response of a bandpass filter versus lock-in amplifier. Left) 2nd order bandpass filter, centred at 1000 Hz with a bandwidth of approximately 300 Hz. Right) 2nd order lowpass filter with a cutoff (bandwidth) of 10 Hz.

the filter output due to its lowpass nature. This sets an upper limit on how fast the sample can be moved in a continuous scan or the settling time you must wait during a stepped scan.

Both analog and digital implementations of the lock-in amplifier shown in Figure 3.1 are possible. Analog configurations typically consist of a preamplifier, analog mixer to perform multiplication, a PLL circuit to perform generation and phase-shifting of a reference signal, and a low-pass filter to extract the DC output signal. Often, the analog multiplier is replaced with a switched multiplier. In this simplified configuration, the general purpose two-channel mixer is replaced by a simpler circuit which multiples the input signal by +1 for half a period and −1 for the second half period. The reference signal is therefore a square wave with 50% duty cycle at the frequency of interest. As a result, the square wave sets the centre frequency f_0 of the system and the phase sensitive detector will also extract all the odd harmonics of the fundamental, $3f_0$, $5f_0$, $7f_0$, etc. The extraction of odd harmonics is due to the Fourier spectrum of a square wave having strong components at odd harmonics of the fundamental frequency [25].

The performance of an analog lock-in is critically dependent on the noise, offsets, stability and temperature variation of the various components. Expert knowledge of analog circuit design may be required to obtain signal-to-noise-ratios useful for photon experiments. As a result, many researchers choose to purchase commercial units, sacrificing cost and the flexibility of a custom design.

Digital implementations offer many advantages over analog designs, but also present significant drawbacks and challenges. A basic digital lock-in amplifier schematic is shown in Figure 3.3. The preamplifier boosts the signal to obtain greater dynamic range before sampling. The anti-aliasing (lowpass) filter is designed to suitably attenuate input signal frequencies above half the sampling rate to satisfy the Nyquist criterion. The analog-to-digital converter (ADC) produces a binary coded value representing an

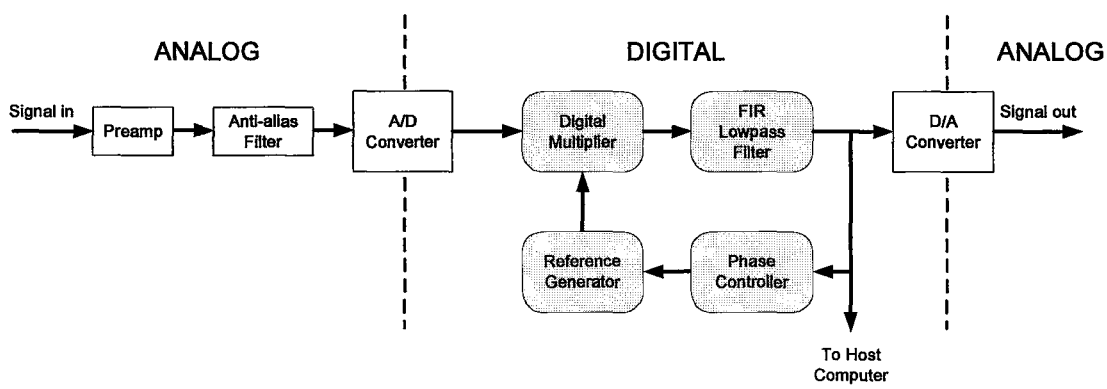


Figure 3.3 – Functional block diagram of a digital lock-in amplifier. The preamp boosts the signal to efficiently use the A/D converter's code range. The anti-aliasing filter is necessary to satisfy the Nyquist ratio and prevent high frequencies being aliased as low frequencies. The A/D converter uses a sample-and-hold technique to convert the voltage to a binary coded value. The multiplier creates sum and difference frequencies. The reference generator and phase controller create a frequency and phase matched reference signal. The FIR filter extracts the DC component. The optional D/A converter produces a voltage signal proportional to the DC component.

approximation of the input voltage during the sampling period. The reference generator provides the binary code values for the reference signal. The digital multiplier performs signed binary multiplication of the input and reference sample values. The finite impulse response (FIR) lowpass filter attenuates high frequencies and provides the output mean. The output can be sent to a host computer for data processing or to a digital-to-analog converter (DAC) to provide the output as a real-time voltage.

A digital implementation can significantly reduce noise by limiting the number of active analog components. Once the analog-to-digital step is performed, the remaining functions are completely impervious to common analog problems such as noise and error associated with electrical offsets, drifts, phase distortion, nonlinearities and temperature variation. However, all digital signal processing systems are subject to quantization error both in signal magnitude (round off from finite number of bits) and in time (sampling rate). The system designer must understand the impact of quantization error and ensure it is below an acceptable threshold. The analysis of quantization error in digital phase sensitive detection is presented in Chapter 4, the architecture of a novel digital lock-in amplifier is presented in Chapter 5, and its performance characterized in Chapter 6.

Chapter 4. Quantization Error in Phase Sensitive Detection

4.1 Introduction

A significant advantage of digital phase sensitive detection is its total immunity to errors (excepting the preamplifier and ADC functions) caused by voltage offsets, phase drifts, gain error, and other analog electrical effects caused by non-linearities and variation of voltage and temperature. However, quantization errors are introduced by practical limitations of finite sampling rate and precision. Sampling frequency and bits per sample are typically decided based on tradeoffs between circuit speed, resources, cost, and analog/digital converter availability. In order to have an optimal design while meeting device requirements the error induced by sampling rate and finite precision must be known for the particular application. Fortunately, this error is predictable and deterministic and can therefore be arbitrarily minimized by choosing a sufficient hardware implementation. Once the error analysis is complete, the design problem can be approached several ways. The maximum tolerable error can be specified and the resulting hardware requirements established, the available hardware resources can be determined and the error performance predicted, or a combination of both. For the DOP experiment we wish the overall error to be less than 0.1% DOP_y since this has been the precision previously achieved in the lab. However, it is difficult to determine exactly how much quantization error this corresponds to since the DOP is calculated as the ratio of the polarizer and chopper responses. Recall from Eq. (2.2) that the DOP_y is the polarizer response divided by the chopper response (for this argument we will assume the calibration constant is unity). Typically, a DOP_y maximum of only a few percent in

magnitude is found in the InP samples used for the research of this thesis. At 100% DOP_y, a 1% error in polarizer response yields a 1% error in DOP_y and at 10% DOP_y, a 1% polarizer error results in approximately 0.1% DOP_y error. The same argument applies to the chopper response (PL). At 100% DOP_y, a 1% error in the chopper response results in a 1% error in DOP_y. At 10% DOP_y, a 1% error in chopper yields a maximum error of approximately 0.1% DOP_y. Thus a 1% quantization error in both the polarizer and chopper responses yields a maximum 0.2% error in DOP_y in the worst case. Thus to ensure the quantization error does not exceed 0.1% DOP_y for samples with maximum 10% DOP_y, we must achieve a quantization error of less than 0.5%.

4.2 ADC Primer

In any digital signal processing system the continuous time signal must be quantized into a discrete time representation. This is carried out by an analog-to-digital converter typically through a sample-and-hold [26,27] and successive approximation [28] process. Before considering quantization error, it is worthwhile to consider the mechanism of this final analog process. Sample-and-hold is typically performed by storing the input voltage on a capacitor at the designated sampling instant, and then the input to the capacitor is disconnected while the voltage-to-binary conversion takes place through successive approximation. No switching can be performed in zero time (sampling aperture time), and no real capacitor can have zero leakage current (hold voltage droop), so the faster the conversion takes the place, the more accurate the result. Devices capable of fast conversions can be advantageous when used to sample at a

frequency much higher than the minimum required by the Nyquist rate. The Nyquist rate is defined to be twice the frequency of the fastest varying component in the input signal [29]. It is critical in a digital signal processing that this criterion be satisfied; failure to do so will erroneously alias high frequency components into low frequencies. If these aliased signals fall within the bandwidth Δf of the lock-in amplifier, significant measurement error can result. Oversampling has been used [30] to spread the fixed noise power of the converter itself over a wider frequency range, thus providing better signal-to-noise ratio in the band of interest and ease the requirements on the preceding anti-aliasing filter. It is beneficial to use the fastest ADC available even for low frequency sampling applications since the leakage of the capacitor is reduced and the option to implement oversampling is available.

4.3 Sampling Rate and Quantization Error

The Nyquist theorem states that if the sampling rate is twice that of the fastest varying component, then all the information of the continuous time-signal is contained within the discrete time representation (assuming infinite precision) [29]. However, when two separately discrete signals are combined in an operation such as phase sensitive detection, there may be additional considerations. The digital multiplier in Figure 3.3 computes the product of two discrete-time signals by multiplying their sample values for each time step. This is described mathematically as:

$$psd[n] = signal[n] \times reference[n], \quad n \geq 0 \tag{4.1}$$

where n is an integer index into the discrete sequence, *signal[n]* and *reference[n]* are the discrete sample sequences of the input signal and reference signal respectively. The output of the multiplier is the sequence *psd[n].*

Recall that our definition for phase sensitive detection from Eq. (3.6) describes sum and difference frequency generation but to accurately measure the input signal's amplitude both frequency and phase must be matched. Sampling a monotone sinusoid of frequency ω at a rate of 2ω allows us to determine the frequency and phase information in accordance with the Shannon Sampling Theorem but restricts any phase-shift operation (sample n to $n\pm1$) to a multiple of π radians. Given that we want to measure the amplitude of the input frequency of interest, any phase mismatch between the multiplier inputs in Eq. (4.1) will create an error in this measurement, always lesser in magnitude than the true value (which occurs at equation maximum) in accordance with Eq. (3.7).

The effect of finite a sampling rate can be determined by taking the average value of Eq. (4.1) over one period of the reference signal wave when the input signal is frequency matched.

$$\overline{psd}[n] = \frac{1}{n_s}\sum_{n=0}^{n=n_s-1} signal[n] \times reference[n], \tag{4.2a}$$

$$\overline{psd}[n] = \frac{1}{n_s}\sum_{n=0}^{n=n_s-1} A\cos[\omega n] \times \cos[\omega n] = \frac{A}{2}, \tag{4.2b}$$

Eq. (4.2a) calculates the mean value (or DC component) from the phase sensitive detection of *signal[n]* and *reference[n]* with a sampling rate of n_s samples per period. Eq. (4.2b) is evaluated for frequency and phase matched signal and reference cosine waves,

where A is the amplitude of the input signal and ω is the angular frequency in radians per sample. In Figure 4.1, Eq. (4.2b) is plotted for $\omega = 2\pi/n_s$, where n_s is evaluated for even numbers from 2 to 256 and the initial angle of the first sample is zero. For $n_s = 2$ the average value is A, but for a phase difference of zero we should get $0.5A$ in accordance with Eq. (3.7). For integer sampling rates greater than two, the result is $0.5A$ and the error is zero. We conclude that for frequency and phase matched discrete sinusoids at frequency ω, the error in phase sensitive detection is zero when the sampling rate is $> 2\omega$. For the $n_s = 2$ case, the PSD average is plotted in Figure 4.2 by varying the initial phase

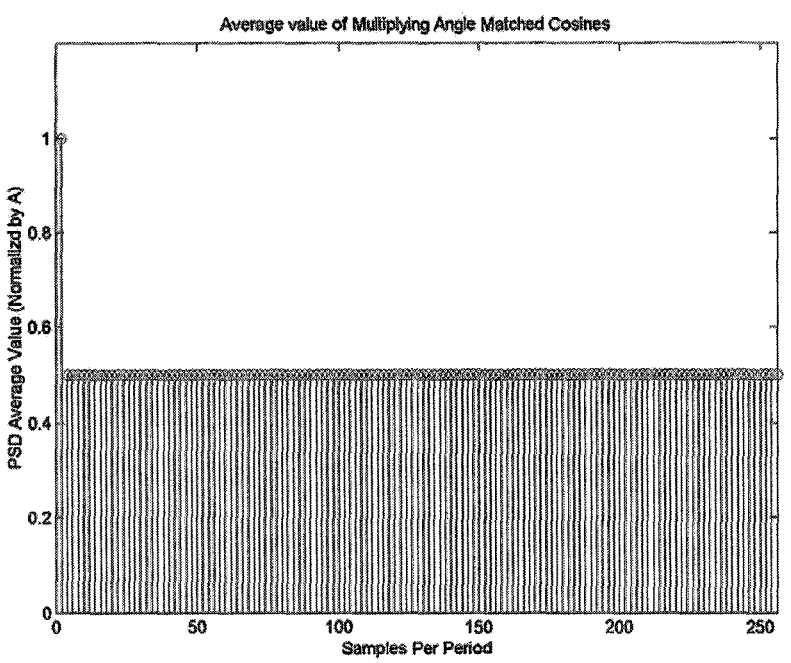


Figure 4.1 – Stem-plot of phase sensitive detection for frequency and phase matched cosines versus samples per period. Equation (4.2b) is evaluated for integer values of n_s from 2 to 256. At $n_s = 2$, the average value is A. For all sampling rates greater than or equal to 3, the average value of 0.5A exactly matches the continuous-time result. Powers of two are often selected to simplify any divide operations to binary shifts in hardware.

of the first sample from zero to 2π radians. We can see that in this case the average value is dependent on the initial sample phase and is only correct for initial phases $\pi/4$, $3\pi/4$, $5\pi/4$ and $7\pi/4$ radians. Since it is difficult to guarantee the phase angle of a sample, a sampling rate of 2ω is not sufficient and oversampling must be used.

The analysis above is based on the reference signal being generated such that the input sinusoid angle at the sampling instants is guaranteed to be at the same angle for the reference signal samples, hence the phase match condition. Usually this will not be the case, however we can easily 'phase-tune' our reference signal by offsetting its index n to

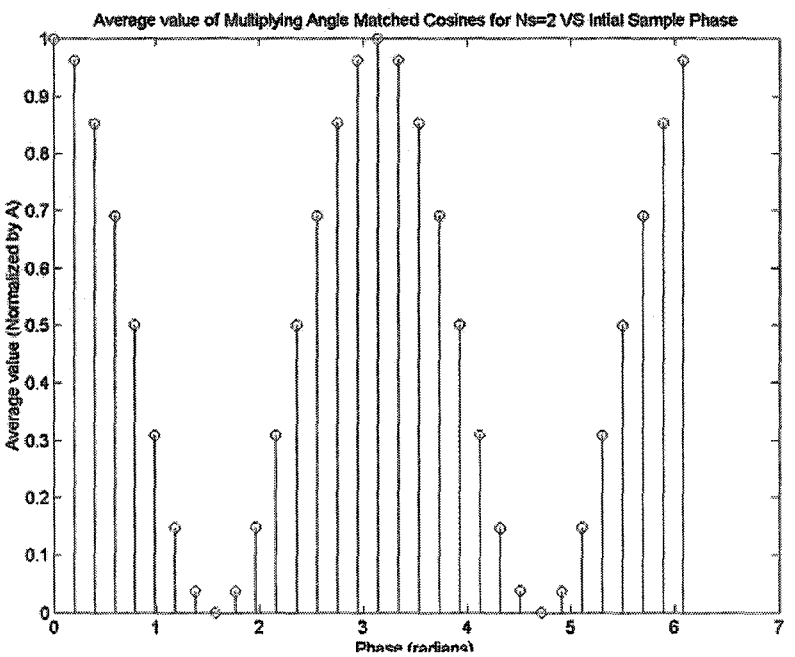


Figure 4.2 – Stem-plot of phase sensitive detection for frequency and phase matched cosines at two samples per period versus initial phase. For two samples per period, the average value of phase sensitive detection of two angle matched cosines is a cosine function of the initial sample phase. When an even number of samples per period is desired, at least four samples per period should be used to eliminate the initial phase dependence.

reduce the phase mismatch. Unfortunately, this is only guaranteed to get the phase angle difference to less than (assume less than or equal to) one sampling angle, therefore we must consider the error caused by the worst case scenario. For this case, at n_s samples per period the maximum phase mismatch is $2\pi/n_s$ radians. Figure 4.3 plots Eq. (4.2b) and % error (where error is deviation from the continuous-time value of 0.5) versus samples per period. We can see that as n_s is increased, the error rapidly approaches zero

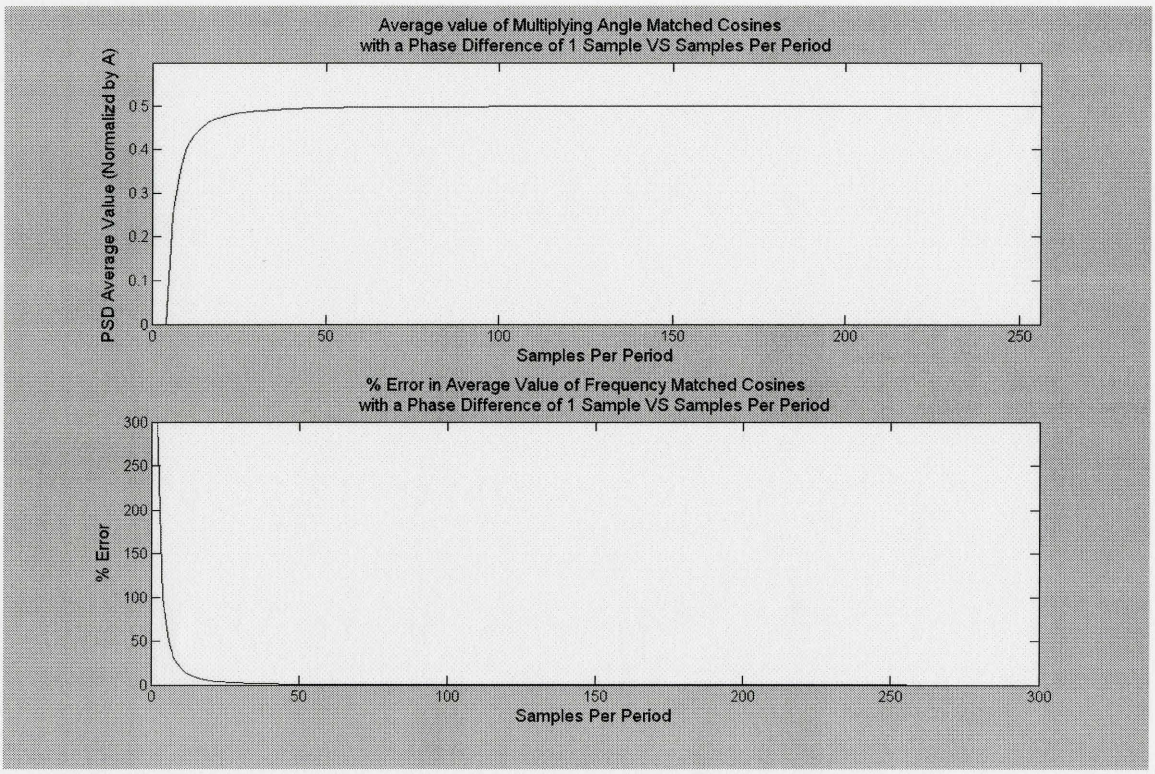


Figure 4.3 – Plot of phase sensitive detection average value and error versus samples per period. In a practical application, the uncertainty in the angle of the sampled input sinusoid will be equal to $2\pi/n_s$ where n_s is the radians per sample. Top) As the samples per period increases, the maximum phase mismatch is reduced and the PSD average asymptotically approaches the continuous time value of 0.5A. Bottom) The data is replotted in terms of % error as a function of samples per period. At n_s=32, the error is less than 1.93%. At n_s=64, the error is less than 0.49%.

asymptotically. The error for 32 and 64 samples per period is less than 1.9% and 0.49% respectively. By further increasing n_s, the error caused by phase mismatch can be made arbitrarily small. To achieve a guaranteed quantization error of less than 0.5% in PSD output, at least 64 samples per period should be implemented.

4.4 Finite-Precision Quantization Error

The analysis for discrete-time quantization above assumed infinite precision, though the data was actually calculated using the maximum precision of the software program MATLAB, which is 64-bits for floating-point numbers. No practical analog-to-digital converter provides 64-bit samples so we must investigate the error caused by finite precision. Typical ADC's for digital signal processing applications offer 8-bit to 18-bit encoding. For this analysis, we will once again assume that the input and reference cosines are precisely frequency and phase matched. Quantization error is caused by the real value being rounded (though some ADCs will truncate instead) to some binary number. The maximum error of a sample is half the continuous-value range represented by one bit for a rounding ADC or equal to the range for a truncating ADC. The resolution is a function of both the bits per sample and the maximum input signal. The resolution of a signed binary sample that is b bits wide is

$$resolution_b = \frac{1}{2^{b-1}} \times \max\_amplitude, \tag{4.3}$$

where *max_amplitude* is the full-scale signal amplitude. Thus $resolution_b$ is twice the maximum error (for rounding ADC) and equal to the maximum error (for truncating ADC) that any given binary value can have due to finite precision. The exponent is *b-1*

because the signal is signed (dual polarity) thus we only have *b-1* bits to encode the amplitude, with the most significant bit encoding the polarity. We can see the effect of the finite precision rounding ADC on Eq. (4.2b) by rewriting it as:

$$\overline{psd}_b[n] = floor\left\{\frac{1}{n_s}\sum_{n=0}^{n=n_s-1} round\left\{A\cos\left[\frac{2\pi n}{n_s}\right]\times 2^{b-1}\right\}\times round\left\{\cos\left[\frac{2\pi n}{n_s}\right]\times 2^{b-1}\right\}\right\},$$

(4.4)

where *round{}* represents the value in the curly braces is to be rounded to the nearest integer and *floor{}* represents the value in the curly braces is to be rounded in accordance with right-shifting a two's complement (see Appendix A) binary number by $log_2(n_s)$. It is important to differentiate between these two different types of rounding. The analog-to-digital converter outputs the nearest binary code to the input signal. Hardware averaging is accomplished by dividing the digital multiplier output sum by n_s via truncation. This means positive values are rounded to a lesser magnitude and negative numbers to a greater magnitude. The result is $avg(psd_b[n])$, the average binary value for phase sensitive detection with quantization of *b* signed bits.

It is important to note that the actual quantization error can be deterministic when the signal to be quantized is absolutely precise and contains no random components. This is the case with our reference sinusoid, and for the purpose of investigating finite precision we will assume it is true for our input signal. If a quantization level happens to coincide with the same value as the continuous-time signal value at a given sampling instant, then no error is introduced. Thus the actual error for any given quantization is a function of not only bits per sample, but also the samples per period. This effect is

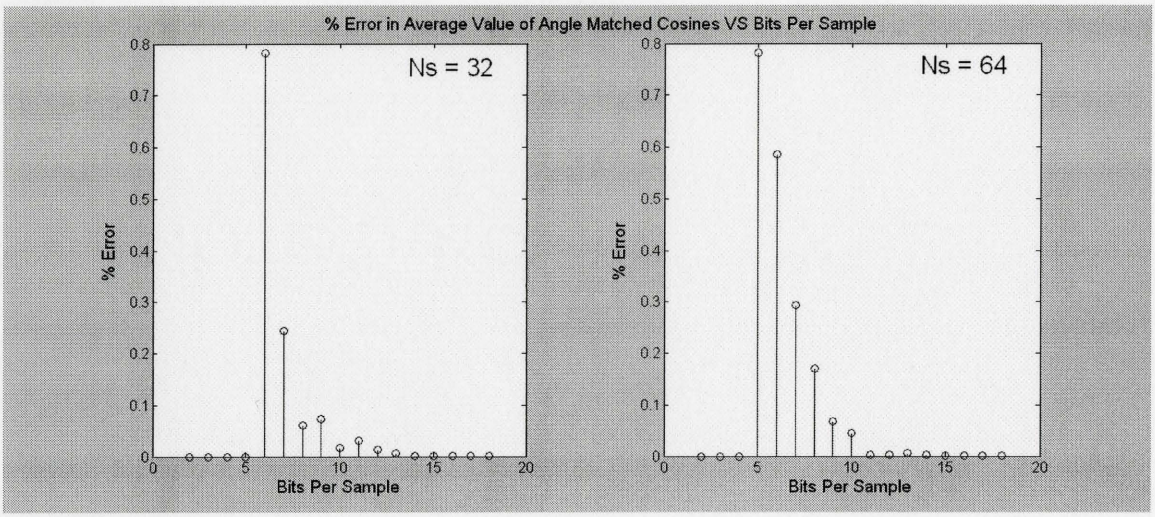


Figure 4.4 – Eq. (4.4) is evaluated for Ns = 32, 64 and the % error versus bits per sample is plotted. Note that for low bits per sample, the system resolution is so coarse that incorrect intermediate values are rounded to the correct value resulting in zero error. We also note that for lower sampling rates, the error function's envelope is oscillating, representing the presence of aliasing in our error calculations. When aliasing is present, the error calculation is not independent of phase and may be unreliable.

illustrated in Figure 4.4, where the % error in the average value is plotted versus sample width in bits for both $n_s = 32$ and $n_s = 64$. Note that in our analysis the golden value for expected average (*0.5*) can be precisely represented in any uniform binary coding scheme. This has an important consideration. In low-precision systems (i.e., less than 8-bits) it is possible for the output of the multiplier to be in error, but the output of the averaging operation appears to be error free. This occurs when the division and rounding operations result in the same code value as the error-free code value (*0.5*) due to truncation. The error function versus bits per sample should decrease monotonically, but due to the dependence on sample phase and truncation, the actual error can appear artificially reduced. When designing a system, it is usually desirable to choose a combination of

samples per period and bits per sample such that estimates of maximum error are always accurate and not aliased to appear smaller than then system actually provides.

Based on Figure 4.4, to achieve an error of less than 0.5% for 32 or 64 samples per period, it appears only 7-bits are needed. The implication here is that finite precision error has significantly less effect than the phase mismatch discussed earlier with respect to samples per period. However, by using internal delay lines discussed in Chapter 5, we can further tune the phase match to an arbitrary precision without increasing the samples per period. Moreover, there is nothing that can be done to reduce finite precision error other than increase the bits per sample. Therefore, if we want to ensure the finite precision error estimate is accurate (reduced error in the estimate means less envelope oscillation), based on Figure 4.4, at least twelve bits per sample is recommended.

4.5 Actual Error from ADCs

In the above analysis we assumed that that the ADC operation was fully deterministic without random components. In reality, the ADC is not guaranteed to provide a perfectly deterministic output. This means that a fixed DC input will not result is the same code value each time. This is an unavoidable consequence of the analog circuit's nonzero internal noise. The internal noise will manifest itself as a DC input being mapped to an output that is a random variable as described in Eq. (4.5).

$$ADC\_OUT = round\left\{\frac{v_{in}}{v_{\max}} \times 2^{b-1}\right\} + round\{X(v_{in})\}, \tag{4.5}$$

where *ADC_OUT* is the equivalent integer out of the ADC, v_{in} is the input voltage to be converted, v_{max} is the voltage magnitude corresponding to the maximum positive code value, *round{}* is the ideal rounding function to the nearest integer, b is the bits per sample, and X is a random variable with approximately zero mean, unknown variance and is a function of the input voltage. For the Analog Devices AD977A 16-bits per sample analog-to-digital converter the manufacturer has provided a histogram over 10,000 conversions of the same DC input. This is shown below in Figure 4.5 which is copied from the manufacturer's datasheet [31]. The output of this ADC can be represented as

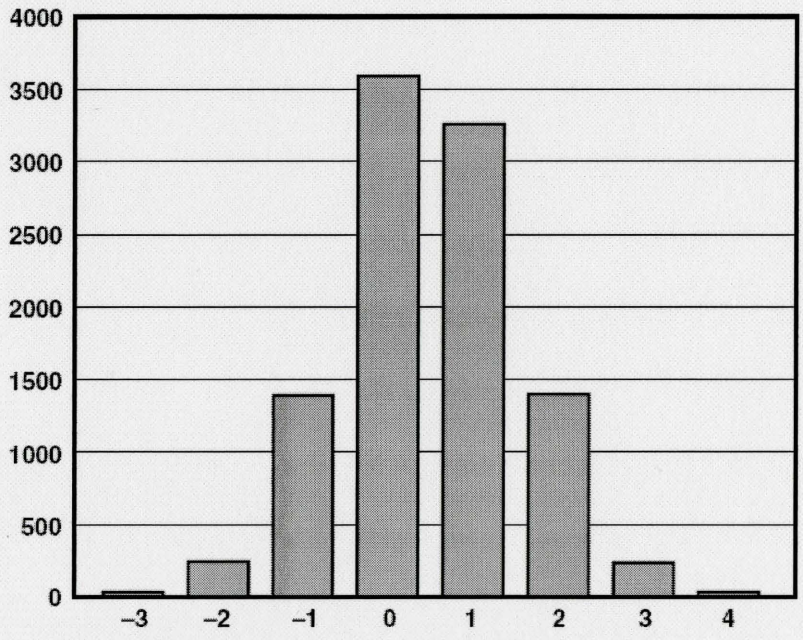


Figure 4.5 – Histogram of Analog Devices AD977A binary output for a constant DC input. The output of the converter can be modelled as the rounded binary code value plus a random noise component. The noise component is plotted above as a histogram for 10,000 conversions of a DC voltage. We can see from the histogram that the error signal has a standard deviation of approximately one bit with a Gaussian profile consistent with thermal noise. It is implied that the noise component is independent of input signal level. Copied directly from [31].

the ideal rounded value plus a quantized random noise variable with the properties shown in the figure above. Note that the input voltage is not specified so it is implied that the random variable X in Eq. (4.5) is independent of input voltage and is valid over the entire code range for this device. The histogram in Figure 4.5 indicates the random variable does not have a mean of zero, but instead is somewhere between 0 and +1.

4.6 Summary of Quantization for Phase Sensitive Detection

This chapter has identified finite sampling rate and finite precision as fundamental sources of error in digital phase sensitive detection. It has been established that special considerations arise in PSD that show quantization error can be artificially reduced due to a dependence between sample timing and precision. As a result, estimates of maximum error may be unreliable or have undesirable phase dependencies. It is recommended that 64 samples per period and at least 12 bits per sample be implemented for phase sensitive detection when quantization error is desired to be less than +/- 0.5% and accuracy in the error estimate is critical.

This chapter provided a fundamental analysis of both the independent and dependent relationship between sampling rate (samples per period) and bits per sample on phase sensitive detection. However, further analysis may be required when the actual implementation of a digital phase sensitive detector is considered. The implementation-specific analysis for the digital phase sensitive detector designed in this thesis is presented in Chapter 6 with the device characterization.

Chapter 5. Digital Lock-In Amplifier Architecture

5.1 Introduction

Digital lock-in amplifiers (DLIAs) have been discussed and reported in the literature for nearly forty years. E.D Morris *et al* first published the concept of digital phase sensitive detection in the journal *Review of Scientific Instruments* in 1968 [32]. Nearly twenty years later in 1987, L. G. Rubin submitted a letter [33] to the editor of the same journal in response to the continuous stream of publications on the topic that was ignorant of previous work. All 'new' work claimed to outperform existing analog commercial devices while offering novel techniques. Rubin's argued that researchers were continuously reinventing the wheel, based upon the apparent lack of citations in the publications. In the 20 years since Rubin's letter, more publications on DLIAs have appeared in the literature [34 - 40].

To avoid reinventing the wheel in this thesis, a focus on upgradeability and flexibility is applied to a Field Programmable Gate Array (FPGA). This thesis presents a versatile and reusable FPGA modular platform from which new instruments (not exclusively DLIAs) may be constructed providing the researcher with enhanced ability to create advanced digital scientific instruments. Without a doubt, the complete characterization and optimization of the DOP experiment presented would not have been possible without the FPGA-based DLIA platform.

This chapter presents the architecture and design of a general purpose digital lock-in amplifier. The existing analog design used previously in the DOP experiments [7,16] delivers exceptional performance, however it is difficult to repair, upgrade and future-

proof due to a dependence on knowledge of analog electrical circuit theory. By implementing the lock-in amplifier in a programmable logic device such as a FPGA, the design itself (hardware description language code) becomes easy to maintain software intellectual property. If the design is highly modular then individual components can be added, improved or replaced as needed by future researchers to expand or improve its capabilities. While some specialized knowledge is required in order to design with FPGAs, the reuse of previously designed modules (especially free design cores from public Internet sources) significantly improves the learning curve.

5.2 Architecture

The system architecture of the DLIA is composed of a number of self-contained modules that perform a specific function. Synchronous resets are used to avoid state machine flops coming out of reset on different cycles due to skew in the reset tree. A functional block diagram of the system is shown in Figure 5.1. The *UART* is a 1-byte serializer/deserializer intended to operate over the standard RS-232 protocol [41]. The UART connects to an external RS-232 Transceiver (MAX3232) which provides the electrical bridge between the 3.3V FPGA and the +/– 12V serial interface. The *host decoder* implements the DLIA command protocol shown in Figure 5.2. The *control status register* contains configuration parameters for the system. The *ADC controller* connects to an external analog-to-digital (A/D) converter from which it receives digitized samples of the input signal. The *programmable delay line* allows precision phase shifting of the input sync signal prior to the PSD operation. The *lowpass filter* provides anti-

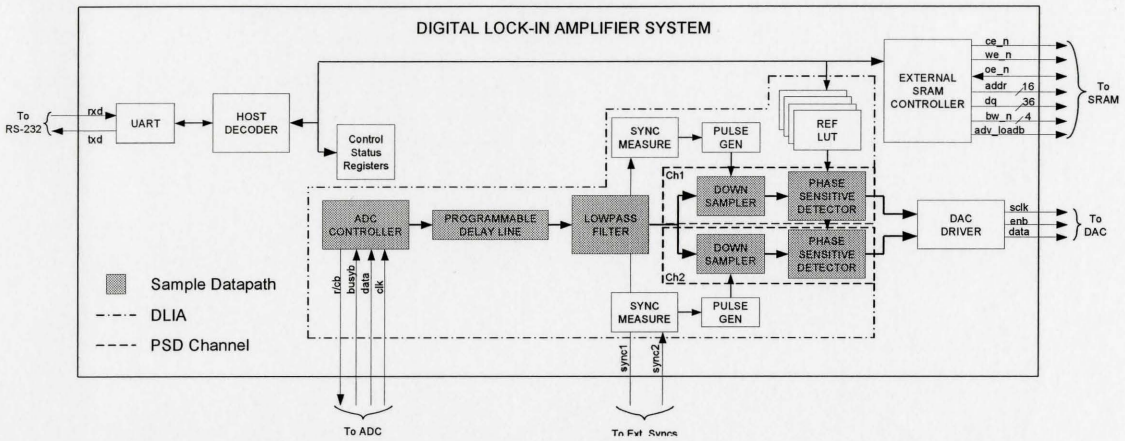


Figure 5.1 – Functional Block Diagram of the DLIA System. The shaded blocks represent the datapath of the sampled analog signal. Each block is self-contained and carries out a particular function. An individual DLIA is outlined in the diagram which contains two PSD channels. The system is created by stitching all the blocks together. The system can support an arbitrary number of DLIAs simply by instantiating more of them and connecting them into the system. For clarity, the connections between the control status registers and the various functional blocks is not shown.

aliasing prior to downsampling. The *downsampler* provides hardware-calculated downsampling of the data signal. The *sync measure* and *pulse gen* control the downsampling rate by measuring the channel's sync signal period and matching the samples per period to a user defined value. The *phase sensitive detector* performs digital PSD as described in Eq. (4.2b). The *reference LUTs* provide the digital samples of the reference signal for demodulation. The *DAC driver* can optionally drive any point on the sample data path to an external digital-to-analog converter for measurement as a real-time voltage waveform. The *external SRAM controller* interfaces to an external SRAM device and is used for large-scale temporary storage of data. For example, data samples from anywhere in the data sample path can be sequentially written to external SRAM in real-

time, then downloaded for offline analysis. All power spectral density plots in Chapter 8 were created by downloading captured datasets into MATLAB.

Note that the full system can contain one or more DLIA sub-systems, with each sub-system containing two PSD channels for demodulating a single analog input signal at two difference frequencies. Multiple analog input channels can be supported by creating multiple instances of the DLIA and stitching them into the system appropriately. For the DOP experiments, two DLIA sub-systems are used. The first demodulates the PL, DOP and ROP from a single analog input, the second demodulates the reflected HeNe from a second analog input. For simplicity, only one DLIA is shown in Figure 5.1.

In summary, the platform is hierarchical. From the top down, the system can contain multiple lock-in amplifiers, where each is connected to a dedicated A/D converter. Within a single lock-in amplifier, two phase sensitive detectors are provided. Each detector demodulates the input signal at a different frequency (e.g., 200 Hz and 1000 Hz). Each phase sensitive detector contains two demodulation channels, each channel has its own demodulation wave form (e.g., inphase and quadrature cosines).

5.3 Command Protocol and System Memory Map

The DLIA is designed to be interfaced as a memory-mapped device. This means that all configuration parameters, measurements and general purpose data storage are accessed at an address in the system map making the device appear as one large virtual memory. The memory map is composed of banks. Each bank is 65,536 addresses deep (16-bit) with a 16-bit data word width. The system is accessed via the *host decoder*

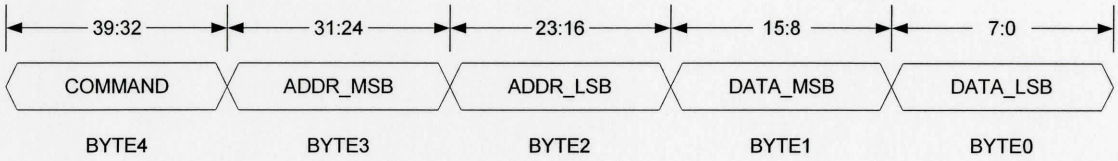


Figure 5.2 – Command protocol for the host decoder. Each packet is sent one byte at a time, starting with the DATA least significant byte, and ending with the COMMAND byte. A complete packet is five bytes (40-bits).

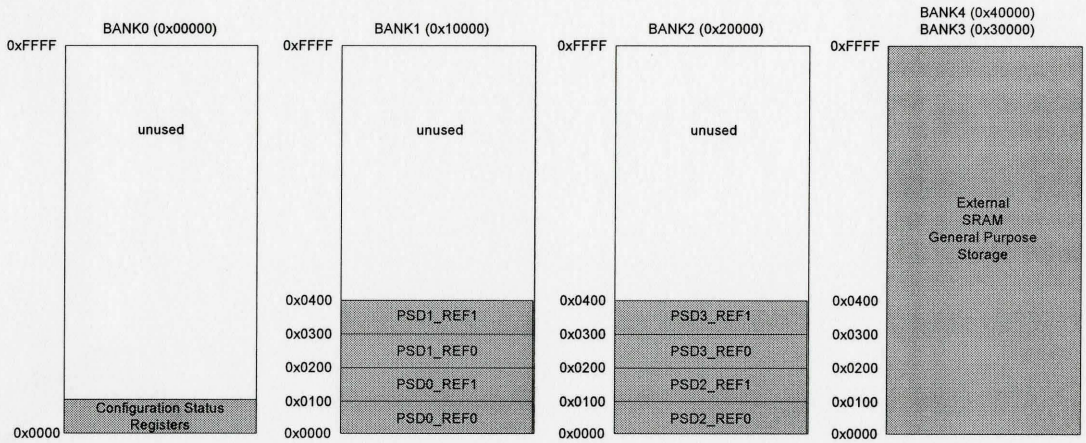


Figure 5.3 – Memory Map for the DLIA. The Host Decoder uses a 16-bit address depth for the system limiting the amount of virtual memory. To extend the memory map, the design contains multiple BANKs, each with 16-bit address depth. The memory is sparsely populated in order to demonstrate the multi-bank architecture. Unused areas represent virtual memory, thus commands to unused addresses will be ignored by the system.

protocol. A protocol packet will contain the command to be executed, as well an address and optionally a datum as shown in Figure 5.2. For example, to write information to the device, a WRITE command is sent to a particular bank, with the datum to be written at the specified address. In order to retrieve information on demand a READ command is

Table 5.1 – DLIA host interface commands

Command	Command Byte	Description
NOP	0x00	No operation. This command ensures the reset condition does not invoke an undesired action.
WRITE0	0x01	This command writes <data> to <address> in BANK0.
READ0	0x02	This commands reads <address> from BANK0.
READ0_RESP	0x03	This is a response packed from a READ0. The requested data is returned in <data>.
WRITE1	0x04	This command writes <data> to <address> in BANK1.
READ1	0x05	This commands reads <address> from BANK1.
READ1_RESP	0x06	This is a response packed from a READ1. The requested data is returned in <data>.
WRITE2	0x07	This command writes <data> to <address> in BANK2.
READ2	0x08	This commands reads <address> from BANK2.
READ2_RESP	0x09	This is a response packed from a READ2. The requested data is returned in <data>.
WRITE3	0x0A	This command writes <data> to <address> in BANK3.
READ3	0x0B	This commands reads <address> from BANK3.
READ3_RESP	0x0C	This is a response packed from a READ3. The requested data is returned in <data>.
WRITE4	0x0D	This command writes <data> to <address> in BANK4.
READ4	0x0E	This commands reads <address> from BANK4.
READ4_RESP	0x0F	This is a response packed from a READ4. The requested data is returned in <data>.
INT0	0x20	This is an unsolicited packet used to synchronize the digital system with the analog system for characterization of concurrent scans. It indicates the analog computer scanning beam has moved to the next spatial location.
INT1	0x21	This is an unsolicited packet containing the data capture on interrupt channel 1. This is the inphase and quadrature values from PSD0 (PL).
INT2	0x22	This is an unsolicited packet containing the data capture on interrupt channel 2. This is the inphase and quadrature values from PSD1 (DOP/ROP).
INT3	0x23	This is an unsolicited packet containing the data capture on interrupt channel 3. This is the inphase and quadrature value from PSD2 (HeNe).
INT4	0x24	This is an unsolicited packet used to synchronize the digital system with the analog system for characterization of concurrent scans. It indicates the analog computer controlled beam has been blocked or unblocked (toggled) for calibration of the zero level.

sent with the requested address. Note in this case the DATA field would be irrelevant in the request packet, but contain the return data in the response packet. Using this architecture, up to 256 commands can be defined. If more than 2^{16} addresses are required, each READ or WRITE command is mapped into separate banks, artificially

extending the memory map beyond 16-bits up to a maximum of 24-bits. In the DLIA implementation, there are five banks defined with WRITE, READ and RESPONSE commands for each bank. Figure 5.3 above shows the memory map implemented for the system.

Table 5.1 above shows the list of supported commands. New commands can be added by simply updating the host decoder to support the desired function.

5.4 Functional Description of System Modules

UART

This block is an RS-232 compliant serializer/deserializer. It is a common function and is not necessary to be designed from scratch. A pre-verified UART was obtained from *www.opencores.org* [42] under the public GNU license [43].

Host Decoder

This block contains a finite state machine (FSM) and command encoder/decoder (CED) with interrupt mechanism (Figure 5.4). When a byte is ready in the UART, it is retrieved by the FSM and placed in a 5-byte command register one byte at a time until the packet is complete, at which time the command is executed via the CED. The CED will decode the packet information and drive the memory interface of the system accordingly. Interrupt lines from the system can be used to send unsolicited data to the host computer. For example, the interrupt mechanism is used to send PSD data for HeNe, PL, DOP and

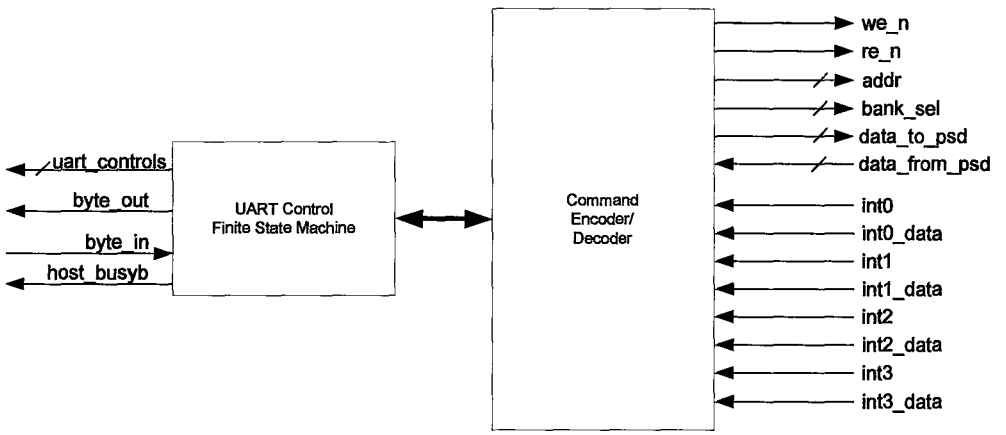


Figure 5.4 – Functional block diagram of the host decoder. A finite state machine controls the assembly of packets from the UART and the transmission of packets from the command encoder. The interface decodes packets and executes memory accesses to the system. The command encoder also assembles packets for transmission from the interrupt lines.

ROP back to the host in real-time. The CED will build a packet for transmission to the host following a read request once the data is available, or when an interrupt line asserts. The UART control FSM primarily handles two tasks: assembling a 5-byte packet from the UART, or transmitting a 5-byte packet one byte at a time to the UART. The state transition diagram is shown in Figure 5.5 and Table 5.2 provides a description of the states. Note that the host decoder cannot perform both reception and transmission simultaneously. An active-low busy signal can be used by a buffered UART for flow control.

Configuration Status Registers

This block is a collection of addressed registers. These registers contain all the configuration parameters for the system that are distributed to the various modules.

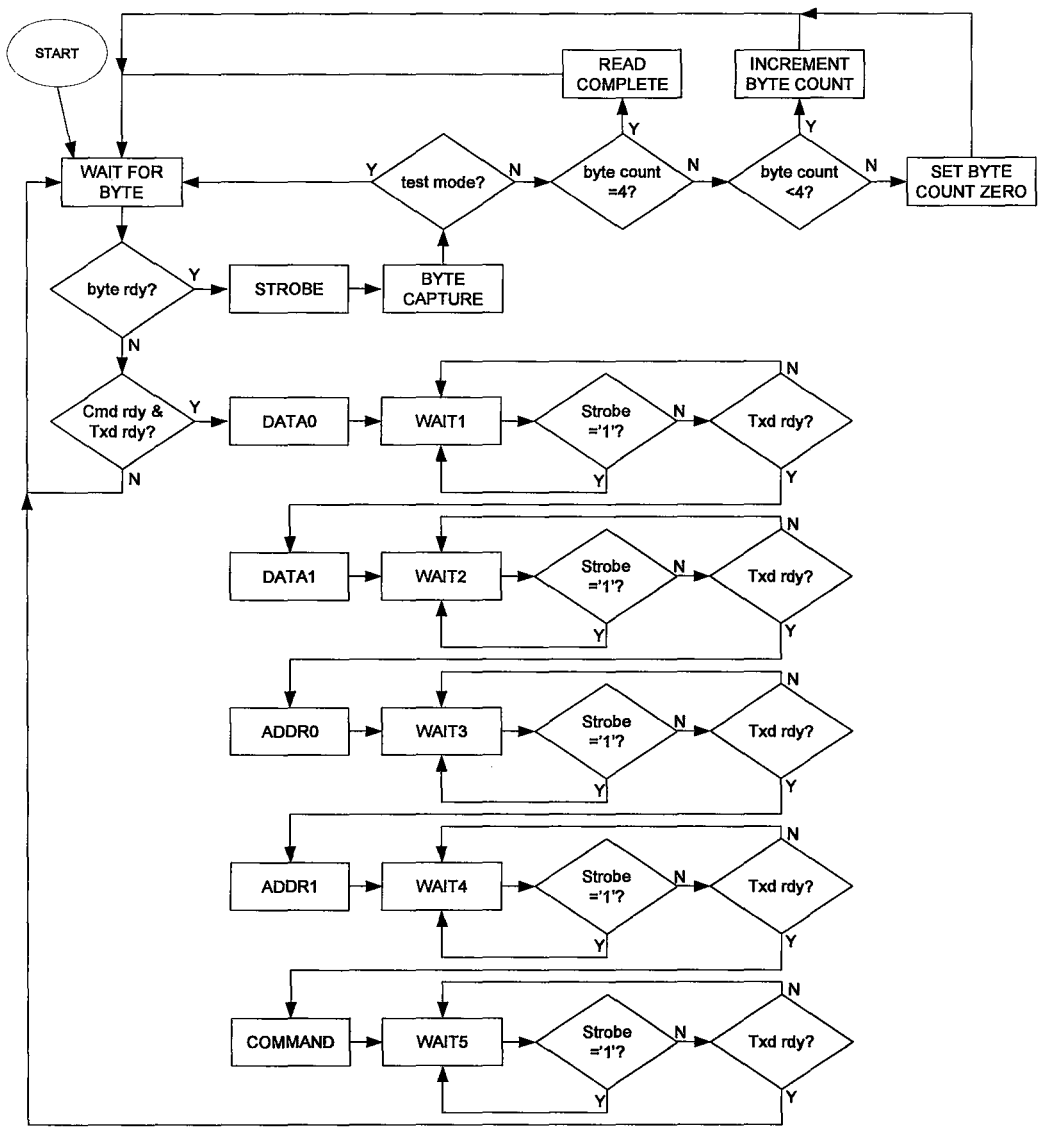


Figure 5.5 – State transition diagram for the UART control FSM. The diagram contains two branches; one handles reception of bytes to build a packet, the other handles transmission of an assembled packet one byte at a time.

Registers which are configuration parameters can be both written and read. Status registers driven by other modules in the system are read only. Externally, the CSR looks like a small memory with typical read/write controls and data in/out busses. The

Table 5.2 – UART Control FSM States

State	Description
WAIT FOR BYTE	Idle state
STROBE	A byte is ready in UART. Assert strobe to receive it.
BYTE CAPTURE	Capture the byte from the UART into the packet register.
SET BYTE COUNT ZERO	Reset the packet byte counter.
INCREMENT BYTE COUNT	Increment the packet byte counter.
READ COMPLETE	Packet is complete, send to Command Encoder/Decoder for execution.
DATA0, DATA1, ADDR0, ADDR1, COMMAND	Transmit the specified byte to the UART by activating the strobe.
WAIT1, WAIT2, WAIT3, WAIT4, WAIT5	An extra clock cycle is needed for UART to update it's transmit ready signal. This is done by de-asserting the strobe and waiting to confirm de-assertion.

implementation uses a multiplexer controlled by the address to select which register is being read. A one-hot address decoder (gated with write enable) controls the enables on the registers ensuring only the selected register is written. The architecture is shown in Figure 5.6.

ADC Controller

This block is designed to control an analog-to-digital converter (Figure 5.7). It is specifically designed to operate with an Analog Devices AD977A serial ADC. When a conversion request is made to the ADC, the sample is received serially, and shifted into a

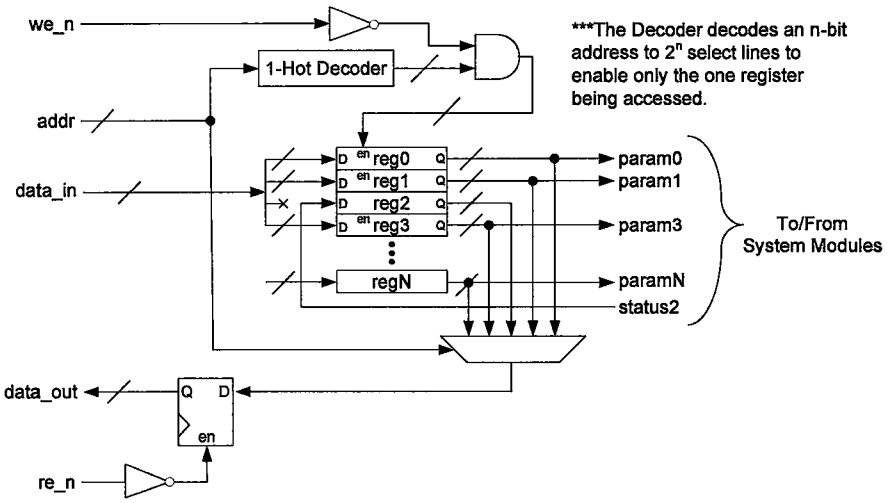


Figure 5.6 – Architecture of the control status registers. The block contains N+1 parameters which are accessed via the address bus. The outputs of the address decoder are gated with the write enable to allow data to be written to a register. The register outputs are muxed via the address to a single data_out bus when read enable is asserted. Note that status parameters are read-only.

16-bit register by the ADC's serial clock. The serial clock is discontinuous, active only while serial data is being transmitted. Once the sample is completely received it is driven onto the sample datapath indicated in Figure 5.1. The timing diagram for this is shown in Figure 5.8. Configuring the sampling rate is accomplished via a simple rollover counter. The counter counts up to a configured value at which point the counter will reset and assert a strobe to the pulse width counter. The second counter is used to create a conversion pulse that is a designer-defined width in clock cycles. This is necessary to ensure the pulse width meets the minimum assertion time requirements of the ADC. Note that the shift register is on the serial clock domain, while the parallel register is on the

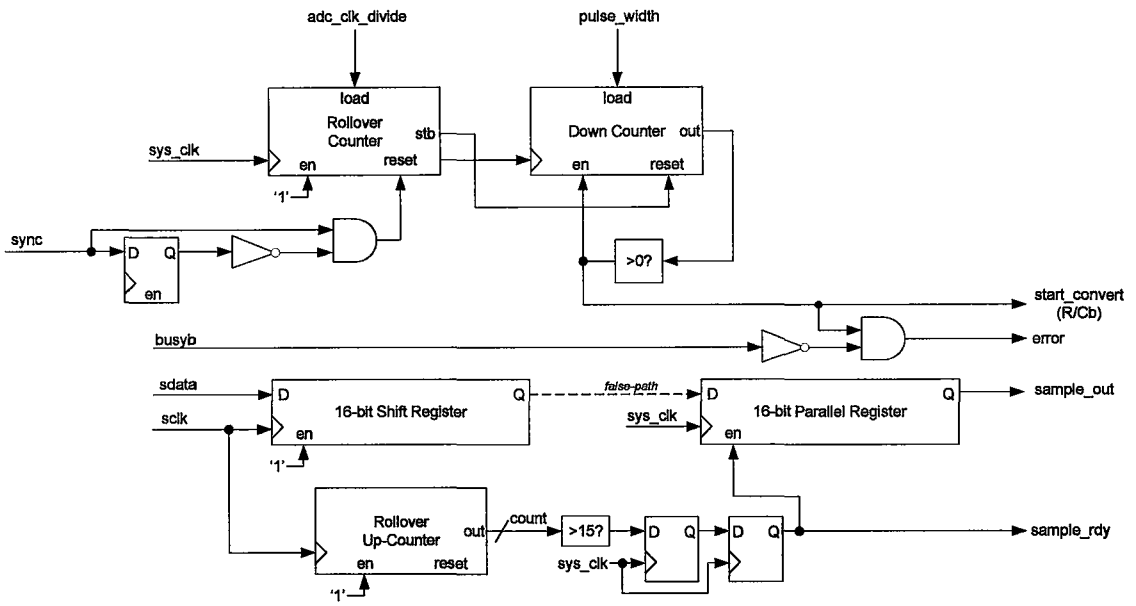


Figure 5.7 – Architecture of the ADC controller. The top half of the circuit creates a multi-cycle convert pulse to the external ADC. The sampling rate is the system clock frequency divided by adc_clk_divide. When the rollover counter strobes, a down counter is reset to 'pulse_width' and begins counting down to zero. While the count is greater than zero, start_convert is asserted. The lower half of the circuit contains a shift-register clocked by the external sclk from the ADC. Once the sample is complete it is registered and sample_rdy is asserted after crossing clock domains.

system clock domain. Since this register will be stable by the time the counter rolls over, the path can be considered a false path to the system clock domain. However, the control signal from the comparator could be metastable so two flip-flops are employed for the clock domain crossing [44]. This will help ensure a single edge is detected on the system clock domain.

Programmable Delay Line

As described in Chapter 4, the error in the system is dependent on the phase resolution in minimizing the phase mismatch between signal and reference. Since the

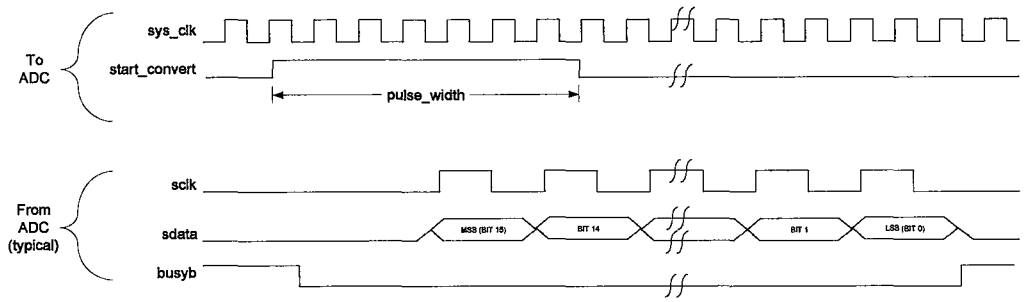


Figure 5.8 – Timing Diagram of ADC inputs/outputs for a single conversion cycle. The ADC is asynchronous to the system clock so the start_convert must be held for some minimum amount of time (device dependant). The serial data is clocked in on the discontinuous sclk.

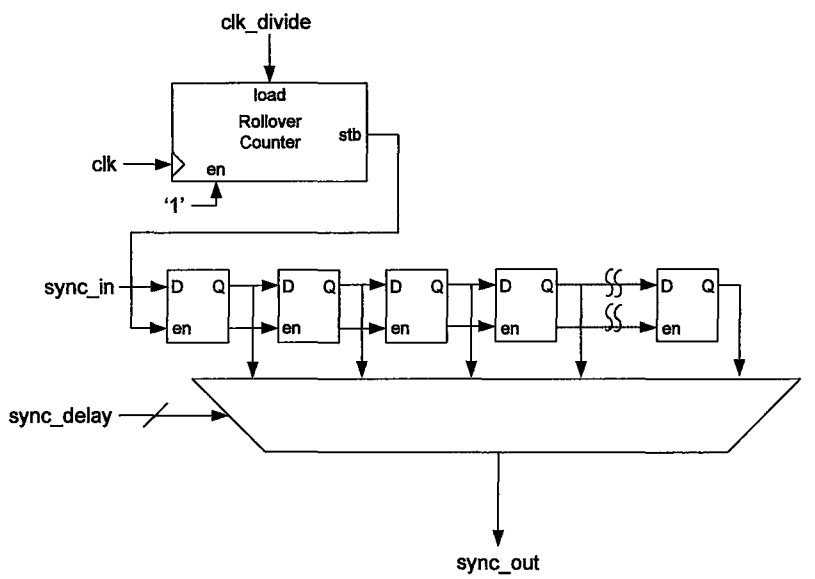


Figure 5.9 – Architecture of the programmable delay line. The circuit consists of a series of flip-flops whose output is multiplexed to the sync_out based on the desired delay. The rollover counter controls the temporal resolution of the phase shifting.

reference wave in the PSD function is usually synchronized to an external sync signal we need a way to align the data samples phase with the sync. Rather than use $16n$ flip-flops to delay a sample by n-cycles, we can delay the single-bit sync signal instead. This

allows us to achieve extremely high precision in phase shifting without resorting to a large number of samples per period (which will increase the hardware resources required in the processing modules). The resolution of the sync-pipe is set by using a pulse generator to activate the shift operation. The pulse generator period can be any multiple of the system clock period. Since the sync signal can be delayed in time but not advanced, we can obtain the equivalent result by changing the PSD phase offset instead (decreasing the reference wave index) by one sample and then delaying the sync signal in the delay line.

Lowpass filter

The lowpass filter is a standard symmetric FIR filter with a configurable number of taps (at compile/synthesis time). The frequency response of this filter is desired to contain the demodulation reference frequency for both PSDs in the passband, but has a stopband frequency below half the downsampled frequency (to prevent aliasing) of Channel 1 or Channel 2, whichever is lower. The architecture for an *N+1* tap filter (Figure 5.10) requires *N+1* sequential sample registers, *N/2-1* two-input adders, and *N/2* multipliers. One large accumulator with *N/2* inputs is also required. The division is accomplished by a binary right-shift, which requires the total weight of the filter coefficients to be a power of 2. The weight of the filter and output is defined as,

$$filter\_weight = \sum_{j=0}^{j=N} c_j , \tag{5.1}$$

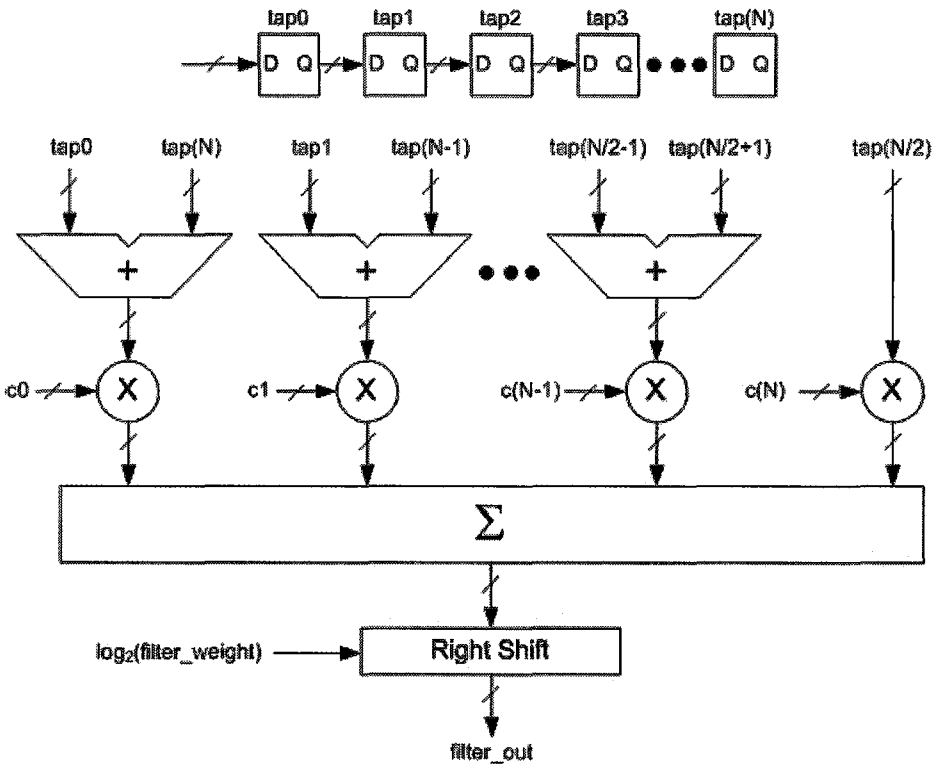


Figure 5.10 – Architecture of a symmetric N+1 tap FIR filter. The filter consists of N+1 sequential sample registers that form the tap points. Each two tap points that are symmetric about the centre tap are added together and multiplied by the corresponding coefficient. The multipliers are summed together and divided by the filter weight where the weight is a power of 2. Note that this architecture requires an even number for N (odd total number of taps).

$$\begin{aligned} filter\_out[n] &= \frac{1}{filter\_weight}\sum_{j=0}^{j=N} c_j tap_j \\ &= \frac{1}{filter\_weight}\sum_{j=0}^{j=N} c_j filter\_in[n+N-j], \end{aligned} \tag{5.2}$$

where c_j is the j^{th} filter coefficient, tap_j is the $(n+N-j)^{th}$ sample from the filter input sequence, $N+1$ is the number of taps, and *filter_out[n]* is the n^{th} sample out of the filter. The input sequence has been zero padded by N samples assuming the taps are reset.

Sync Measure and Pulse Gen

This block calculates the system clock divide value that will give a downsampled frequency that yields the desired samples per period with respect to the sync signal from Figure 5.9. In other words, by continuously measuring the period of the external sync signal, a downsampling trigger is generated that will result in precisely the desired number of samples per sync period. This allows dynamic frequency tracking of the input signal.

The block contains an up-counter that counts system clock cycles that on the low-to-high transition of the sync signal, the count value (representing the sync period in system clocks) is captured by flip-flops and the counter is reset. The sync period is then divided by the desired samples per period via an n-bit right shift (divide by 2^n) and loaded

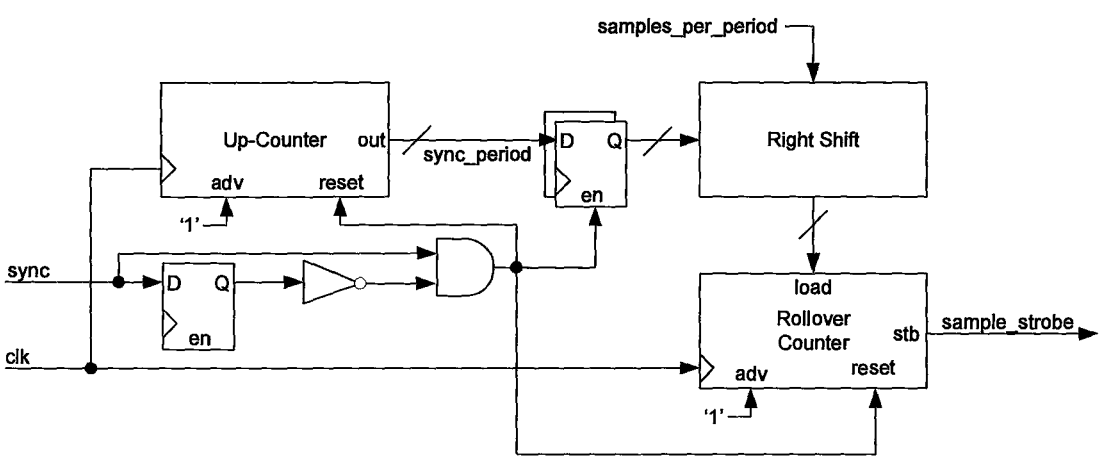


Figure 5.11 – Architecture of the sync measure and pulse generator functional blocks. The up-counter measures the period of the sync signal in system clock cycles. For each low-to-high transition of the sync signal, the count value is captured by flip-flops, divided by the samples per period and fed to a down-counter that generates the sample strobe to the downsampling block.

to a down-counter in accordance with Eq. (5.3b) below. This counter will produce a strobe signal which indicates the desired downsampling time-instants. The sync period is calculated once per sync period, thus this circuit enables dynamic tracking of frequency changes in the signal of interest with a maximum latency of one sync period (measurement of wave period N gets applied to wave period $N+1$).

$$Downsampled\_rate = \frac{sys\_clk}{clk\_divide\_value} \tag{5.3a}$$

$$clk\_divide\_value = \frac{sync\_period}{samples\_per\_period} \tag{5.3b}$$

In Eq. (5.3b), the required clock divide value is the sync period (measured in system clock cycles) divided by the samples per period. Figure 5.11 shows the architecture of the sync measure and pulse gen.

Downsampler

This block is very simple, consisting only of a single parallel register with the data input connected to the output of the lowpass filter stage, and the enable connected to the 'sample_strobe' from the sync measure block. When the sync measure and pulse gen circuit indicates a desired sampling instant, the block captures the current sample output from the lowpass filter and passes it as a valid sample down the sample datapath. Since the lowpass filter structure naturally provides a weighted averaging scheme, no further averaging between samples is necessary for downsampling as long as the new reduced sampling rate exceeds the Nyquist rate with respect to the spectral signal content.

Reference Look-Up Tables

The LUTs contain the samples of the reference demodulation waves. The table width is the same as the sample width in bits, and the table depth is the number of samples per period. For each PSD channel, the two tables (one for each demodulation wave) are implemented as embedded SRAM made available in most modern FPGAs. Implementing the tables as RAM rather than ROM allows in-system programming of the demodulation waves by the host computer. Refer to the manufacturer's FPGA guide on how to utilize the device's embedded SRAM resources [45].

Phase Sensitive Detector

The PSD performs the discrete phase sensitive detection described in Chapter 4 by

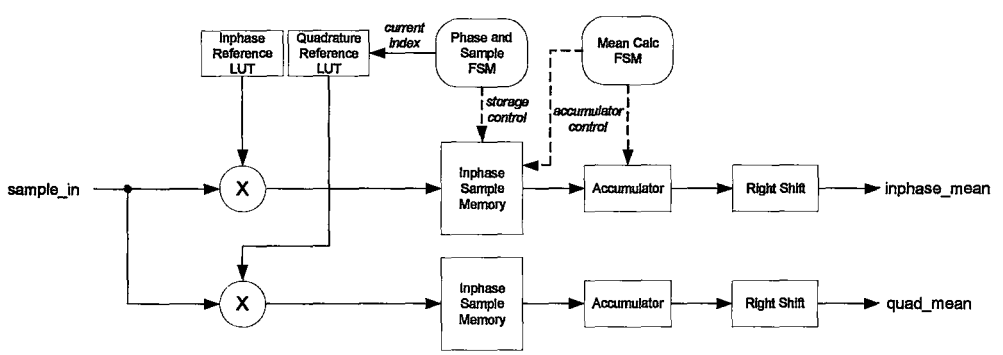


Figure 5.12 – Functional block diagram of the phase sensitive detector. The PSD is dual channel for demodulating both the inphase and quadrature components. LUTs provide the reference samples and the multiplier outputs are stored in memory. The number of stored samples is designer-configurable to a power of 2. Once the storage is filled, the contents are summed in an accumulator and the result is divided by the number of samples. The result is the DC component of the multiplier output. The phase and sample FSM controls the retrieval of reference LUT samples and storage of multiplier outputs. The mean calc FSM controls the accumulation sum.

Eq. (4.1) and its block diagram (Figure 5.12) is similar to the digital portion of Figure 3.3. The module accesses two lookup tables which store the digital samples of the demodulation waves as described above. Typically, the LUTs will be programmed with the same wave, phase shifted by $\pi/2$ radians to yield inphase and quadrature demodulation. However, both PSD multipliers are independent so there is no requirement that the reference waves be related. The module is capable of operating in two distinct modes. In *external sync mode*, the demodulation function is synchronized to an externally provided digital signal which contains precisely the same period as the signal of interest. If no external sync is available, *auto sync mode* is provided to search for and lock-on to the correct phase relationship using digital feedback locking techniques.

A phase and sample control FSM handles the task of retrieving the correct reference samples and providing them to the multipliers, furthermore it sequentially stores the multiplier output in the sample memories (implemented as embedded SRAMs). A second FSM for performing the mean calculation is used to retrieve the multiplier samples from the memory and load them into an accumulator. The sum is then right shifted to accomplish a divide by 2^n where 2^n is the total number of samples to be averaged and n is an integer. This result is the final value from the PSD. Typically the memory will store many signal periods worth of data in order to reduce the rate at which output samples are produced (lower bandwidth requirements on host/PC connection). It is worth noting that the FIR lowpass filter in Figure 3.3 has been replaced with an accumulator and shifter. This allows maximum ease in calculating the mean. An accumulator/shifter is a special case of a FIR filter where the number of samples

accumulated is the number of taps, and all coefficients are the same value, equal to the reciprocal of the number of taps. Alternatively, using an explicit architecture similar to the FIR filter in Figure 5.10 for extracting the mean would be far less efficient then a simple accumulator and shifter.

A significant portion of the complexity of this module is contained in the state machines. The detailed state-transition diagrams for the phase/sample and mean calc FSMs is shown below in Figures 5.13 and 5.14 and their state description tables in Tables 5.3 and 5.4, respectively.

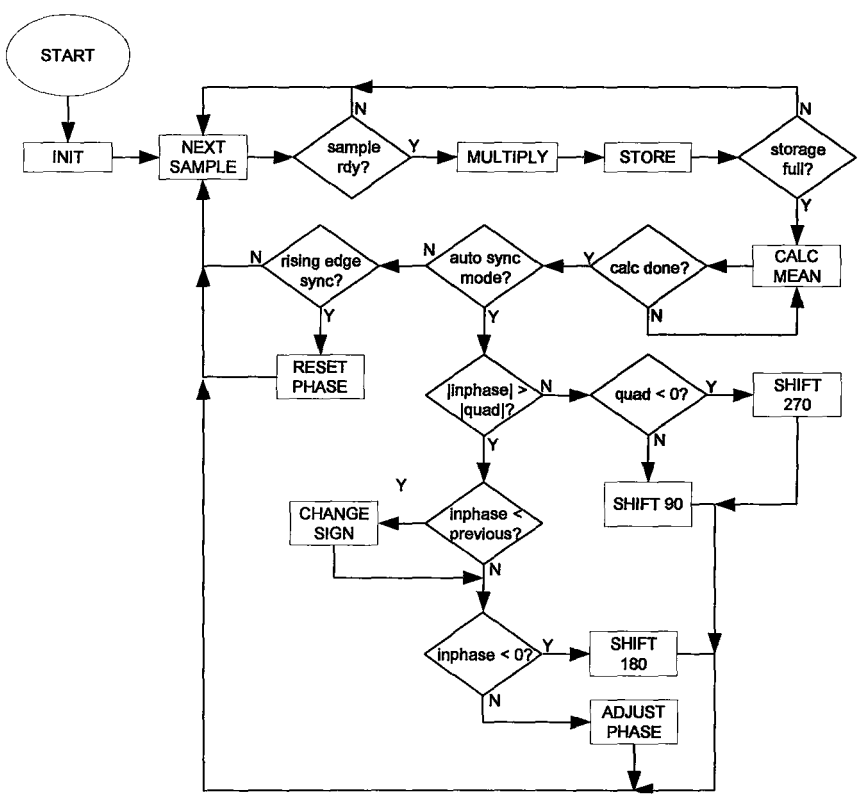


Figure 5.13 – State transition diagram for the phase and sample FSM. The circuit controls the operation of the multiplier and storage memories as well as the reference LUT index depending on whether external or automatic synchronization is configured.

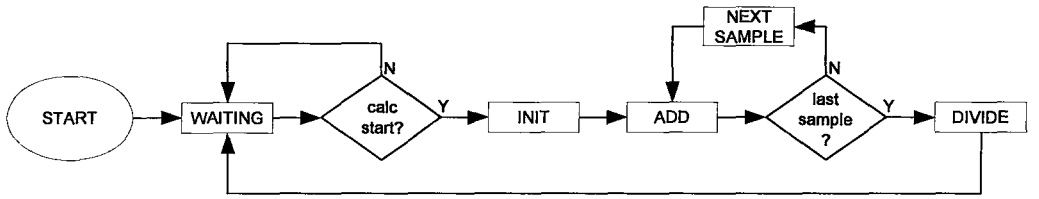


Figure 5.14 – State transition diagram for the mean calculation FSM. The circuit controls the retrieval and accumulation of the sample summation as well as triggering the divide (right shift) once completed.

Table 5.3 – PSD Phase and Sample FSM States

State	Description
INIT	Reset and initialize all state variables.
NEXT SAMPLE	Wait for the next sample to arrive at the PSD inputs.
MULTIPLY	Multiply the input sample by the reference sample.
STORE	Store the multiplier result into the storage memory.
CALC MEAN	Trigger the Mean Calc FSM to begin.
RESET PHASE	Reset the reference phase to the user defined phase value. This forces phase realignment for every new sync period.
CHANGE SIGN	The absolute value of the new inphase mean is smaller than previous mean. The reference is diverging from phase-match, so the direction of phase shifting is reversed. (Auto sync mode only).
SHIFT 90	The quadrature component is larger in magnitude than the inphase component and the quad component is positive. The reference is roughly 90 degrees out of phase. (Auto sync mode only).
SHIFT 270	The quadrature component is larger in magnitude than the inphase component and the quad component is negative. The reference is roughly 270 degrees out of phase. (Auto sync mode only.)
SHIFT 180	The absolute value of the new inphase mean is larger than the previous, and is larger than the quadrature component, but is negative. The reference is roughly 180 degrees out of phase. (Auto sync mode only).
ADJUST PHASE	The reference phase index is adjusted by either +1 or -1 depending on current polarity set by the CHANGE SIGN state. (Auto sync mode only).

Table 5.4 – Calc Mean FSM States

State	Description
WAITING	Idle state. Waiting for a signal from the Phase and Sample FSM to indicate to begin calculation.
INIT	Initialize all variables and accumulation sum to zero.
ADD	Add the current sample from the storage memory to the accumulation sum.
NEXT SAMPLE	Retrieve the next sample in the storage memory.
DIVIDE	Accumulation is complete. Right shift the sum by the required number of bits.

Chapter 6. DLIA Optimization and Characterization

6.1 Overview

To obtain the most useful results from the digital lock-in amplifier it is necessary to understand how the various digital parameters affect the results, and how these results may differ from the analog lock-in amplifier used in the past for the DOP experiments [7, 16]. A fundamental analysis of the digital quantization error introduced through finite bit-precision and finite sampling rate was presented in Chapter 4. However these parameters influence the system in other ways when the implementation architecture (presented in Chapter 5) is taken into account. When treating the entire digital lock-in amplifier as a linear time-invariant system, there are five parameters that will determine the system response; bits per sample, sampling rate, filtering, samples per period, and demodulation wave form. These parameters influence each other and manifest themselves in the calibration constant, input and output stage filtering, and overall accuracy and noise suppression. Based on the results of Chapter 4, the effect of changing bits per sample is negligible when greater than eight bits are used; however we will see this still impacts the design of the anti-aliasing filter. For this chapter, all data are obtained using sixteen bits per sample.

6.2 Input Stage Characterization

The input stage of a lock-in amplifier conceptually begins at the signal source and ends at the input to the PSD multiplier. Input stage filtering is regularly used in many signal processing systems and lock-in amplifiers are no exception. Figure 5.1 has been

redrawn with the preamplifier circuit shown in Figure 6.1 below. The input stage now includes the analog and digital signal path up to and including the downsampler.

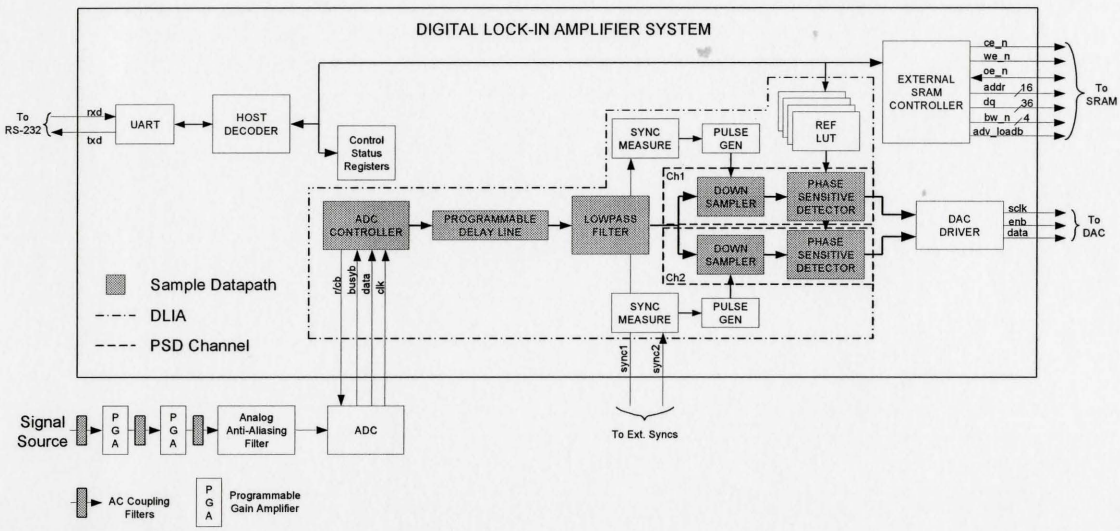


Figure 6.1 – DLIA functional block diagram including input stage. Figure 5.1 has been redrawn to include the AC coupling stages, preamplifiers and ADC. The input stage starts at the signal source and ends at the output of the downsampler.

For optical signal applications, the output from the photodetector will usually require significant amplification, often divided into several sub-stages. Since each amplifier will have some non-zero output offset voltage, it is recommended to AC couple each stage to remove any unwanted DC offsets. Failure to do so may cause the DC component to be amplified by each stage eventually beyond the last amplifier's power supply rail and saturate, thus destroying the information in the signal. Removal of the unwanted DC component is accomplished by using one or more inter-stage highpass filters with a cutoff frequency above zero Hertz. It is important to note that different

design choices were made for the preamplifier and AC coupling circuit for the new digital system with respect to the existing analog lock-in design. As a result, each system will not have the same high-pass response, estimates of which are plotted below in Figure 6.2.

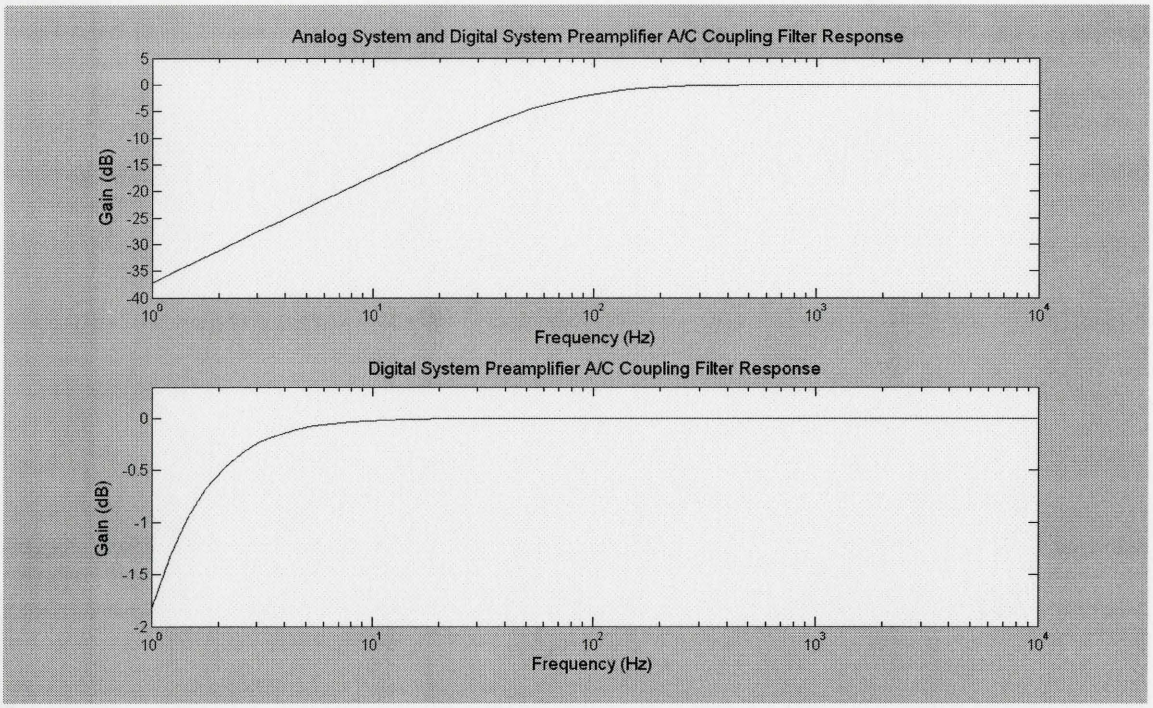


Figure 6.2 - Filter response for 1st order AC coupling highpass filter. Top) Analog system. R = 10 kΩ, C = 220 nF. The cutoff frequency of 72.3 Hz was intended to help reduce the power line interference at 60 Hz. At 200 Hz, the filter's response is -0.53 dB, or 0.94. Bottom) Digital system. R = 1 MΩ, C = 220 nF. The cutoff frequency is 0.72 Hz. The author feels the DLIA will sufficiently reject power line harmonics and mild AC coupling is sufficient. Note the response accuracy in the 0 ~ 1 Hz range is suboptimal.

The digital system additionally requires that an analog anti-aliasing filter be used before sampling in order to satisfy the Nyquist criteria discussed previously in Chapter 4. Ideally we want a filter that has unity gain at the DOP experiment frequencies of interest

(rotating polarizer and optical chopper) but has sufficient attenuation such that frequencies above half the sampling rate are completely eliminated.

To achieve total effective elimination of aliasing (meaning any spectral component above the half the sampling rate is attenuated to less than one bit amplitude), the filter must satisfy Eq. (6.1) below for a two's complement system and a full-scale input.

$$\left|Filter\left(\frac{sampling\_rate}{2}\right)\right| < \frac{1}{2^{bits\_per\_sample-1}}, \tag{6.1}$$

where *Filter()* is the filter magnitude response to be evaluated at half the sampling rate. Eq. (6.1) follows directly from the definition above. Expressed in dB, Eq. (6.1) can be rewritten as,

$$Filter\_dB\left(\frac{sampling\_rate}{2}\right) < 10\log_{10}\frac{1}{2^{bits\_per\_sample-1}} \tag{6.2}$$

Eq. (6.2) evaluated for 16-bits per sample yields an attenuation requirement of approximately -45 dB at half the ADC sampling rate. Since a second-order (two-pole) filter achieves a frequency roll-off of 40 dB/decade [46], we want a decade between the chopper and the cutoff (-3 dB point) to ensure unity gain, then more than another decade to achieve the -45 dB. Considering the chopper typically operates around 1 kHz, this implies the cut-off rate should be about 10 kHz, the half sampling greater than 100 kHz and finally the sampling rate greater than 200 kHz. Not only is this impractical, but we shall see that the sampling rate will directly affect the requirements of the downsampling lowpass FIR filter shown in Figure 6.1.

When designing a practical anti-aliasing filter, we must often choose a trade-off between less than unity gain (and non-equal response) at our two frequencies of interest (polarizer and chopper) versus the possibility of non-zero aliasing effects. However, if the input signal is unlikely to have strong components above half the sampling rate (i.e., the detector is sufficiently band-limited), then it is reasonable to consider aliasing effects to be negligible and a less aggressive filter will be suitable. The response of the Sallen-Key [47] filter shown in Figure 6.3 achieves a response of -35.2 dB at 50 kHz which provides approximately 99.97% aliasing attenuation for a 100 kHz sampling rate.

In order for the digital phase sensitive detection to work, the samples per demodulation wave period and the samples per period of the signal to be extracted must be precisely matched otherwise phase errors described in Chapter 4 will result. In order

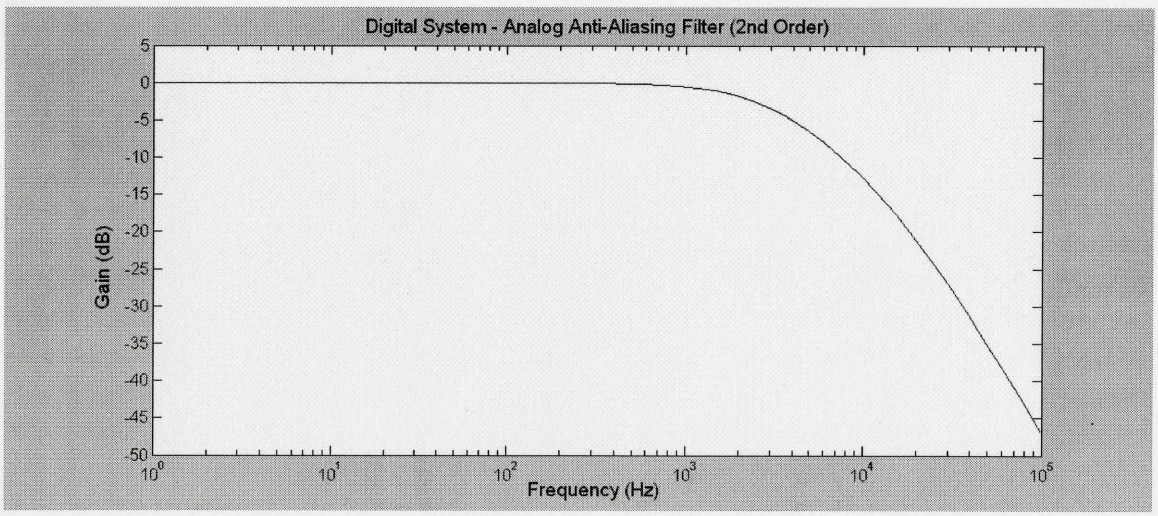


Figure 6.3 – Magnitude response of the analog anti-aliasing filter for the digital lock-in amplifier. The 2nd order filter is constructed using the Sallen-Key lowpass configuration with R1=R2=16.6 kΩ, C2=2C1=2.0 nF. The response at 1000 Hz is -0.54 dB, or 0.94. The response at 50 kHz is -35.2 dB, or 0.017.

to decouple this synchronization requirement from the sampling rate of the external ADC, the digital signal is downsampled to the correct lower rate (samples per period × reference sync frequency) by the sync measure and downsampler blocks shown in Figure 6.1. Define this automated lower sampling rate to the *reduced* rate. The downsampling FIR digital lowpass filter should be configured such that once again, the Nyquist criterion is satisfied for the new reduced rate while maintaining unity gain at the frequencies of interest. The frequency response roll-off of FIR filters is a function of the coefficients and the number of taps in the filter. The actual frequencies that correspond to the filter's response is a function of the temporal sampling period, thus the sampling rate has an effect on the actual frequency response and filter design. A higher input sampling rate will require more taps in order to meet the frequency-specific design requirements. Conversely, a lower input sampling rate can create the same absolute cutoff frequency with fewer taps. The magnitude response of two example 64^{th} order (65-tap) FIR filters are shown below in Figure 6.4. FIR filters do not have the smooth monotonically decreasing stopband that is typical of analog filters; they have a point of diminishing return. Once a certain frequency is reached, the response begins oscillating with a slightly decreasing or generally flat envelope. In Fourier Theory, this is due to the Gibbs phenomenon [48] and is a function of the passband width. This is visible in the figure below where the filter on the top is designed for lowest cutoff while the filter on the bottom is designed for unity gain in the passband. The former does not have a flat passband which will result in unequal attenuation of our frequencies of interest. However, the latter may not provide suitable attenuation for anti-aliasing, depending on

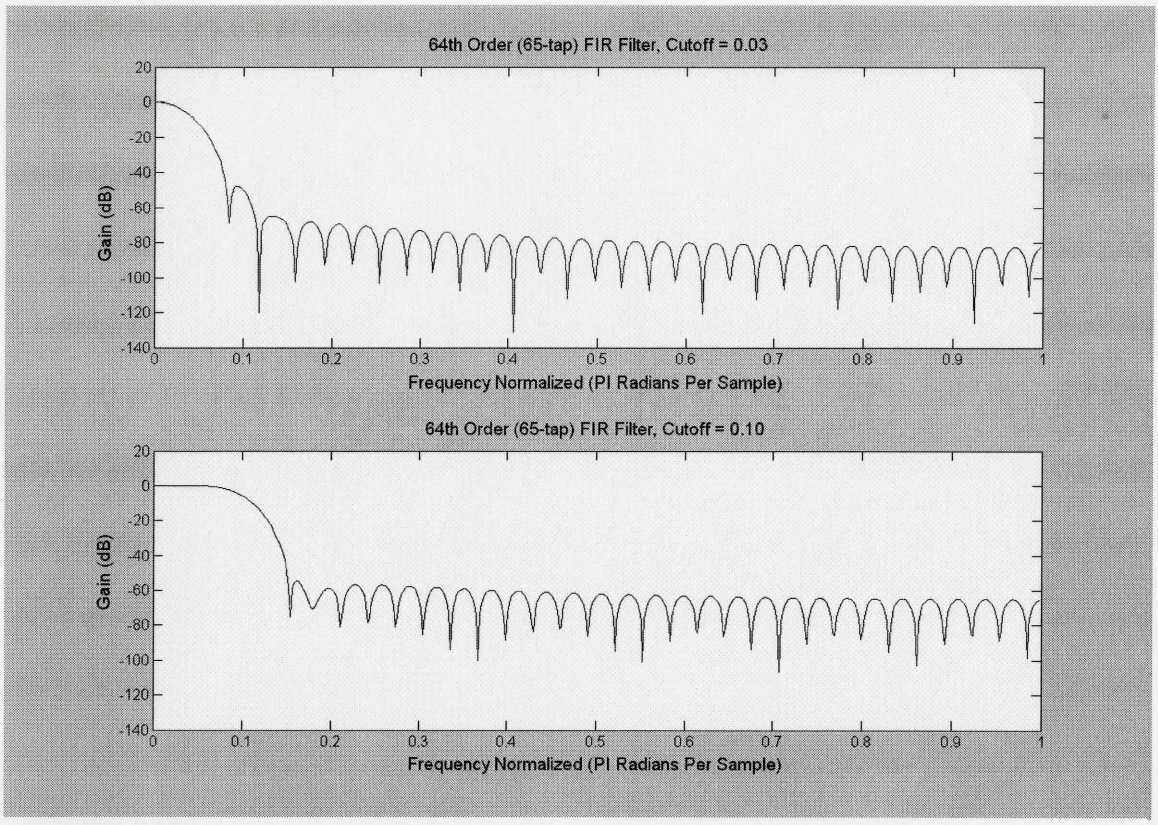


Figure 6.4 – Magnitude response for a 64th order (65-tap) lowpass filter. The frequency axis is normalized such that 1.0 is half the sampling rate. Top) Cutoff set to 5% of the half-sampling rate. The filter reaches an average attenuation of -80 dB in the stopband. Bottom) Cutoff set to 10% of the half-sampling rate. Average attenuation of -64 dB in the stopband.

the input sampling rate and reduced sampling rate. In order to determine the necessary cutoff, we must calculate the reduced half-sampling frequency created by the hardware following the downsampling block.

$$reduced\_half\_rate = \frac{sync\_freq \times samples\_per\_period}{2}, \qquad \text{Eq. (6.3)}$$

where *reduced_half_rate* is the half the sampling rate at which phase sensitive detection is to be performed, *sync_freq* is the external frequency of the reference sync signal and

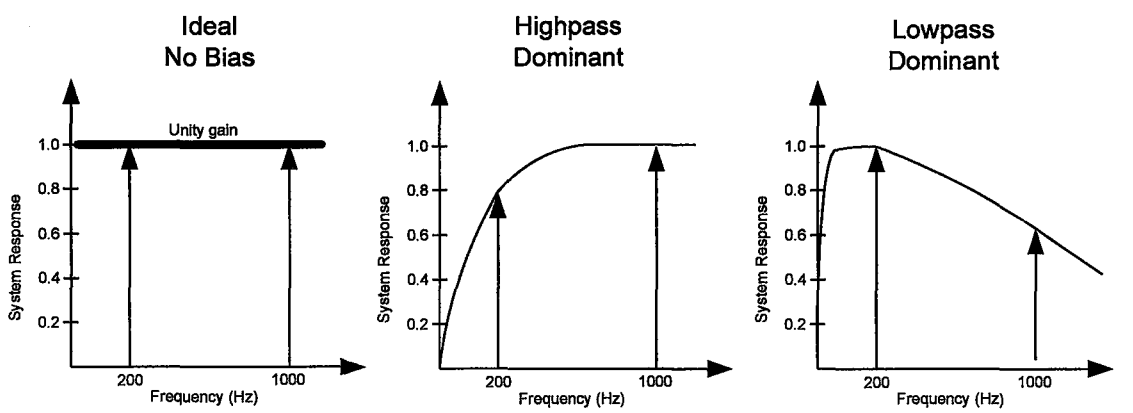
samples_per_period is the number of samples in one period of our reference demodulation wave.

For example, a 200 Hz polarizer at 32 samples per period has a reduced sampling rate of 6.4 kHz and a reduced half-rate of 3.2 kHz, but we also want unity gain at the chopper frequency of 1000 Hz. This is an almost impossible requirement to meet and trade-offs are inevitable. Similar to the analog anti-aliasing filter preceding the ADC, if we do not anticipate significant power in the higher frequencies, then aggressive anti-aliasing will not be necessary and a filter similar to the bottom of Figure 6.4 can be used. However, if aliasing is present we must use a filter similar to the top of Figure 6.4, and accept non-equal response at our frequencies of interest.

In summary, our choice of sampling rate imposes competing requirements on the analog anti-aliasing and FIR downsampling filter requirements for the digital lock-in amplifier. Increasing the sampling rate relaxes the requirements of the analog filter while decreasing the sampling rate relaxes the requirements of the digital filter. Since it is far easier to create a high-order digital filter than an analog counterpart, it is advisable to choose a higher sampling rate to relax the analog anti-aliasing requirements. The design of these two filters sets the overall lowpass response for the input stage of our linear system while the AC coupling in the pre-amplifier stages sets the highpass response for the input stage. As a result, the overall input stage response for the polarizer at 200 Hz and the chopper at 1000 Hz may not be of equal gain. The calibration constant described experimentally in Chapter 2 is a direct consequence of this non-uniform response. This concept is explained in the following section.

6.3 Calibration Constant

The calibration constant is determined by the overall bandpass response of the input stage by combining the highpass and lowpass characteristics. Unfortunately, the calibration constant cannot be accurately calculated simply by examining the AC coupling and anti-aliasing filters due to the non-infinite input impedance and non-zero output impedance of the preamplifiers. Despite this, the input stage can still be generalized as highpass or lowpass dominant, and filter design changes will be evident in the measured calibration constant of the system. This concept is illustrated in Figure 6.5. The value of the calibration constant is determined such that it corrects the DOP reading to become valid for any reasonable combination of highpass and lowpass system



$$calib = \frac{1.0}{1.0} = 1.0$$

$$calib = \frac{0.8}{1.0} = 0.8$$

$$calib = \frac{1.0}{0.66} = 1.5$$

Figure 6.5 – Impact of the input stage response on the calibration constant. The calibration is a consequence of the input stage not having a perfectly flat magnitude response across all frequencies of interest. An ideal (unbiased) system will have a calibration constant of unity. A highpass dominant system will have a constant less than unity, while a lowpass dominant system will have a constant greater than unity.

response. This is possible because the entire system is linear and time invariant. The DOP for a particular spatial location is calculated as,

$$DOP_y = \frac{psd\_polarizer}{calib\_const} \times \frac{1}{psd\_chopper}, \qquad \text{Eq. (6.4)}$$

where DOP_y is the fractional degree of polarization, *calib_const* is the calibration constant, *psd_polarizer* and *psd_chopper* are the DC phase sensitive detector outputs at the polarizer and chopper frequencies, respectively. As described in Chapter 2, before performing a DOP scan the calibration constant is first determined by placing a static polarizer in the optical path with the polarization axis aligned to the 100% direction in accordance with Figure 2.2. In this scenario, a uniform gain input stage would theoretically yield equal measurements from both the chopper PSD and the polarizer PSD yielding a DOP_y of 1.0 or 100%. Depending on the true response of the input stage, the calibration constant will fall into one of the scenarios depicted in Figure 6.5 above.

An easy way of observing the effect of input stage response on calibration constant is to modify the coefficient values for the FIR downsampling filter. The DOP maps for four different scenarios are shown below in Figure 6.6. Note that '*digital/analog preamp*' refers to the different circuit designs for the two systems; the digital system still has analog AC coupling and preamplifiers. The first two are scans done with FIR filters similar to those in Figure 6.4 above with a 100 kHz input sampling rate. The calibration constant for these two scenarios was measured to be 1.95 and 1.54 for cutoffs of 1.5 kHz and 5.0 kHz respectively. Note that some chopper attenuation is unavoidable due to the analog anti-aliasing preceding the ADC. However, with a low cutoff frequency in the FIR filter, the chopper frequency is heavily attenuated and this is

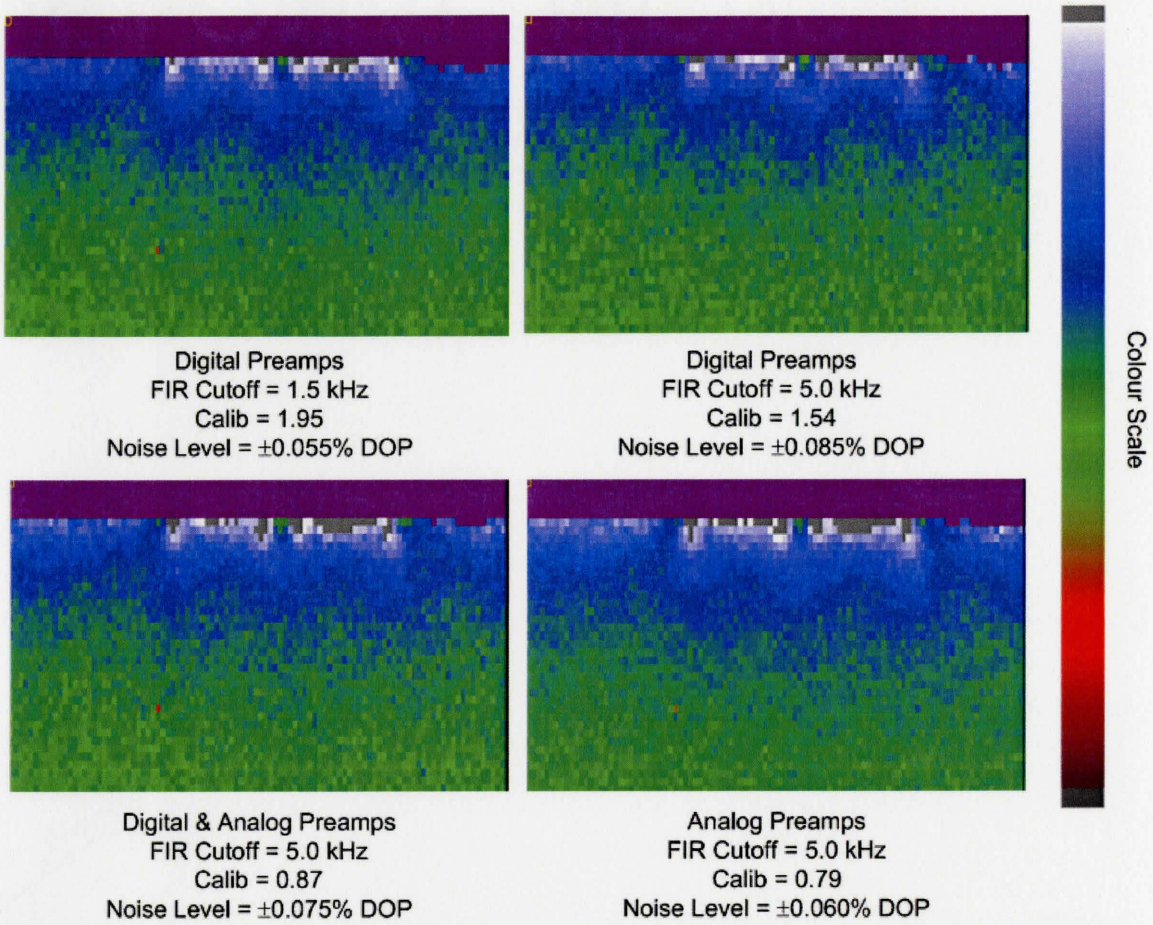


Figure 6.6 – Effect of input stage response on calibration constant and noise. 100 kHz input sampling rate. Top Left) FIR cutoff = 1.5 kHz. Chopper is highly attenuated, very little aliasing noise. Top Right) FIR cutoff = 5.0 kHz. Chopper is mildly attenuated, aliasing noise is present. Bottom Left) Analog and Digital system preamplifiers are used. Polarizer is attenuated. Bottom Right) Analog System scan. Polarizer is heavily attenuated.

observable in the measured calibration constant. With the higher cutoff frequency, the FIR filter does not further attenuate the chopper response and the calibration constant is closer to unity. In order to force a highpass dominant response on the digital system, the analog and digital preamplifier circuits were connected in series. The analog AC coupling filter response (Figure 6.2) is now the dominant effect, and the calibration

constant has dropped below unity to 0.87. For comparison, an analog lock-in map is shown as well, indicating a baseline calibration of 0.79 for the highpass domination of its preamplifier circuit. These results indicate the analog board has an aggressive highpass characteristic with significant attenuation at low frequencies due partially to the 60 Hz rejection of the AC coupling circuit. Conversely, the digital board has a mild AC coupling circuit (just enough to remove DC) but has an aggressive lowpass response due to both the analog and FIR anti-aliasing filters. The result is the analog system attenuates the polarizer response but not the chopper response resulting in a constant less than unity. Conversely, the digital system attenuates the chopper but not the polarizer, resulting in a constant greater than unity. When the digital system is connected in series with the analog preamplifier, the constant will be between the normal values for the two systems. Despite these differences in calibration constant, the maps essentially look the same which validates the hypothesis that any reasonable bandpass system can be used for DOP measurements.

The noise level can be obtained by calculating the standard deviation over a suitable area that should have zero DOP variation. For the maps in Figure 6.6, noise was measured in the lower left corner over a 15×4 sample region. It is worth noting that the noise level in the digital scans was superior when the lower cutoff frequency was used, despite a larger deviation from unity for the constant. This can be explained if we recall that the reduced rate in the polarizer PSD for this system configuration is 6.4 kHz and aliasing occurs above 3.2 kHz. The results indicate the 5.0 kHz cutoff is inadequate and aliasing is being realized as noise in the measurements. The 1.5 kHz filter is attenuating

our signal of interest but it is attenuating the higher (unwanted) frequencies to a greater degree. The purpose of the calibration constant is to normalize the DOP readings such that they are valid for any input stage frequency response. While the actual value of the calibration constant does not impact the validity of the DOP map, it may provide insight into why certain configurations are noisier than others.

The calibration constant may vary slightly from test to test (± 0.01 for both analog and digital systems) but will also vary on the digital system depending on the demodulation wave form by ± 0.02. This is due to the small but measurable changes in the PSD output for different demodulation wave spectrums. The effect the demodulation function has on PSD magnitude will be discussed in detail in section 6.6.

6.4 Output Stage Characterization

The output filtering stage in a lock-in amplifier is used to perform the lowpass filtering of Eq. (3.6) to get Eq. (3.7). In the analog lock-in amplifier system, the output of the switched multiplier is sent to a 2^{nd} order Sallen-Key lowpass filter. The output of the filter is then digitized by an ADC for computer processing of the DOP map. Conversely, in the digital phase sensitive detector from Chapter 5, the output samples of the multiplier are stored in a memory array. An easy way to extract the DC component from the sample set is to compute the arithmetic mean. Other options are available, such as weighted averaging but this would require more hardware resources, or a higher bandwidth connection to a host PC for real-time software processing. We can model the frequency response of the uniformly weighted arithmetic mean operation by investigating the

response of an equivalent FIR filter. For an n-sample arithmetic mean, it is equivalent to an n-tap FIR filter where the coefficient for each tap is $1/n$. For the digital system, the frequency response of the mean-equivalent FIR filter is a function of both n, and the sampling rate in the filter. The analog lock-in uses a filter with a time constant of 100 ms to extract the DC component. For the DOP maps shown above in Figure 6.6, the polarizer and chopper were set at 230 Hz and 1070 Hz respectively. PSD was done with 32 samples per period and 3072 samples per spatial location. Using Eq. (6.3), the reduced rate for the polarizer PSD is 7.36 kHz which will be the sampling rate for the mean-effective FIR filter. The magnitude response of both the Sallen-Key filter and 3072-point arithmetic mean are plotted in Figure 6.7. Notice the superior rejection of the analog filter versus the mean-equivalent FIR response.

An analog filter's response decays quickly as a function of $e^{(t/RC)}$ where RC (resistance × capacitance) is the time constant and t is time in seconds. If the time constant is not significantly less than the scanning time per spatial location, information from a particular location will be smeared across several following sample points in the DOP map with exponentially decaying amplitude. This will appear in the map as a smoothing out of noise but also may reduce the resolution of fine spatial detail. The primary frequency response of the analog filter is of course fixed at design time (with further digital averaging by the host computer), whereas the response of the digital mean calculation is directly dependent on how many samples are included in the mean, and thus how long a particular location is scanned. As a result, when scanning time is short there is a minimum bandwidth associated with the analog system that is not achievable in the

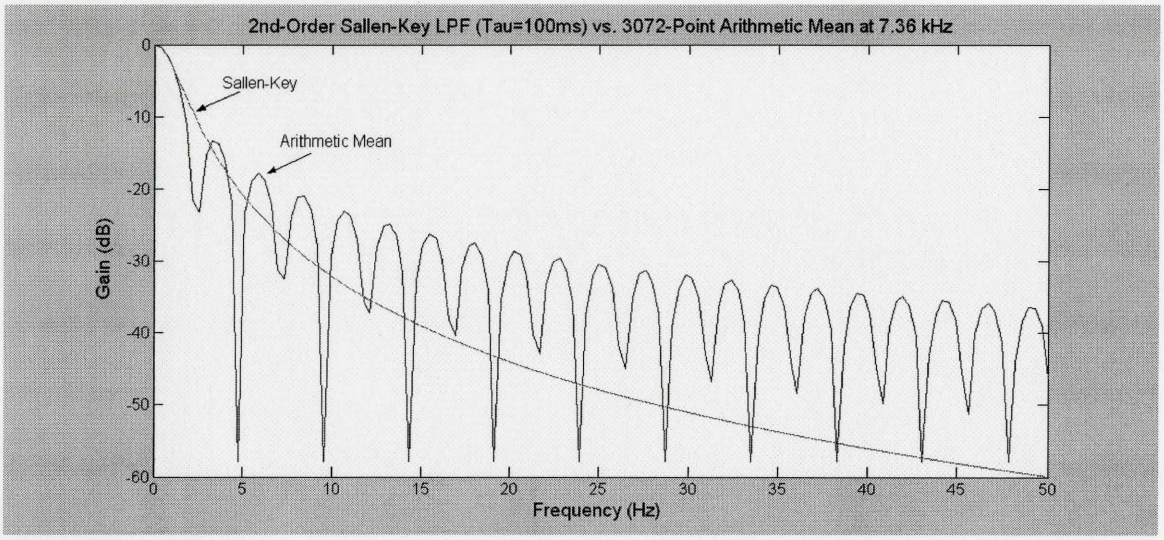


Figure 6.7 – Magnitude response of 2nd order analog LPF versus mean-equivalent FIR filter. Time constant = 100 ms for the analog filter, 3072 points in the FIR filter at 7.36 kHz sampling rate.

digital system. In Figure 6.7 above, the analog filter has a -3dB point at 1.6 Hz and reaches -40 dB at approximately 16 Hz. The digital system also reaches -3dB around 1~2 Hz but does not have a sharp roll-off, reaching only -35 dB at a much higher 50 Hz. The noise present in a DOP map is a direct result of the non-zero rejection of low frequencies near DC. Thus, if the scan time per location is not significantly larger than the time constant of the analog filter, it is not possible for the digital system to achieve the noise reduction performance of the analog system. For comparison, plotted below in Figure 6.8 is the analog filter response versus 1024 and 5120 point arithmetic means for a 7.36 kHz reduced rate.

If there is a lot of signal content in the low frequencies near DC, it will be very difficult (even with long spatial sampling times), to guarantee better performance from the digital system based purely on evaluating the performance of the output stages. This

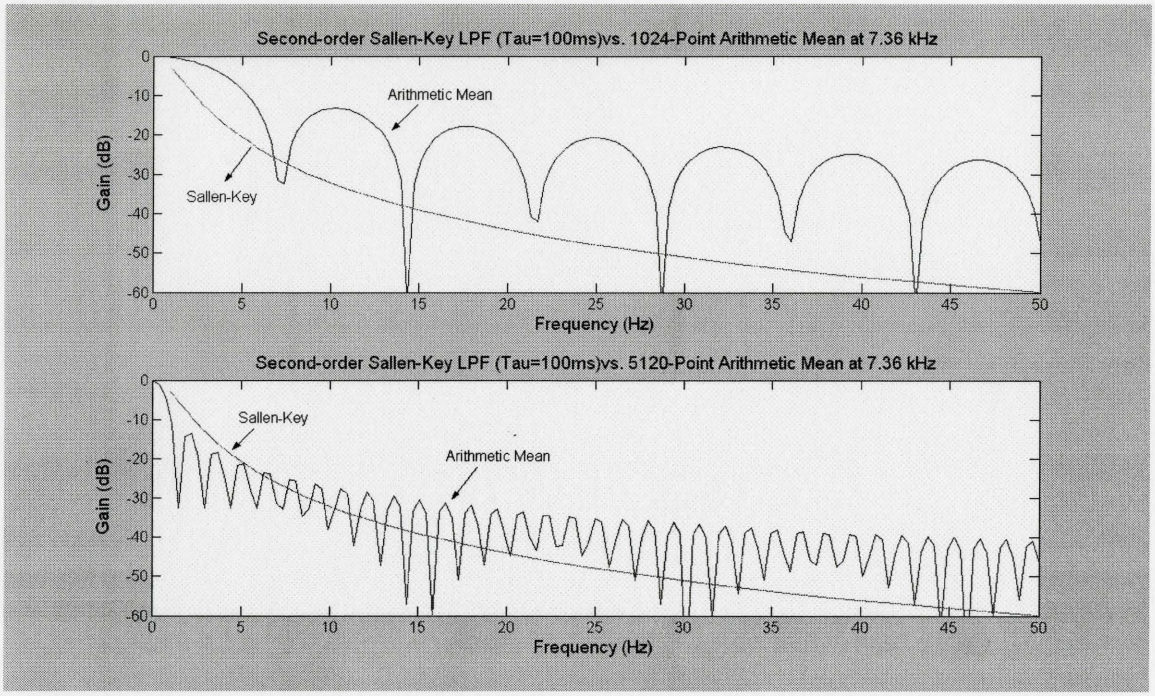


Figure 6.8 - Magnitude response of 2nd order analog LPF versus mean-equivalent FIR filter. Time constant = 100 ms for the analog filter. Top) 1024 points in the FIR filter. Bottom) 5120 points in the FIR filter.

pessimistic view will not always be the case however, since the digital system can gain a noise advantage by using an optimal demodulation wave form discussed later in this chapter. However, despite the added noise in the digital DOP maps, they are fundamentally more accurate (if not precise) since no information from the previous spatial locations is included in the measurement for the current location. This is accomplished by flushing the contents of the sample memory array within the PSD multiplier for each new spatial location. In order to achieve both high accuracy and low-

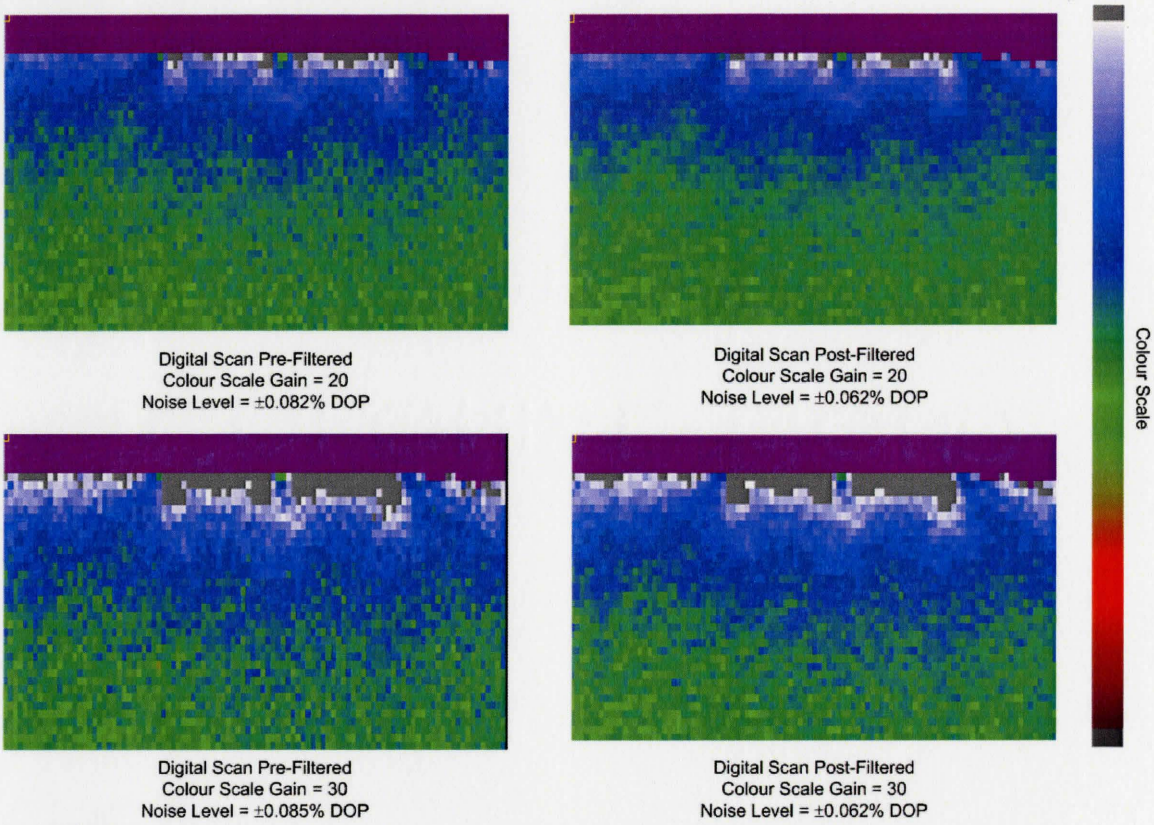


Figure 6.9 – Post-Processing of noisy data. Top Left) Noisy digital scan with default colour scaling. Top Right) Scan has been post-processed with a 3-tap LPF {0.15, 0.70, 0.15} to achieve 25% reduction in noise figure. Bottom Left) Noisy scan with 50% greater colour scaling to emphasize noise. Bottom Right) Filtered scan with 50% greater colour scaling.

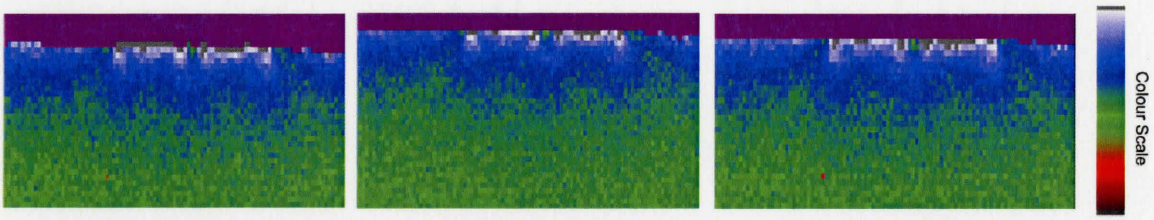
noise we must unfortunately utilize long scanning times per spatial location regardless of choice of analog or digital lock-in. Be aware that if the system contains $1/f$ noise, increasing the averaging time will not be totally effective and only the Gaussian noise in the system will continue to diminish. Fortunately, if we are simply interested in a clean, low noise scan with as short a scan time as possible, then the smoothing effect of the analog filter by smearing data across multiple locations can be approximated from a noisy digital scan by post-processing the data with a simple spatial lowpass filter after the map

(or dataset) is completed. Note that the effective filtering by the analog filter is not symmetric, only previous samples are included, not future samples. Alternatively a more complex, higher order 2D spatial filter can be used if desired in order to smooth out a DOP map. The before and after results of post-processing with a simple one-dimensional horizontal filter {0.15, 0.70, 0.15} are shown in Figure 6.9 above. A noise reduction of approximately 25% has been achieved without significantly reducing the detail in the image. Other image processing techniques may be applied to further enhance an image. See [49] for examples.

6.5 Samples Per Period Results

According to Chapter 4, increasing the samples per period reduces the maximum phase error during demodulation, recommending 64 samples per period to ensure < 0.5% error. However, the delayline in Figure 5.1 is used to shift the sync signal phase in increments far less than the period of one sample. In theory, the ability to fine tune the phase should allow us to achieve near perfect phase matching, and only 4 or 8 samples per period would be necessary. However, in order to guarantee phase alignment, the delayline must be capable of a net delay of an entire sample period. Decreasing the samples per period by a factor of two in general will require the length of the delayline to double while maintaining the temporal resolution of each delay step. Furthermore, we have already seen from studying the input and output stages that the samples per period plays a significant role in the design of the various filters, especially the requirements regarding anti-aliasing. At 8 samples per period, a 200 Hz polarizer will result in a

reduced rate of only 1.6 kHz, and the chopper itself would create aliasing at 1 kHz. For this reason 8 samples per period or less are simply not feasible. The results of 16, 32 and 64 samples per period are presented below in Figure 6.10. All scans were done with a 100 kHz input sampling rate and a 5 kHz cutoff in the FIR downsampling filter.



16 Samples Per Period
Calib = 1.55, FIR Cutoff = 5.0 kHz
Noise Level = ±0.092% DOP

32 Samples Per Period
Calib = 1.54, FIR Cutoff = 5.0 kHz
Noise Level = ±0.082% DOP

64 Samples Per Period
Calib = 1.54, FIR Cutoff = 5.0 kHz
Noise Level = ±0.081% DOP

Figure 6.10 – Comparison of samples per period. Changing the samples per period had little impact on the overall DOP map. The noise figure for 32 and 64 samples per period was the same, however at 16 samples per period the noise increased by 12% but is statistically insignificant. The vertical offset between the images is due to cumulative slip in the translation motors over a large number of scans.

It becomes clear that the phase error expected for 32 and 16 samples per period (approximately 8% and 2% respectively from Chapter 4) are negated by the precise phase tuning of the delayline. For 16 samples per period, some aliasing may have caused the noise figure to increase by 12% versus 32 samples. At 64 samples no significant gain in noise reduction is apparent. In order to determine if 12% is statistically significant, a confidence interval was calculated over the 15×4 sample region from the 32 sample per period map assuming the variance has a χ^2 distribution. For 59 degrees of freedom the 95% confidence interval is $[72.4 \leq \sigma \leq 104.2]$ where σ is the standard deviation so the change in noise level between 16 and 32 samples per period is statistically insignificant.

This analysis indicates that 8 samples per period should not be used however the user may choose 16, 32 or 64 samples per period with high confidence that phase error will not be a problem. As a final note, in the 32 sample map in Figure 6.10, the red dot visible in the other two scans is obscured by a random noise of the opposite DOP polarity and appears suppressed in the colour-scaled mapping.

6.6 Demodulation Waveform

With the existing analog system, we are limited to a switched-multiplier. This implementation limits our demodulation wave to be a square wave. Furthermore, if the reference sync signal is not precisely 50% duty cycle, it will need to be regenerated by a phase-locked loop in order to avoid introducing error. Since the digital system only uses the rising edge of the sync signal, the period must be stable however the duty cycle is irrelevant. Furthermore, the look-up table based reference generation allows us the flexibility to choose any demodulation wave function we desire. Typical functions for demodulation are cosine, triangular and square wave forms. A cosine will have a narrowband response at a single fundamental frequency. The triangle and square waves will also contain odd-harmonics of the fundamental frequency, with the harmonic being weaker for a triangular wave. The three wave forms were tested and the results are shown below in Figure 6.11.

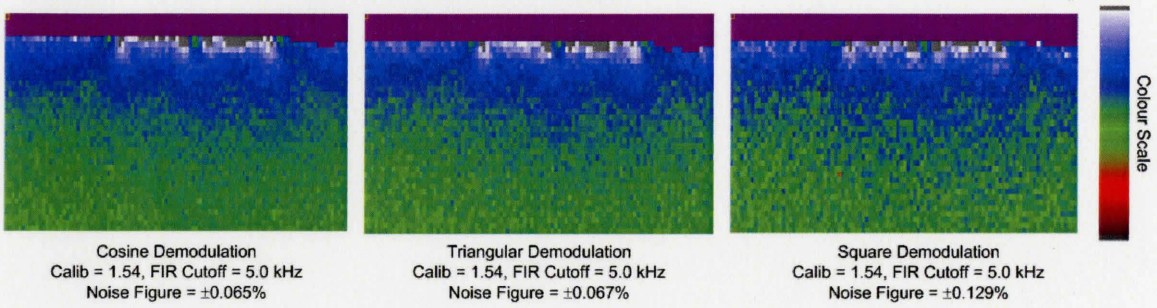


Figure 6.11 – DOP maps for cosine, triangular and square wave demodulation.

In examining the results it appears that for lowest to highest noise level, the functions rank as cosine and triangle (essentially same noise level), and finally square. This can be explained by examining the power spectrum of these functions, plotted below in Figure 6.12. As mentioned above, the sine wave contains a single frequency component at the fundamental frequency. Triangle and square waves contain a number of sidebands at odd multiples of the fundamental frequency, however the sidebands of a triangle wave have much less power than the sidebands of a square wave. During demodulation, all frequency components in the wave forms will shift the corresponding frequencies in the signal of interest down to DC, and will contribute to the DOP measurement. If the signal of interest actually contains more information in those spectral sidebands than noise, then the final measurement will be greater in magnitude and achieve better signal-to-noise ratio. However, if only noise is present at those frequencies, then the final measurement will be noisier if the unwanted frequencies have varying amplitudes (constant amplitude would appear as an offset and be calibrated out).

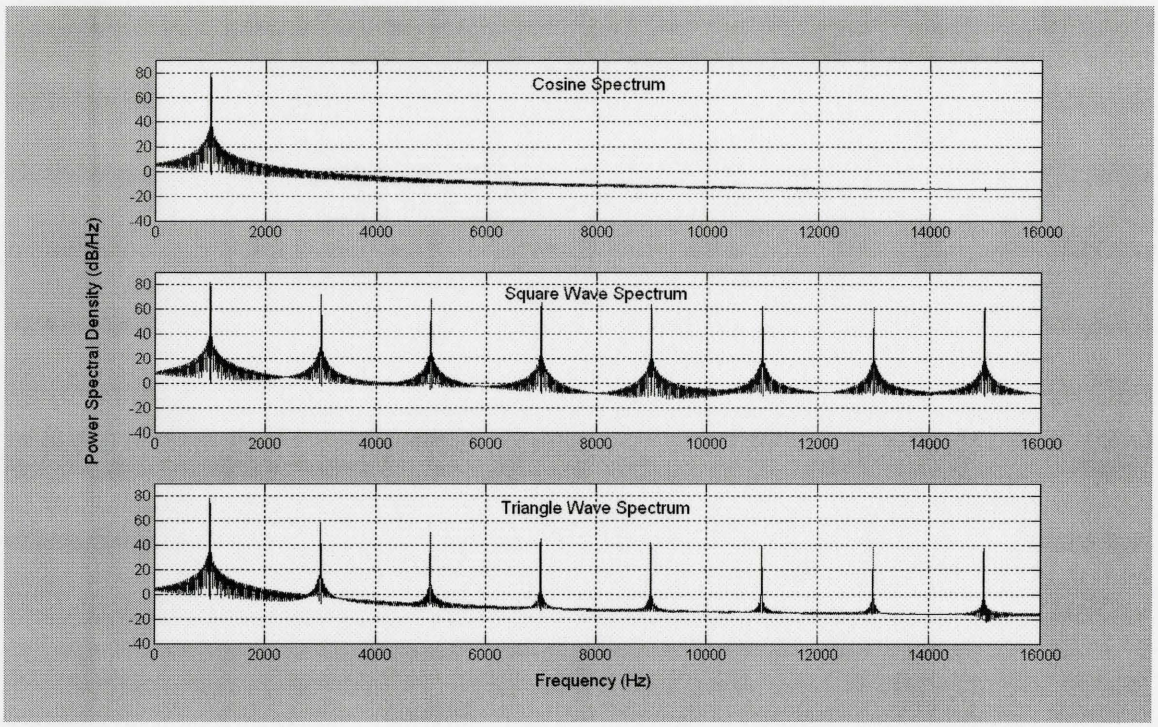


Figure 6.12 – Spectrums of 1000 Hz cosine, square and triangle demodulation wave forms. Square and triangle waves have odd harmonics of the fundamental frequency. The power of the triangle sidebands is much weaker than the square wave.

The square wave is the most likely to capture noise and shift it down to DC and has the worst performance as seen in Figure 6.11.

This suggests that the optimal solution for general lock-amplifiers is to generate a wave form with precisely the same spectral profile as the signal of interest. We can do this on our digital system by using the data capture functionality to obtain the time-domain function of the chopper and polarizer. By statistically processing these periodic signals we can create a best approximation to its spectral content. Unfortunately this method is not applicable if the signal of interest has its fundamental frequency below the frequency at which we wish to demodulate it. For example, the rotating polarizer is

expected to have strong responses at both the mechanical frequency and twice the mechanical frequency, though we are only interested in the 2nd harmonic. Even so, it is not possible to include the 1st harmonic in the demodulation since its period would be two reference sync periods. In the DOP experiment, since we only wish to demodulate the 2nd harmonic of the polarizer a cosine wave is the best choice. It may be better to use a custom wave for the chopper PSD, but since it appears for the DOP experiment that very little gain is achieved between the cosine and triangle waves, it is unlikely a profiled wave would provide significant improvement. For this reason, it was not included in the characterization.

6.7 Performance Summary

The five user controllable parameters described at the beginning of this chapter can have a significant impact on the performance of the system far beyond that described in Chapter 4. Best results were obtained using 32 samples per period (16 requires larger delayline and 64 more PSD resources), 16-bits per sample, 100 kHz sampling rate, 1.5 kHz cutoff in the FIR lowpass filter and cosine demodulation.

For quick scans where high precision is not required, shorter scanning times per location (approximately 0.25 seconds) can be utilized in combination with post-processing to reduce noise. High precision scans with excellent reduction can be obtained using moderate scan times (approximately 0.55 seconds), while scans where noise level is the highest priority, there is no choice but to use long scanning times per location (approximately 1.0 seconds).

An example of the results achieved using the above configuration with 0.55 second per location scan time is shown below in Figure 6.13 including PL, DOP, ROP and HeNe maps. Optimal polarizer and chopper frequencies of 230 Hz and 1070 Hz respectively are chosen based on data presented in Chapter 8. The device scanned in Figure 6.13 is a ridge waveguide laser on an InP substrate. The HeNe scan provides accurate detail of the central rib, the channel to each side and the uneven etch on the right side (these features are visible on scans of all devices from the sample bar). The HeNe

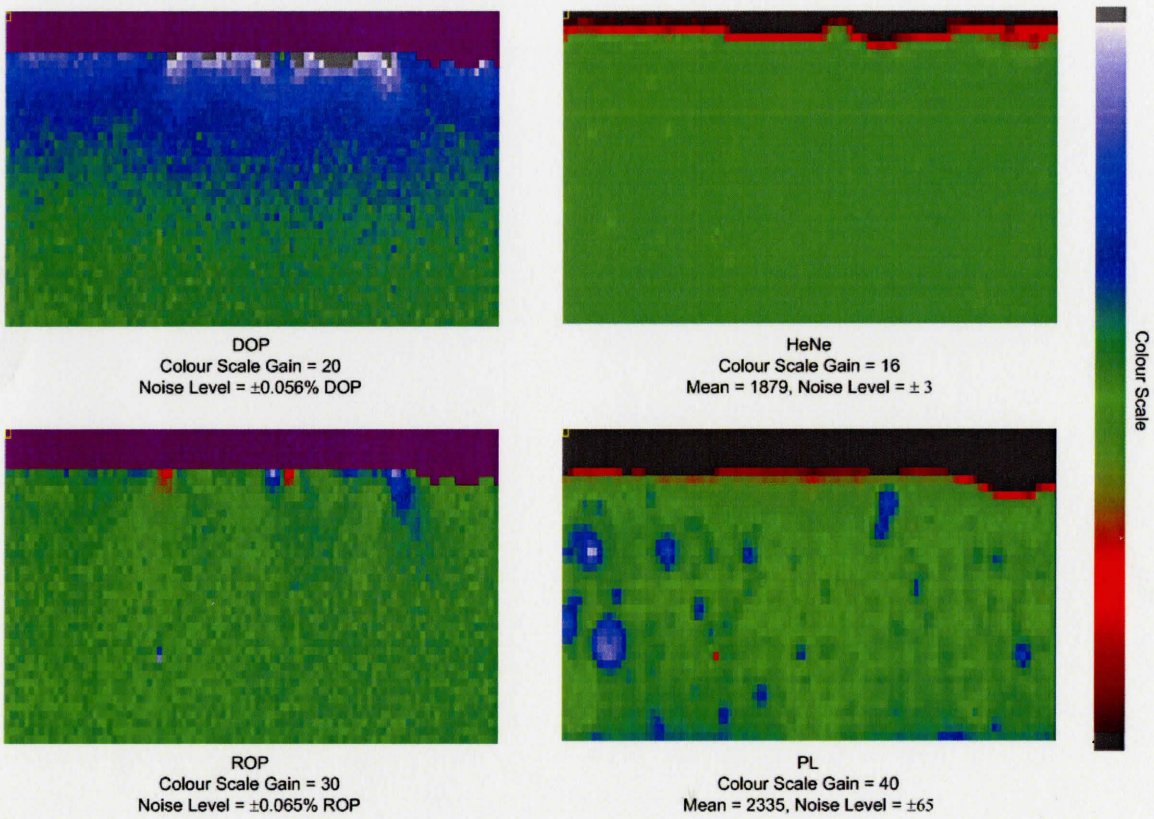


Figure 6.13 – Optimized Results. DOP, ROP, HeNe and PL maps are presented. The colour scale offset for the DOP/ROP maps was set by calculating the mean in a uniform, unstrained area. The scale offset for the HeNe and PL maps was set by taking the mean of the entire map.

also reveals that the substrate reflects light with high uniformity. The PL map reveals areas of peaked intensity with larger Gaussian areas that have been previously overheated by the pump laser during alignment. The DOP map shows the relative direct strains caused by the channel etchings. Three pronounced lobes appear corresponding to etching boundaries. The ROP map reveals the shear strain, indicating opposite polarity at each side of the central rib as well as at the outer channel edges.

Chapter 7. Reflection HeNe Maps

7.1 Concept

A new addition to the DOP experimental apparatus is a secondary photodetector and the associated optics to collect the reflected HeNe pump light from the sample under study. Referring to Figure 2.3, a neutral density filter (optical density = 0.1) is inserted between the pump laser and the optical chopper. This filter must be placed in this location in order to spectrally separate the pump light from the reflected light. The cold mirror in front of the primary photodetector will reflect most of the HeNe light and transmit most of the IR wavelengths. Thus, the light returning all the way back to the neutral density filter should contain only reflected HeNe from the sample facet and very little photoluminescence. It will have passed through the optical chopper so it will be modulated at the chopper frequency. Unfortunately, some pump (non-reflected) light will also make its way to the secondary photodetector so we cannot simply take the photodetector output to be the reflected signal. Since the pump light will be a DC component, and the reflected light will be at the chopper frequency (approximately 1000 Hz), we can once again use a lock-in amplifier to extract the signal of interest and reject all others. Since the digital FPGA platform is expandable and upgradeable, the second DLIA is easily added to the design as described in Chapter 5. The only other modifications necessary are to add a second input stage similar to the one in Figure 6.1. This requires another series of AC coupling and preamplifier circuits, as well as another ADC. It is important to note that if the neutral density filter were placed at any other

point in the optical path, both pump light and reflected light will be modulated by the chopper and spectrally isolating them would be impossible.

7.2 Mapping Device Structure

There is no guarantee that the device structure on the sample will have a 1:1 relationship with photoluminescence or even any relationship with the PL at all. Conversely, as long as the device structure is reflective, it will be visible in a reflection HeNe map. In addition to resolving fine detail in the etching, other features such as metallization layers and bonding solder may be visible with different reflectivites. Figures 7.1 and 7.2 below show some comparisons between reflection HeNe maps and their associated PL maps. The devices shown in the figures above are ridge waveguide lasers creating by channel etching. In Figure 7.1, the central ridge is visible as well as channels to either side. The channel on the right is unevenly etched (concave instead of flat) with respect to the left channel. None of these details are discernable in

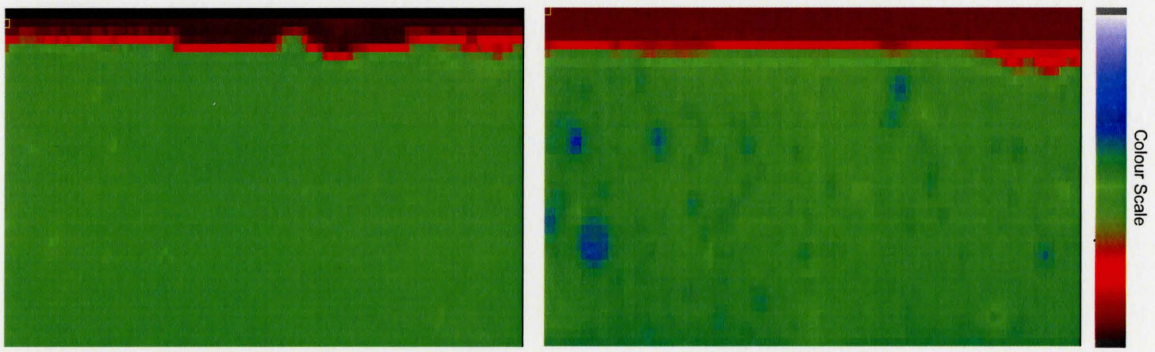


Figure 7.1 – Wide angle view of a single device with reflection HeNe and PL maps. The colour gain has been set such that the darkest regions will not saturate the scale.

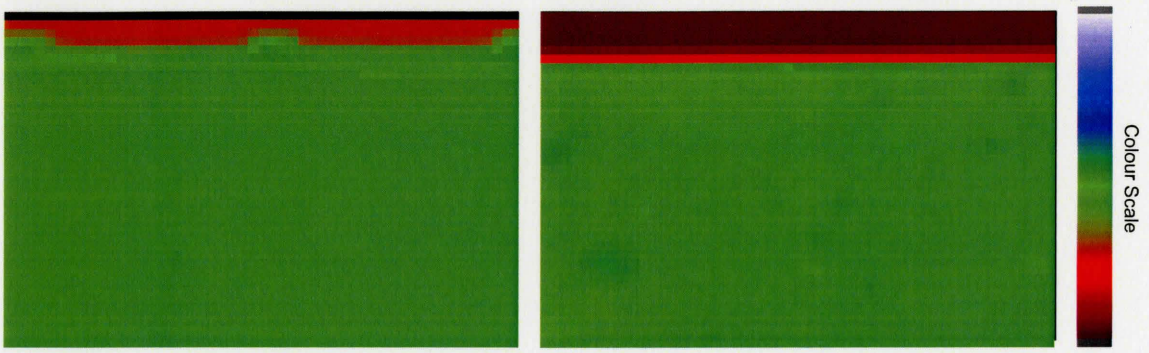


Figure 7.2 – Close-up view of a single device with reflection HeNe and PL maps. The colour gain has been set such that darkest regions will not saturate the scale.

the PL map. In Figure 7.2, a close-up scan of a similar device with symmetric channel etching is shown. Once again, the device structure (central ridge and channels) has high visibility in the reflection HeNe scan but is completely invisible in the PL map. Since the DOP/ROP map patterns are often caused by the manufactured device structure, the ability to see this structure explicitly in the HeNe map versus inferring it from the DOP/ROP maps is a useful tool. For comparison, an optical photograph of the device in Figure 7.2 was taken with a 125× objective lens to show the channel structure along the middle of the device (Figure 7.3).

7.3 Sample Alignment

Another use of the reflection HeNe scan is to confirm sample alignment once a scan is completed. This can become useful if very quick scans of the sample are done first over a small region where only the reflection HeNe is evaluated. The map will provide feedback with much greater resolution than typical PL aligning procedures. Once

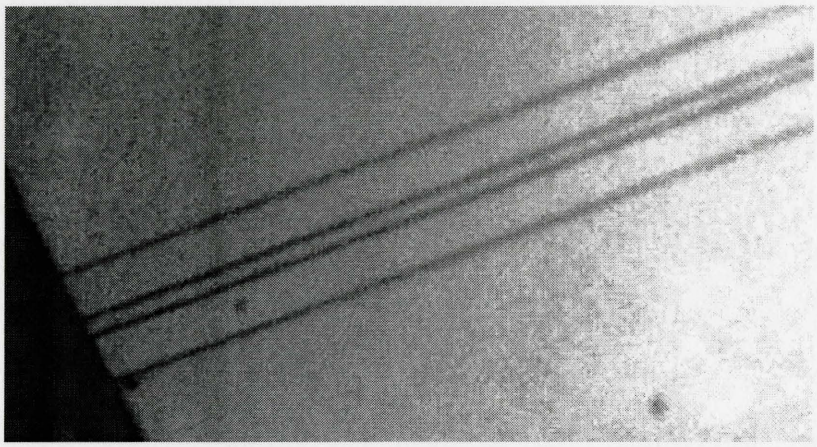


Figure 7.3 – A photograph of the device in Figure 7.2. Top view, shown using a 125× objective lens.

suitable alignment is achieved, the longer full scan can be initiated with confidence. Figure 7.4 below once again compares reflection HeNe and PL maps in order to assess sample alignment. The sample was aligned in the traditional fashion using PL readings. The figure below indicates that it takes extreme gain in the colour scale in order to discern that the alignment of the sample is slightly tilted on the diagonal. The gain on the PL map can be turned up to nearly the point of saturation and the tilt angle is not obvious. The HeNe map is far less susceptible to variations that are not alignment dependent which can translate into higher reliability when taking alignment readings versus PL. Using the traditional method of manually assessing PL readings and making stage rotation adjustments to align a sample will provide suitable map results. However, due to alignment-independent variation in the PL a lot of human judgment is required to discern the difference between facet anomalies and misalignment. It would be difficult to create a computer algorithm to simulate these judgment calls. Fortunately, the increased precision and consistency of the HeNe map could be used to automate the alignment

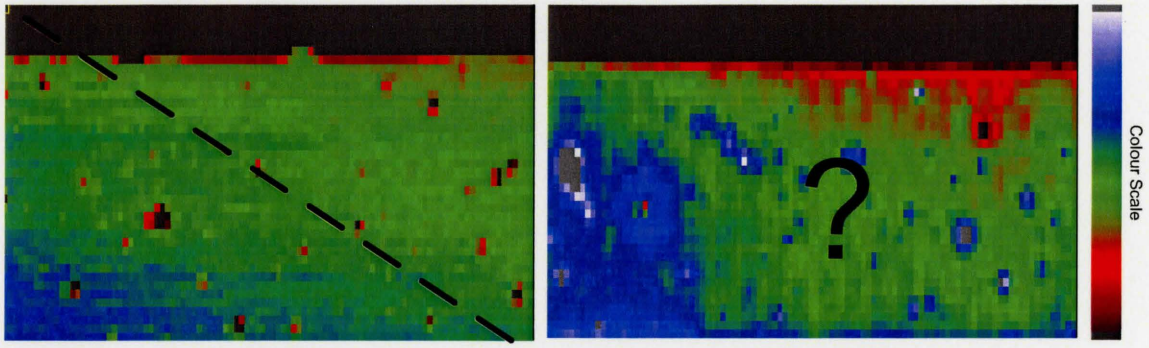


Figure 7.4 – Sample alignment maps of reflected HeNe and PL. Left) with the colour gain set at a very high level a slight diagonal tilt about the indicated axis is evident. This sample is not perfectly aligned but is likely acceptable. Right) with the colour gain set near saturation, the precise alignment of the sample is not obvious.

process entirely with a simpler algorithm if the 3-axis rotation stage can be adjusted under computer control using servo motors.

Chapter 8. Spectral Analysis

8.1 Frequency Spectrum of the DOP Experiment

In Chapter 3, it was shown (Figure 3.2) that the bandwidth of a lock-in amplifier is equivalent to the bandwidth of the post-multiply lowpass filter. For DOP experiments we are primarily interested in the DC output of phase sensitive detection, and any nonzero frequencies within the filter passband Δf are considered to be noise. Thus, we can maximize the signal-to-noise ratio by ensuring no interfering signals are within Δf from the frequency (signal) of interest. To analyze the frequency spectrum, 2^{16} sequential samples were captured from the output of the analog-to-digital converter (@ 100 kHz) and used to calculate the power spectral density. Figure 8.1 shows a power spectral density plot of the photodetector response (after the preamplifier stage and ADC) when the optical beam is focused on a sample location far from any strain inducing effects. Since the DOP/ROP should be zero (the light is equally polarized in every possible x-z direction) then the rotating polarizer should have no effect other than a uniform attenuation across all polarizations. Accordingly, we expect the power spectral density to contain components at the chopper and polarizer frequencies, as well as their sum and difference sidebands in accordance with Eq. (3.4) representing amplitude modulation. It is clear from Figure 8.1 that there are many more interfering components that are too strong and regularly spaced to be Gaussian noise or electrical power line harmonics. The stronger, unidentified components appear to be about 25 to 30 dB less than the chopper signal.

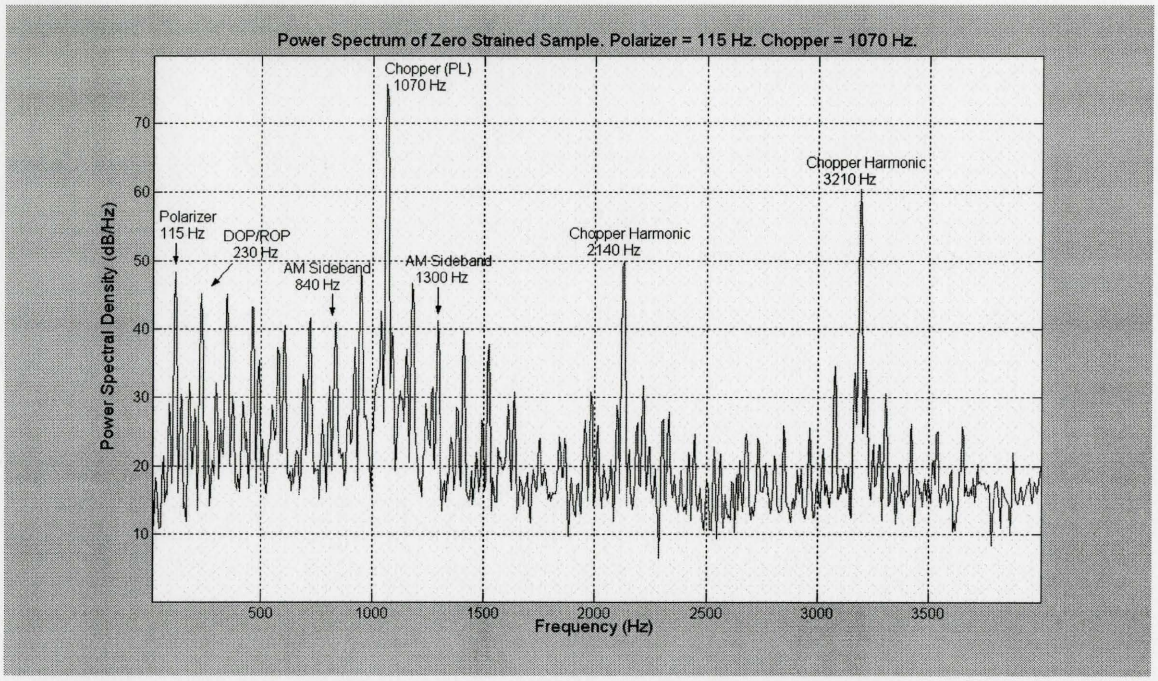


Figure 8.1 – Power spectrum of zero-strained sample. The chopper signal is amplitude modulated by the rotating polarizer producing AM sidebands. Two harmonics of the square wave chopper are indicated. The remaining signals are too strong and regularly spaced to be random noise.

In Figure 8.2, the laser beam has been blocked and the power spectral density will identify the electrical noise (power line harmonics) present in the detector and lock-in amplifier. In Figure 8.3, the measurements were repeated with the beam on, but the chopper and rotating polarizer glass are removed (polarizer motor itself is still present and rotating). The power spectrum is very similar to Figure 8.2 since the unmodulated beam is predominantly a DC signal. This figure also includes any other harmonics that may be present in the pump laser itself caused by cavity instabilities and the electrical noise in the laser power supply. It appears that the spurious components in Figure 8.1 are not caused by the laser, detector or lock-in amplifier.

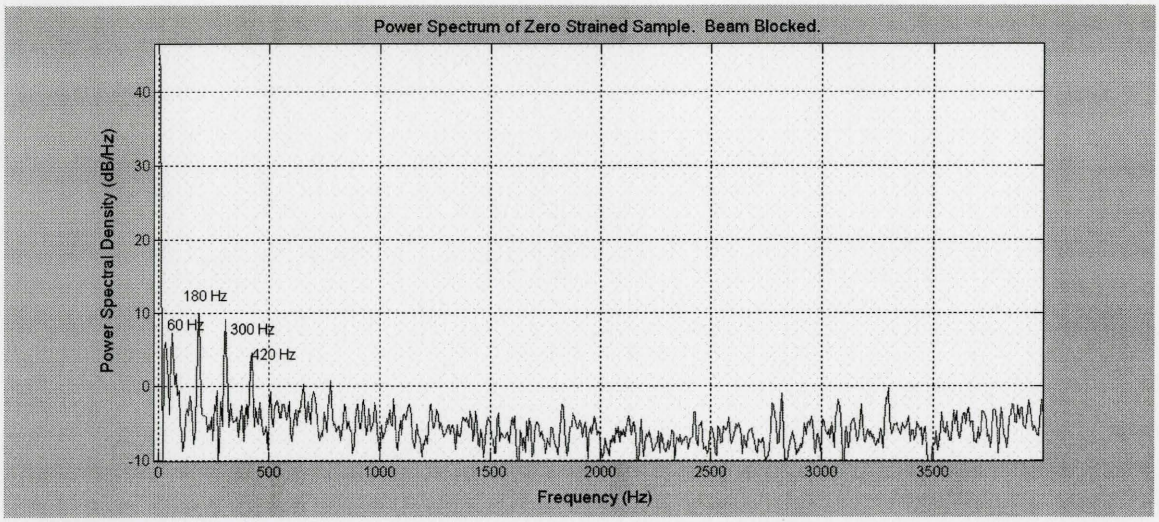


Figure 8.2 – Power spectrum of a blocked beam. The odd harmonics of the electrical power line are the most strongly present. The first four odd harmonics are labeled.

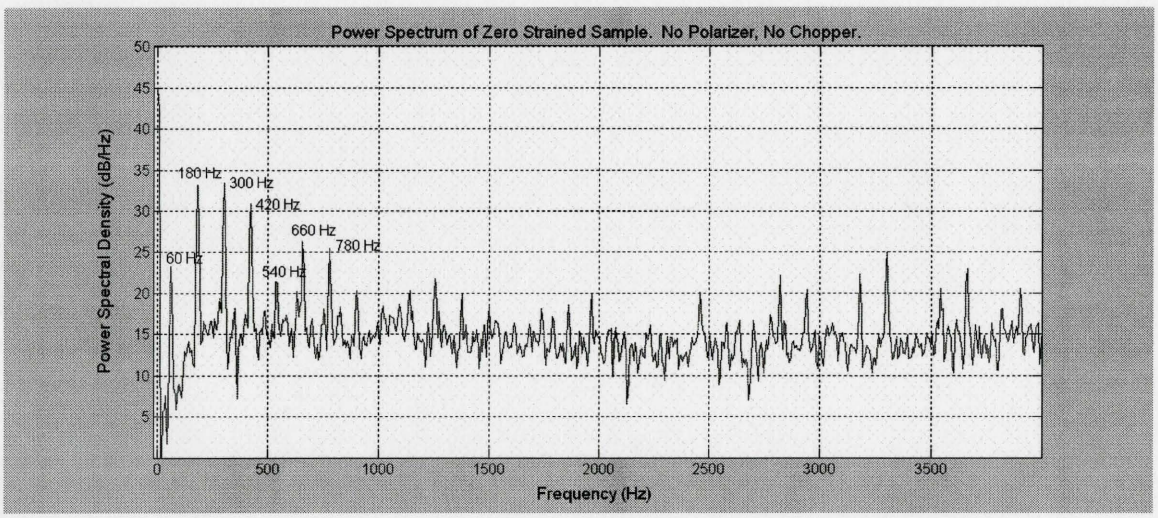


Figure 8.3 – Power spectrum with no chopper and no polarizer. Many regularly spaced components are present but careful inspection shows they are all power line harmonics. Several frequencies are labeled.

In Figure 8.4, the chopper has been put back in the optical path with a mechanical frequency of 1070 Hz. The beam is now amplitude modulated where the modulating function is approximately a square wave with 50% duty cycle. The modulating wave is not a perfect square wave for two reasons. The optical spot beam has a non-zero area, which means the chopper blade takes some non-zero amount of time to move across the beam profile. Secondly, the photodetector and lock-in amplifier do not have infinite

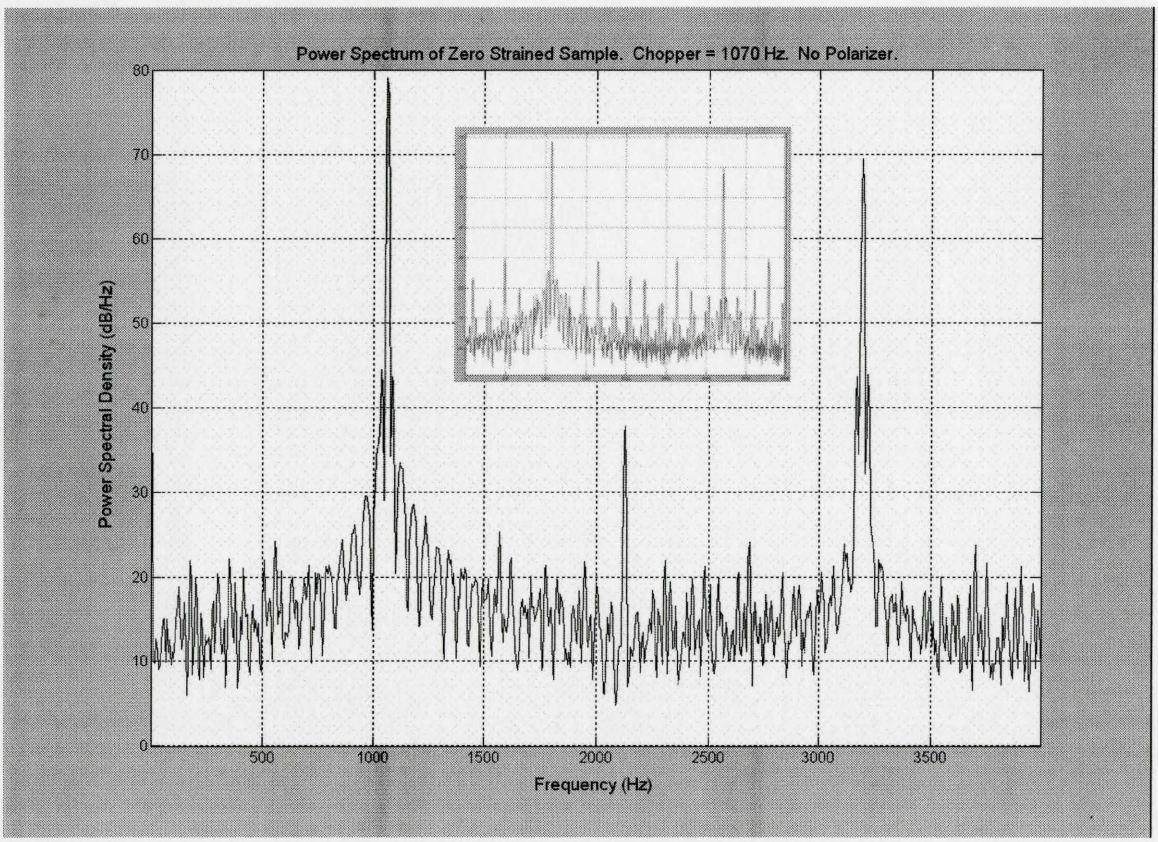


Figure 8.4 – Power spectrum of a zero-strained sample, chopper only. Chopper is 1070 Hz, no polarizer. Only the fundamental frequency and harmonics of the chopper are present. The power spectrum of a pure 1070 Hz square wave of unit amplitude is shown inset. Notice that there are several frequencies in the chopper signal very close to the fundamental between 20 and 45 dB (eg. 1070±25 Hz) caused by the imperfect shape of the chopped wave.

bandwidth, so high frequency components will be attenuated. The power spectrum of a perfect 1070 Hz square wave with a mean of half its amplitude, 50% duty cycle, at 100 kHz sampling rate is shown inset above in Figure 8.4 for comparison. The measured square wave does have a number of components in addition to the fundamental frequency; however it is evident that despite the imperfect chopping and finite bandwidth in our system they are not the cause of the spurious components in Figure 8.1.

We evaluate the power spectrum when the chopper is once again removed, but now the polarizer is placed in the beam with a mechanical frequency of 115 Hz. The test was repeated with similar results on a second independent apparatus whose results are plotted in Figure 8.5. The expectation is once again that the polarizer will only have uniform attenuation on the DC component since the DOP/ROP should be zero. After examining Figure 8.5, it is now evident that the at least some of the spurious components in Figure 8.1 are caused by the rotating polarizer glass. Even if the system does contain a non-zero DOP/ROP offset, it should appear only as a single narrowband component at twice the mechanical frequency of the polarizer. However, components have been generated at all multiplies of the mechanical polarizer frequency where the strength of the components generally decreases with increasing frequency. Their magnitudes appear to have approximately a Bessel function envelope (Figure 8.5 inset). These results strongly suggest the underlying physical effect is a form of frequency or phase modulation.

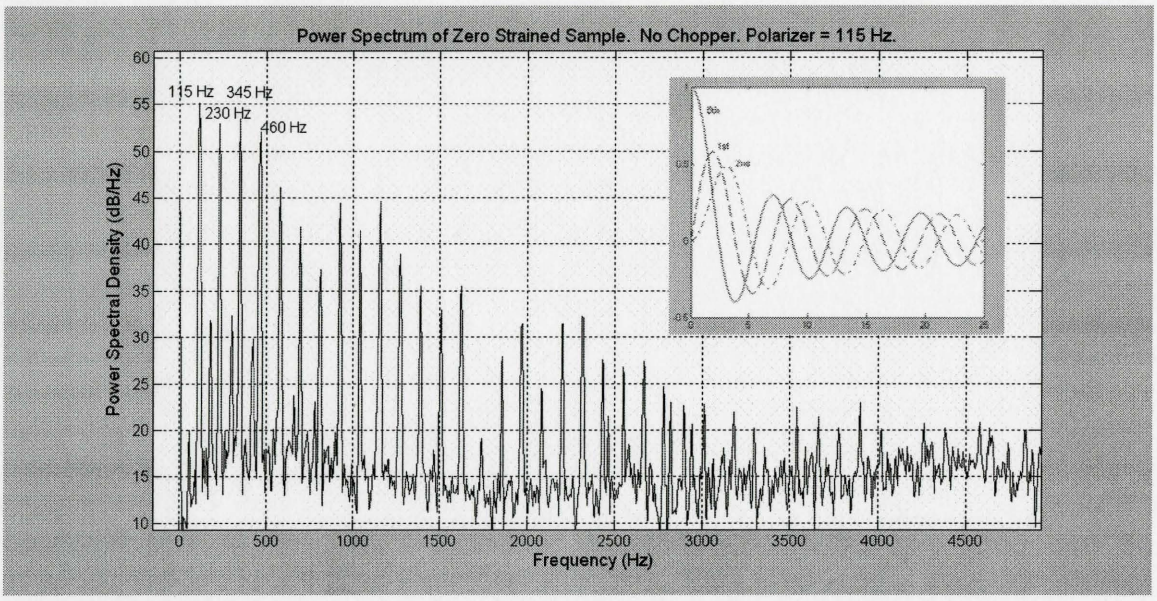


Figure 8.5 – Power spectrum of a zero-strained sample, polarizer only. The polarizer is rotating at 115 Hz. No chopper is present. The spacing of the harmonics is also 115 Hz. The magnitude of the components as frequency increases suggests it is related to Bessel functions (inset shows 0^{th}, 1^{st} and 2^{nd} order functions). Note the strongly attenuated harmonic around 1725 Hz.

8.2 Mechanical Phase Modulation of Optical Waves

An optical EM wave can be modulated by varying the index of refraction of the transparent material it is propagating through. Modulation can also be achieved if the beam is propagating through two or more mediums with dissimilar refraction indices and the ratio of the propagation distance through the mediums is altered in time. If we are interested only in the total change of the wave phase, rather than its actual phase value at a given space and time, this scenario can be simplified to traveling a fixed distanced through the two mediums where the spatial location of their interface is changing in time. In other words, the ratio of distance traveled through each medium is time varying. This

is illustrated in Figure 8.6 for the case of a sinusoidally varying interface location between air and glass.

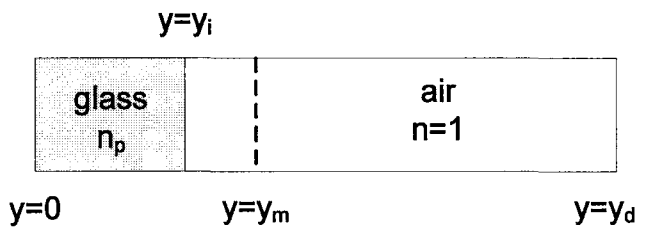


Figure 8.6 - Simplified model of time-varying interface location. Assume a sinusoidally changing position of the interface between two materials with dissimilar indices of refraction. The mean location of the interface is $y=y_m$, the location at any given instant of time is $y=y_i$ and the location of the photodetector is at $y=y_d$. The index of refraction of the glass is n_p.

If we assume an initial phase of zero at time zero and ignore wave polarization then the y-propagating electric field of an optical wave is generally

$$E(y,t)=\mathrm{Re}\{e^{j\phi}\}=\cos(\omega t-ky)=\cos\left(\omega t-\frac{2\pi n}{\lambda}y\right), \tag{8.1}$$

where ω is the angular frequency, n is the index of refraction of the material, λ is the wavelength of light in the material (not necessarily a vacuum) and t is time. We can find an expression for the electric field phase at $y=y_d$ by considering the wave propagating in glass, where the initial phase must be determined at the boundary $y=y_i$.

$$\Phi(y=y_i,t=0)=-\frac{2\pi}{\lambda_p}n_p y_i, \tag{8.2}$$

where n_p is the index of the glass material and λ_p is the wavelength of light in the glass. Similarly, we can find the phase term at $y=y_d$ by using the angle from Eq. (8.2) as an initial condition.

$$\begin{aligned}\Phi(y=y_d,t) &= \left(\omega t - \frac{2\pi}{\lambda_0} n_0 (y_d - y_i)\right) + \left(-\frac{2\pi}{\lambda_p} n_p y_i\right) \\ &= \omega t - 2\pi\left[\frac{n_0}{\lambda_0}(y_d - y_i) + \frac{n_p y_i}{\lambda_p}\right],\end{aligned} \tag{8.3}$$

where n_0 and λ_0 are the index of refraction and wavelength in air, respectively. Substitute for y_i a sinusoidally varying position with amplitude and mean y_m and angular frequency ω_m.

$$\begin{aligned}\Phi(y=y_d,t) &= \omega t - 2\pi\left[\frac{n_0}{\lambda_0}(y_d - y_m(1+\cos(\omega_m t))) + \frac{n_p}{\lambda_p} y_m(1+\cos(w_m t))\right] \\ &= \omega t - 2\pi\left[\left(\frac{n_p}{\lambda_p} - \frac{n_0}{\lambda_0}\right) y_m \cos(\omega_m t) + \frac{n_0}{\lambda_0}(y_d - y_m) + \frac{n_p y_m}{\lambda_p}\right]\end{aligned} \tag{8.4}$$

Since we are primarily interested in how the phase will be changing in time at the detector, we take the derivative,

$$\frac{\partial \Phi(y=y_d,t)}{\partial t} = \omega + 2\pi\left(\frac{n_p}{\lambda_p} - \frac{n_0}{\lambda_0}\right) y_m \omega_m \sin(\omega_m t) \quad . \tag{8.5}$$

Equations (8.4) and (8.5) show that sinusoidally varying the position of the material interface results in sinusoidal frequency modulation (time rate of change of phase) of the propagating electrical field. Eq. (8.4) can be generalized as

$$\Phi(y=y_d,t) = \omega t - C\cos(\omega_m t) + \Phi_i \tag{8.6}$$

where C represents the depth of modulation which is a constant that is dependent on the index of refraction of the two mediums and the mean and amplitude of the modulating signal, and an initial phase constant represented by Φ_i. Since Eq. (8.5) can also be generalized to the same form, for the purposes of this discussion without actual values for the constants, the modulation can be equally considered frequency or phase modulation. These equations provide insight into interpreting the results of Figure 8.5.

If we consider that the rotating polarizer glass may not have uniform thickness, then the actual thickness seen by the propagating beam will be dependent on its location on the glass at a given instant. If the beam is not aimed at the centre of rotation, then the thickness seen by the optical beam will be changing periodically with angular frequency ω_m.

It is unlikely that the actual thickness variation as the glass is spun past the beam is monotonically sinusoidal as in our example above. However, since the thickness is periodic with angular frequency ω_m we can expand it as a Fourier series. We will assume the initial phase constant Φ_i, is zero.

$$\Phi(y = y_d, t) = \omega t - Cm(t) \tag{8.7}$$

where $m(t)$ is some periodic modulating function and C is the modulation depth. Expanding $m(t)$ we have

$$m(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(\omega_m t) + b_k \sin(\omega_m t), \tag{8.8}$$

and Eq. (8.7) can be rewritten as

$$\Phi(y = y_d, t) = \omega t - \left[a_0 + \sum_{k=1}^{\infty} a_k \cos(k\omega_m t) + b_k \sin(k\omega_m t) \right]. \tag{8.9}$$

where the constant C has been absorbed into the Fourier coefficients. Recall that Eq. (8.9) is the phase term of the optical electric field wave, where ω is the frequency of the light wave (~ 10^{14} Hz). Thus, the periodic electric field wave is phase modulated by a sum of low frequency harmonic waves. Mathematical analysis of multi-frequency phase modulation of a single carrier wave is complex and can be found in other references [50,51]. To summarize the findings in [51] and apply them to the propagating electric field, for a sinusoidal wave that is frequency modulated by k harmonics of ω_m, we find the electric field incident on the detector is

$$E(t) = A_c \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \cdots \sum_{n_k=-\infty}^{\infty} J_{n_1}(z_1) J_{n2}(z_2) \cdots J_{n_k}(z_k) \cos\left[\left\{ \omega_c (1 + c_0) + \sum_{i=1}^{k} n_i \omega_m \right\} t + \sum_{i=1}^{k} n_i \theta_i \right],$$

(8.10)

where J_n is the nth order Bessel function, ω_c is now the carrier angular frequency, and z_i and θ_i are defined as

$$z_i = \frac{\omega_c \sqrt{a_i^2 + b_i^2}}{\omega_k}, \quad i = 1, \ 2, \ \ldots k \tag{8.11}$$

$$\theta_i = \tan^{-1}\left(-\frac{b_i}{a_i} \right), \quad i = 1, \ 2, \ \ldots k \tag{8.12}$$

Some important results found by [51] are worth mentioning. The frequency spectrum is centered on ω_c, the fundamental frequency of the optical light wave in our case. Sidebands appear on either side of ω_c separated by ω_m where the harmonic amplitude is

determined from Eq. (8.10). Furthermore, Figure 8.7 below (copied directly from [51]) illustrates through example that frequency modulation can create asymmetry in the sidebands and harmonics can appear missing (or highly attenuated).

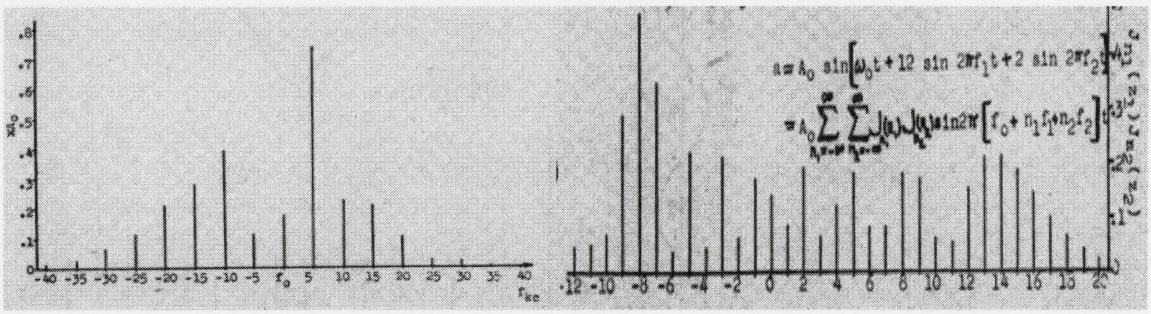


Figure 8.7 – Two-tone frequency modulation. Left) 5 and 10 kHz modulation with 10 kHz deviation. Notice the asymmetry about the carrier f_0. Right) Two-tone modulation can create attenuated harmonics that may appear missing or buried in the measured spectrum. Copied directly from [51].

The evidence from the DOP spectral analysis suggests that the off-centre focus of an optical beam onto a rotating transparent medium that is not uniformly thick will produce multi-tone phase modulation of the electric field wave. This will create a series of sidebands centred on the carrier wave frequency. However this does not explain why the sidebands appear at base band frequencies. In order to understand the low frequency power spectrum in Figure 8.5, we must consider the effect of the photodetector. When the electrical field of the optical beam is incident on the photodetector, the response is proportional to the square of the electric field.

$$V(t) \propto E^2(t), \tag{8.13}$$

where V is the voltage response from the photodetector. The squared response is important since without it the phase modulation sidebands would only exist in the terahertz range. Multiplying the electric field with itself creates sum and difference frequencies of all the sidebands in accordance with Eq. (3.4) for phase sensitive detection. It is obvious in this case that the electric field is frequency and phase matched with itself (homodyning). As a result, we would expect that the terahertz sidebands are shifted down to low frequencies such that the carrier frequency is now zero. As a final note on homodyning, since both sidebands around the carrier frequencies are involved in sum-difference generation, any asymmetry will play a role in determining the magnitudes of the low frequency components.

The overall magnitude (envelope) of the phase modulated sidebands is generally a function of the depth of modulation, which in this case is determined by the thickness variation in the rotating glass. If the variance is not a significant fraction of a wavelength, it is unlikely the effect would be strong enough to be detectable let alone interfere with our DOP measurements.

In order to confirm that the time varying thickness of the glass medium is creating phase modulation harmonics in the base band, a simulation of the model is carried out in MATLAB and presented below.

8.3 Simulation of Mechanical Phase Modulation

This chapter has demonstrated that a light beam passing through a transparent medium with periodically varying thickness will phase modulate the electric field. We

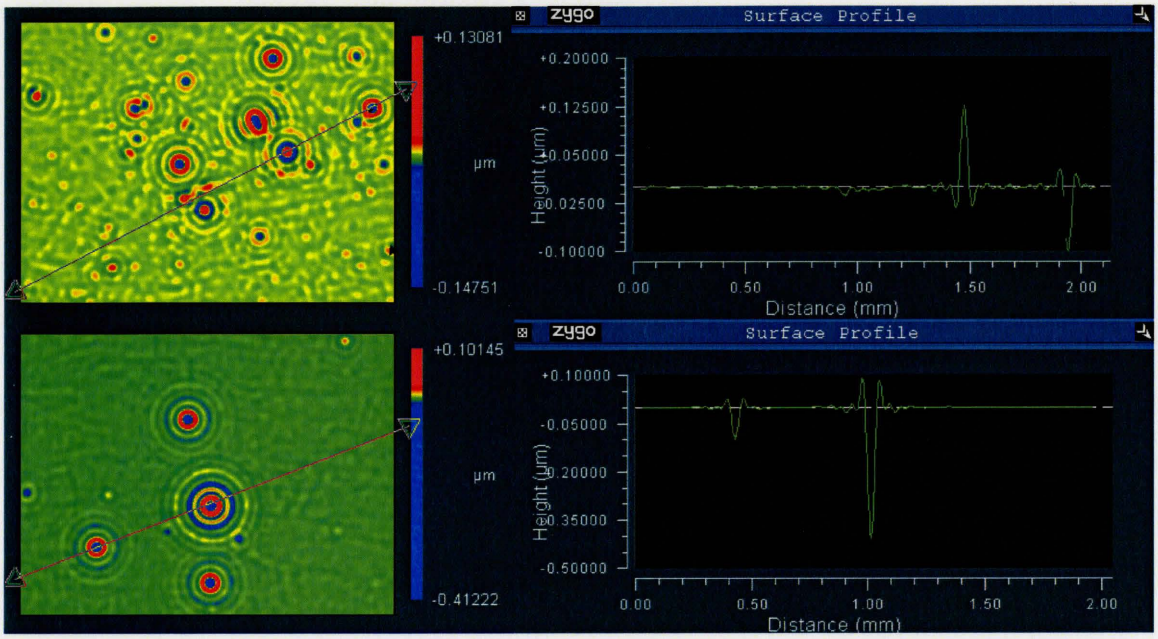


Figure 8.8 – Interferometer reflection scan of the Polarcor polarized glass. The area scanned is approximately 1.3 mm by 1.8 mm. Left) Contour maps of the glass surface. Right) Surface profile along the line indicated in the figure. The surface imperfections appear to be on the order of several hundred nanometers. The ringing around the edges is likely caused by diffraction effects and is unlikely representative of actual surface elevation.

need to discover if the variation in the polarized glass disc used in the DOP experiment is sufficient to not only create harmonics, but shift them to base band as well as seen in Figure 8.5. To investigate this, a reflection scan (Zygo Newview 6000 interferometer) of the glass was used to get an approximation for typical surface features. Figure 8.8 shows a scan of a random location on both the top and the bottom of the glass. It appears that the surface is covered with many individual imperfections, where the magnitude of the imperfection can be several hundred nanometers. Since the wavelength of the photoluminescence in the DOP experiment is approximately 900 nm, the imperfections appear to range from $\lambda/8$ to $\lambda/3$.

We can simulate the effect of the variations in the glass by creating a modulation function similar to the profiles in right side of Figure 8.8. To simulate $\lambda/8$ and $\lambda/3$ variations, the amplitude of the phase modulating function will be $2\pi/8$ and $2\pi/3$ respectively. In order to numerically evaluate Eq. (8.7), we cannot use the actual frequency of the light wave since the carrier frequency ($\sim 10^{14}$ Hz) would take far too many data samples to complete the necessary mechanical cycles at 115 Hz. However, since phase modulation is independent of the carrier frequency value, we can change it to a much lower frequency without invalidating the results. To calculate the spectrum created by the rotating polarized glass we have

$$V(t) = E^2(t) = \cos^2(\omega t - Cm(t)), \tag{8.14}$$

where $C=\{2\pi/8, 2\pi/3\}$ and a sample function for the $2\pi/\omega_m$ periodic wave $m(t)$ is shown inset in Figure 8.9. The choice for ω is arbitrary as long as $\omega >> \omega_m$. Plotted in Figure 8.9 is the spectrum of a phase modulated electric field in accordance with Eq. (8.7). The top of Figure 8.9 shows that the 10 kHz electric field has undergone phase modulation with sidebands separated by 100 Hz. Squaring the electric field has created all the 2ω centred frequencies but not the difference frequencies at base band as we were expecting. It appears that a single pass through of a varying medium is insufficient to create the base band harmonics present in our DOP experiment due to cancellations between the generated base band harmonics.

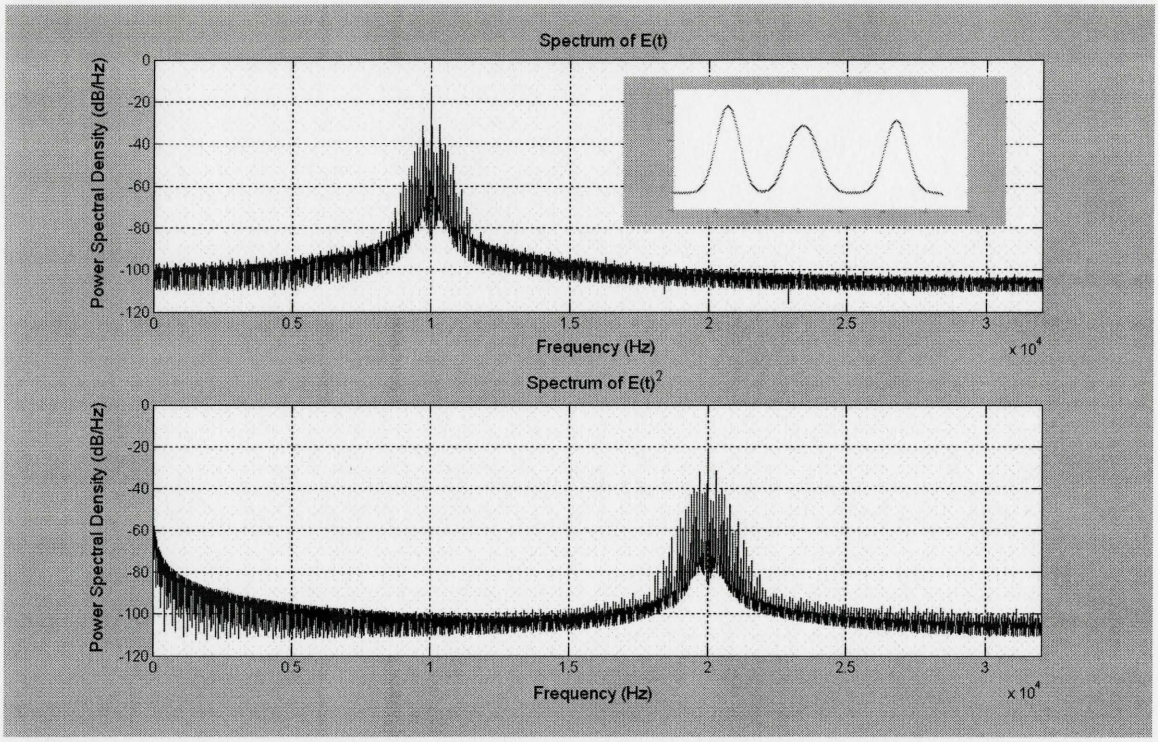


Figure 8.9 – Power spectrum of simulated photodetector output. Carrier is 10 kHz and phase modulation is 100 Hz. Depth of phase modulation is π/8. Top) Spectrum of the phase modulated electric field. One period of the normalized modulating function is shown inset. Bottom) Spectrum of the electric field squared.

However, the glass medium in our DOP experiment forms a Fabry-Perot cavity in which some of the light is reflected twice off the interior sides of the front and back facets before interfering with pass-through light. The equations for the interfering light waves are

$$E_P(t) = \cos(\omega t - Cm(t)), \qquad \text{Eq. (8.15a)}$$

$$E_R(t) = R^2 \cos(\omega t - 3Cm(t)), \qquad \text{Eq.(8.15b)}$$

$$E_T(t) = \cos(\omega t - Cm(t)) + R^2 \cos(\omega t - 3Cm(t)), \qquad \text{Eq. (8.15c)}$$

where $E_P(t)$, $E_R(t)$, $E_T(t)$ are the electric fields of the pass through, reflected and total respectively, R is the reflection coefficient and the factor of *3* accounts for the extra distance the reflected wave travels through the medium. The total electric field wave is then the sum of Equations (8.15a) and (8.15b). For the Polarcor glass, the index of refraction is 1.54, meaning the reflection coefficient is approximately 0.045, calculated from the Fresnel equations. Eq. (8.15c) has been plotted below in Figure 8.10 for a 10 kHz carrier, 100 Hz modulation frequency, and both modulation depths of $2\pi/3$ an $8\pi/3$. A reflection coefficient of 0.3 was used instead of 0.045 in order to emphasize the

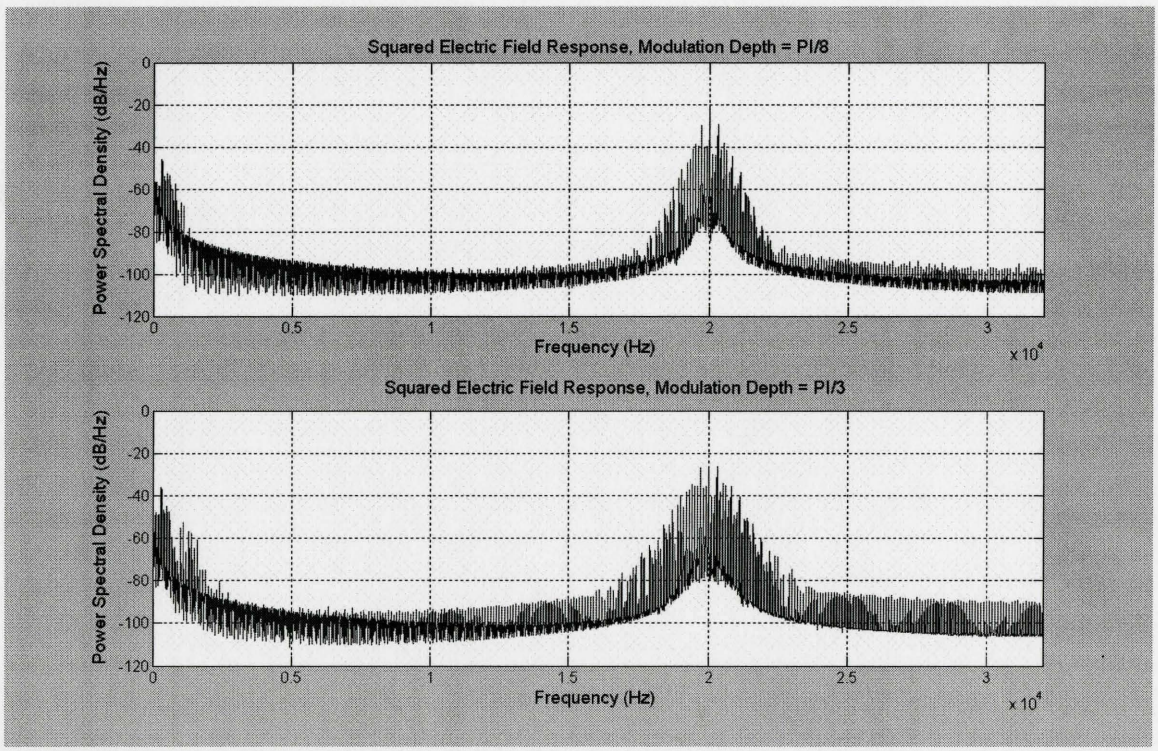


Figure 8.10 – Baseband generation of phase modulation harmonics using Fabry-Perot effect. Reflection coefficient of 0.3 is used to emphasize the contrast with Figure 8.9. Top) Squared electric field response for modulation depth of $\lambda/8$. Bottom) Modulation depth of $\lambda/3$.

contrast with Figure 8.9 that without the Fabry-Perot effect, the base band generation will not occur. Notice that in the figures the spectral width of the carrier is not perfectly monochromatic due to a finite discrete sequence length in MATLAB. As a result, this artificial broadening is also appearing in the base band and obscuring the harmonics we are looking for. With a reflection coefficient of 0.045, they will no longer be discernable in the plot. Fortunately, it is easy to estimate the magnitude. The amplitude of the reflected wave will be 0.045^2 or 2.025×10^{-3} which corresponds to a difference in the power spectral density plot of approximately -27 dB. If we refer back to Figure 8.1, we see the difference in power between the PL and the modulation harmonics is approximately 30 dB, in line with our expectations for a material with an index of refraction of 1.54.

8.4 The Total Spectrum

It is now probable that the rotating polarizer will generate measurable harmonics at integer multiples of its mechanical frequency. These shall be considered *primary harmonics*. Once again employing Eq. (3.4), when the chopper signal (Figure 8.4) is modulated by the transfer function of the polarizer (Figure 8.5) the fundamental 1070 Hz signal will create sum and difference signals with each of the primary harmonics ($115n$ were n is the n^{th} harmonic for 115 Hz rotation), the result of which shall be considered *secondary harmonics*. Furthermore, the chopper signal is neither a pure square wave nor sinusoid, thus additional components near the fundamental are also present in Figure 8.4 (e.g. 1070 ± 25Hz @ 45 dB). Each of these nearby chopper sidebands will also create

Table 8.1 – Harmonic generation for 115 Hz polarizer and 1070 Hz chopper

Harmonic Type	Harmonic Frequencies (Hz)										
Primary	115	230	345	460	575	690	805	920	1035	1150	1265
Secondary (Diff)	955	840	725	610	495	380	265	150	35	80	195
Secondary (Sum)	1185	1300	1415	1530	1645	1760	1875	1990	2105	2220	2335
Tertiary (1095 Hz)	980	865	750	635	520	405	290	175	60	55	170
Tertiary (1045 Hz)	930	815	700	585	470	355	240	125	10	105	220

Table 8.1 – The fundamental frequency for primary generation is 115 Hz. Fundamental frequency for sum and difference secondary generation is 1070 Hz. Difference-only tertiary generation is shown for 1070±25 Hz. The frequencies most likely to create noise in our DOP readings are the secondary harmonics at 195 and 265 Hz for phase sensitive detection at 230 Hz.

sum and difference components with each of the *primary harmonics*, creating *tertiary harmonics*. To be as accurate as possible, the multiples of the chopper frequency must also be considered in sum and difference generation however these will be neglected to keep the analysis manageable since the primary harmonics at these frequencies are of far less strength as can be seen in Figure 8.1. Table 8.1 above calculates the primary, secondary and one pair of difference tertiary harmonics for chopper and polarizer frequencies of 1070 Hz and 115 Hz respectively with chopper sidebands of 1070±25 Hz as seen in Figure 8.1.

Figure 8.11 below shows a zoomed view of the spectrum shown in Figure 8.1 in the vicinity of 230 Hz, the frequency at which we measure the polarizer response. The primary and secondary frequencies predicted in Table 8.1 are present.

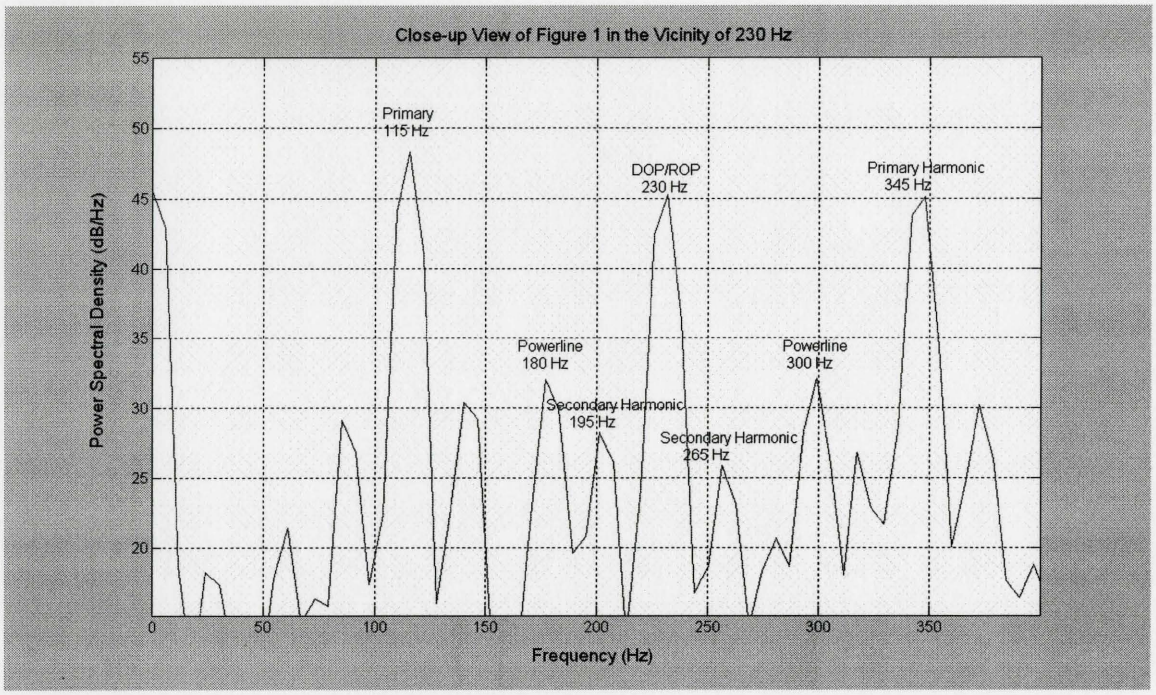


Figure 8.11 – Close-up view of Figure 8.1 in the vicinity of 230 Hz. The predicted primary and secondary harmonics are clearly visible.

8.5 Optimizing the Polarizer and Chopper Frequencies

The DOP maps can be optimized by selecting frequencies for the polarizer and chopper such that the secondary harmonics are spectrally as far as possible from the polarizer and chopper signal. The results of Table 8.1 can be empirically summarized by defining the *beat frequency* to be the difference between the chopper and the nearest primary harmonic. Beat harmonics (separated by the beat frequency) will appear on both sides of every primary harmonic since these are the sources of the beating. Therefore, we can minimize the impact by selecting frequencies such that the beat frequency is $1/3^{rd}$ the mechanical frequency of the polarizer. A beat frequency of anything greater or less than

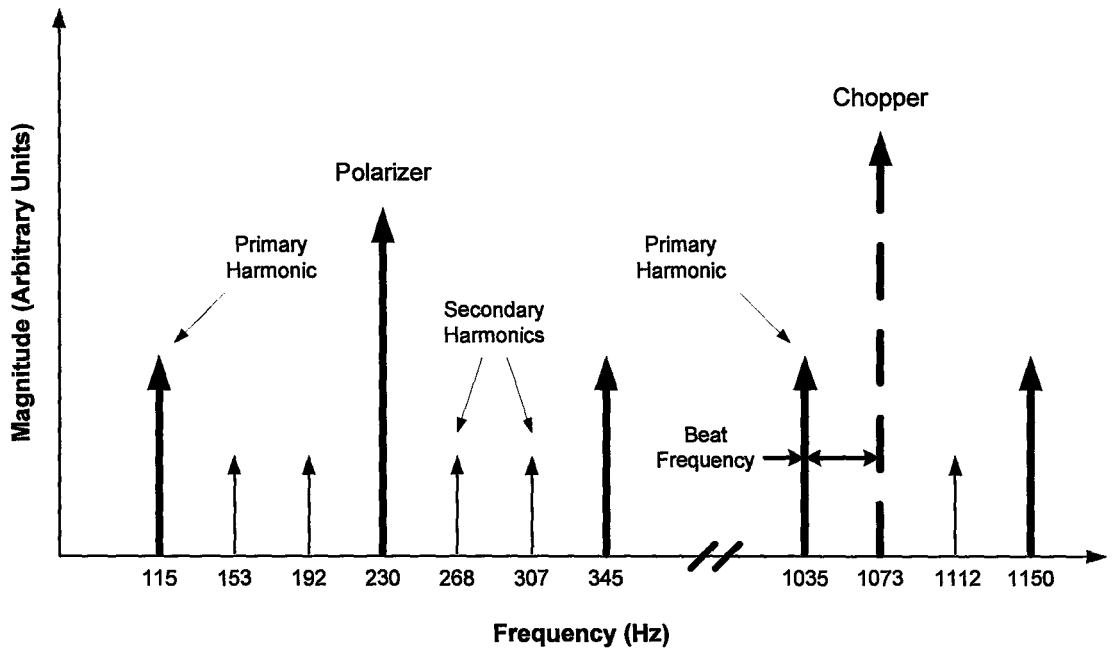


Figure 8.12 – Optimal configuration of the polarizer and chopper frequencies. The beat frequency is determined by the difference between the chopper and the nearest primary harmonic. In the case above, the nearest primary harmonic is 1035 Hz creating a beat frequency of 38 Hz. While increasing the beat frequency will move the secondary harmonics further from the polarizer, it will also move the beat frequency closer to the chopper.

1/3rd of the mechanical frequency will move the secondary harmonics closer to the chopper or polarizer signals of interest. By setting the polarizer rotation at 115 Hz and the chopper at 1070 Hz, the beat frequency is 35 Hz where as optimal beat frequency would be 115/3 = 38.3 Hz. This situation is illustrated in Figure 8.12 for the optimal configuration.

By examining Figure 8.12, we see that if we increase the beat frequency past the 1/3rd mechanical frequency point, the secondary harmonic at 1112 Hz will be getting closer to the chopper signal as it gets further from 1150 Hz. Further enhancement may possibly be gained by finding the location at which the choppers nearest primary

harmonic and secondary harmonic are not at equal distances, but at equal effective magnitudes when the frequency response of the post-multiply lowpass filtering is accounted for. Finally, the power line harmonics must also be considered. In Figures 8.2 and 8.3, the even harmonics of the power line were not significant, however the odd harmonics were. As a result, I have chosen to set the polarizer at 230 Hz and the chopper at 1070 Hz in order to keep the polarizer at a reasonable distance from the power line harmonics as well. It should be mentioned that the step resolution of the chopper frequency is 10 Hz. To illustrate the impact of poor frequency configuration and thus the secondary harmonics on the DOP maps, two DOP maps are presented below.

The left side of Figure 8.13 shows a scan where the polarizer and chopper are set to 230 Hz and 1040 Hz respectively. Using our empirical analysis, this creates a beat frequency of 5 Hz. The secondary harmonic will be well within the passband of the

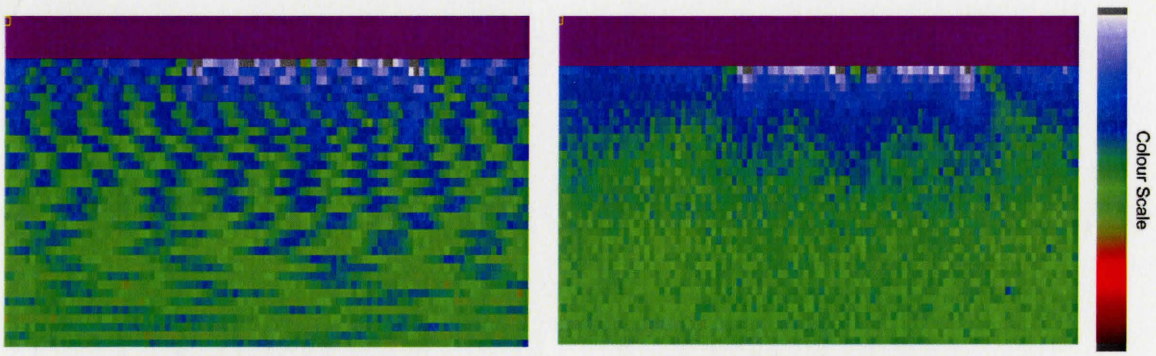


Figure 8.13 – Effect of secondary harmonics in a poor frequency configuration. Left) The PL is focused near (but not at) the centre of the 230 Hz rotating polarizer, but the chopper is configured for 1040 Hz, resulting in a beat frequency of 5 Hz. Right) The chopper is configured for 1070 Hz, but the PL is focused at approximately 2/3rd of the radius from the centre of rotation.

output stage lowpass filtering, and as a result the periodic noise is clearly evident in the scan and the map is of poor quality. The right side of Figure 8.13 has returned the polarizer and chopper to 230 Hz and 1070 Hz respectively, but the polarizer motor has been translated such that the beam is now focused far from the centre of rotation, thus increasing the magnitude of the primary and secondary harmonics. We see that while the noise has increased slightly (typical RMS noise in Chapter 6 was found to be between 0.050% and 0.065%), the map is still fairly accurate however the DOP now has a large offset in the unstrained region. The strength of the secondary harmonics is very strong, however the 35 Hz beat frequency has ensured the interfering signal is not in the passband of the output stage. Moreover, the polarizer demodulation frequency of 230 Hz is itself a primary harmonic, and this harmonic will appear as an offset in the DOP reading. In this case, an offset of over 3% is measured. For contrast, the maximum variation in DOP between the assumed unstrained region (green) and the highly strained regions around the center rib is approximately 1.5%.

With respect to the mechanically induced optical phase modulation present in the DOP experiment, it is important to choose good frequencies at which to operate the polarizer and chopper, as well as ensure that the optical beam is focused on or close to the centre of the rotating polarizer glass. Failure to perform the former will have significant impact on the amount of noise present in DOP maps. Failure to perform the latter will have a much smaller impact on noise, but will heavily influence the average DOP seen from the sample. The basis for locating the centre of rotation on the polarizer is simple; when the HeNe laser is focused on a spatial location with an expected zero degree of

polarization, the reading from the lock-in amplifier should also be zero. A suggested procedure to minimize the effects of the frequency modulation is to:

i) Calibrate the zero level reading by measuring PL/DOP/ROP with the beam blocked. This will remove any offsets introduced by the electronics.

ii) Unblock the beam, and position the sample under study so the HeNe is focused on a location far from any strain inducing effects, where the DOP is expected to be zero but the PL is nonzero. If the DOP/ROP reading is not zero, continue on to step iii).

iii) Translate the polarizer motor along the horizontal x-axis to find the opposite edges of the glass. The PL will drop dramatically at the edge. Estimate and set the motor at the location of the middle.

iv) Repeat step iii) for the vertical z-axis.

v) Fine tune the position of the motor in both x and z directions until the DOP and ROP readings are as close to zero as possible.

Chapter 9. Conclusions

This thesis has reported the results of a thorough investigation of the DOP experiment. Previously unexplained behavior with regards to noise and error as the optical chopper frequency varied has been spectrally examined and found to be the result of sideband generation of beat frequencies between the optical chopper and polarizer. The origin of these frequencies is hypothesized to be caused by mechanical frequency modulation of the terahertz light wave by the rotating polarizer. A physical model explaining the phenomenon was presented. Recommendations and procedures for minimizing the impact of the interfering harmonics have been established. A reduction in noise level of roughly 50% for typical DOP scan times has been achieved on both the original analog and new digital systems by minimizing harmonics.

The research was carried out by replacing the analog phase sensitive detection circuit with a modular digital system. The design of the system was presented along with several novel techniques for synchronizing phase sensitive detectors and expanding their capabilities. The platform style of system construction will provide researchers with a reconfigurable architecture that can be modified for their own research with far less expertise and time required versus a full-custom solution. The platform also provides a feature rich alternative to purchasing expensive commercial equipment.

This thesis covered a broad range of topics including DOP theory, quantization analysis, recommended procedures and provides a design reference for graduate students new to digital system design. Hopefully this contribution will improve both the engineering knowledge and skills of future DOP researchers.

Appendix A. Two's Complement Binary Number System

A firm understanding of two's complement binary numbers is absolutely essential to working with digital signal processing systems. Binary numbers are often referred to as unsigned or signed where unsigned means the numbers are non-negative. Unsigned numbers follow the typical numerical rules for a base-2 counting system. Below are some examples of 8-bit binary numbers and their unsigned base-16 and base-10 equivalents. Note that it is conventional to precede a hex number with '0x' to indicate hexadecimal. The subscript in parentheses indicates the number base.

$$0000\ 0000_{(2)} \rightarrow 0x00_{(16)} \rightarrow 0_{(10)} \quad \text{(A.1)}$$

$$0111\ 1111_{(2)} \rightarrow 0x7F_{(16)} \rightarrow 127_{(10)} \quad \text{(A.2)}$$

$$1000\ 0000_{(2)} \rightarrow 0x80_{(16)} \rightarrow 128_{(10)} \quad \text{(A.3)}$$

$$1111\ 1111_{(2)} \rightarrow 0xFF_{(16)} \rightarrow 255_{(10)} \quad \text{(A.4)}$$

When a binary number system includes negative numbers, 2's complement notation is used. In 2's complement notation, the most significant bit (MSB) of the binary number is the sign bit. A sign bit of '0' means the number is positive, and will follow the conventional number rules. A sign-bit of '1' indicates the number is negative, however the remaining bits do not follow the same convention as the positive numbers. The reason for this is if negative numbers had the same representation as positive numbers, but with the MSB as '1', they would not follow the rules for binary arithmetic, i.e. adding two numbers would give you the wrong answer. 2's complement ensures all arithmetic

rules will apply and give you the correct answer regardless of whether the numbers are actually signed or unsigned.

The following equations describe how to convert a decimal number to a 2's complement n-bit number.

$$2s\_comp\_pos = dec\_to\_bin(positive\_number, n), \tag{A.5a}$$

$$2s\_comp\_neg = dec\_to\_bin((negative\_number + 2^n), n), \tag{A.5b}$$

where *2s_comp_pos* and *2s_comp_neg* are the 2's complement positive and negative numbers respectively, and *dec_to_bin* is a function that takes in a positive number and the number of bits to convert it to.

The following equations describe how to convert a 2's complement binary number to a decimal number.

$$dec\_msb\_zero = bin\_to\_dec(2s\_comp\_ps), \tag{A.6a}$$

$$dec\_msb\_one = bin\_to\_dec(2s\_comp\_ps) - 2^n, \tag{A.6b}$$

where *dec_msb_zero* is to be used if the MSB is '0' and *dec_msb_one* is to be used if the MSB is '1'.

Alternatively, a common method for finding the 2's complement of a binary number is to copy (or keep the same) all bits starting from the LSB until the first '1' is reached. Then all remaining bits after that '1' to the MSB should be inverted. An example of this transformation is shown in Eq. (A.7). This transformation illustrates the 2's complement representation for $-110_{(10)}$.

$$0110\ 1110_{(2)} \rightarrow 1001\ 0010_{(2)} \tag{A.7}$$

Some examples are shown below for an 8-bit 2's complement system.

$$0_{(10)} \rightarrow 0000\ 0000_{(2)} \rightarrow 0x00_{(16)} \quad \text{(A.8a)}$$

$$127_{(10)} \rightarrow 0111\ 1111_{(2)} \rightarrow 0x7F_{(16)} \quad \text{(A.8b)}$$

$$-1_{(10)} \rightarrow 1111\ 1111_{(2)} \rightarrow 0xFF_{(16)} \quad \text{(A.7c)}$$

$$-128_{(10)} \rightarrow 1000\ 0000_{(2)} \rightarrow 0x80_{(16)} \quad \text{(A.7d)}$$

The *range* of an *n*-bit 2's complement number is defined as

$$-2^{(n-1)} \leq range \leq (2^{(n-1)})-1 \quad \text{(A.8)}$$

Appendix B. Glossary of Terms

ADC – analog to digital converter.

CED – command encoder/decoder. Used to process and create communication protocol packets.

CSR – configuration status register. Stores device configuration and status for the system.

DAC – digital to analog converter.

DLIA – digital lock-in amplifier.

DOP – degree of polarization. Refers generally to the experiment or to polarization induced by direct strain.

DSP – digital signal processing. Analog data has been quantized and processed discretely.

FIR – finite impulse response. A type of discrete filter.

FPGA –field programmable gate array. A reconfigurable digital logic device.

FSM – finite state machine. A digital circuit that implements a state-transition diagram.

PL – photoluminescence. The total light emitted from the sample that is excited by a pump laser.

ROP – rotated degree of polarization. Refers to polarization induced by shear strain.

RAM – random access memory.

ROM – read-only memory.

SRAM – synchronous random access memory.

Appendix C. References

[1] G.E. Pikus and G.I. Bir. *Symmetry and Strain-Induced Effects in Semiconductors*. New York: Wiley, 1974.

[2] G.H. Olsen, C.J. Nuese, and R.T. Smith, "The effect of elastic strain on energy band gap and lattice parameter in III-V compounds," *J. Appl. Phys.*, vol. **49**, p. 5523 (1978)

[3] H. Asai and K. Oe, "Energy band-gap shift with elastic strain in GaInP epitaxial layers on (001) GaAs substrates," *J. Appl. Phys.*, vol. **54**, p. 2052 (1983)

[4] H. Shimizu, K. Itoh, M. Wada, T. Sugino, and I. Teramoto, "Improvement in operation lives of GaAlAs visible lasers by introducing GaAlAs buffer layers," *IEEE. J. Quantum Electron.*, vol. **17**, pp. 763-767 (1981)

[5] J. Groenen, G. Landa, R. Carles, P. S. Pizani, and M. Gendry, "Tensile and compressive strain relief in $In_xGa_{1-x}As$ epilayers grown on InP probed by Raman scattering," *J. Appl. Phys.*, vol. **82**, pp. 803-809 (1997)

[6] S. Zhang, M. G. Boudreau, R. Kuchibhatla, Y. Tao, S. R. Das, E. M. Griswold, and U. Sharma, "Influence of the electrical contact on the reliability of InP-based ridge waveguide distributed feedback semiconductor diode lasers for telecommunication applications," *J. Vac. Sci. Technol.*, vol. **22**, pp. 803-806 (2004)

[7] D. T. Cassidy, "Spatially-resolved and polarization-resolved photoluminescence for study of dislocations and strain in III-V materials," *Mat. Sci. & Eng.*, vol. **B91-92**, pp. 2-9 (2002)

[8] D. Lisak, D. T. Cassidy, and A. H. Moore, "Bonding stress and reliability of high power GaAs-based lasers," *IEEE Trans. Pack. Technol.*, vol. **24**, pp. 92-98 (2001)

[9] J.P. Landesman, "Micro-photoluminescence for the visualization of defects, stress and temperature profiles in high-power III-V's devices," *Mater. Sci. Eng.*, vol. **B91-92**, pp. 66-61 (2002)

[10] A. Barwolff, J.W. Tomms, R. Muller, S. Weiss, M. Hutter, H. Oppermann, and H. Reichl, "Spectroscopic measurement of mounting-induced strain in optoelectronic devices," *IEEE Trans. Adv. Packag.*, vol. **23**, pp. 170-175 (2000)

[11] K. Kobayashi, Y. Inoue, T. Nishimura, M. Hirayama, Y. Akasaka, and T. Kato, "Local-oxidation-induced stress measured by Raman microprobe spectroscopy," *J. Electrochem. Soc.*, vol. **137**, pp. 1987-1989 (1990)

[12] P. Van der Sluis, "Determination of strain in epitaxial semiconductor structures by high resolution x-ray diffraction," *Appl. Phys. A*, vol. **58**, pp. 129-134 (1994)

[13] B.G. Yacobi, B. Elman, C. Jagannath, A.N.M. Masum Choudhury, and M. Urban, "Cathodoluminescence observation of metallization-induced stress variations in GaAs/AlGaAs multiple quantum well structures," *Appl. Phys. Lett.*, vol. **52**, pp. 1806-1808 (1988)

[14] K. Rammohan, D.H. Rich, R. S. Goldman, J. Chen, H.H. Wieder, and K.L. Kavanagh, "Study of micrometer-scale special variations in strain of a compositionally step-graded InGaAs/GaAs (001) heterostructure," *Appl. Phys. Lett.*, vol. **66**, pp. 869-871 (1995)

[15] A. Jakubowicz, "Revealing process-induced strain fields in GaAs/AlGaAs lasers via electron irradiation in a scanning electron-microscope," *J. Appl. Phys.*, vol. **70**, pp. 1800-1804 (1991)

[16] P.D. Colbourne, D.T. Cassidy, "Imaging of stresses in GaAs diode lasers using polarization-resolved photoluminescence", *IEEE J. Quantum Elecrton.*, vol. **29**, pp. 62-68 (1993)

[17] D. T. Cassidy and C. S. Adams, "Polarization of the Output of InGaAsP Semiconductor Diode Lasers," *IEEE J. Quantum Electron.*, vol. **25**, pp. 1156-1160 (1989)

[18] B. Lakshmi, D. T. Cassidy, and B. J. Robinson, "Quantum-well strain and thickness characterization by degree of polarization," *J. Appl. Phys.*, vol. **79**, pp. 7640-7645 (1996)

[19] J.F. Nye, *Physical Properties of Crystals.* Clarendon: Oxford, 1985

[20] D. T. Cassidy, S.K.K. Lam, B. Lakshmi, and D. M. Bruce, "Strain mapping by measurement of the degree of polarization of photoluminescence", *Appl. Optics*, vol. **43**, pp. 1811-1818 (2004)

[21] J. Yiang and D.T. Cassidy, "Strain measurement and estimation of photoelastic effects and strain-induced optical gain change in ridge waveguide lasers", *J. Appl. Physics*, vol. **77**, pp. 3382-3387 (1995)

[22] B. G. Streetman, *Solid State ElectronicDevices, 4th Ed.*, New Jersey: Prentice Hall, 1995, p. 219.

[23] J.O. Smith, "Quality Factor," in *Introduction to Digital Filters with Audio Applications*, http://ccrma.stanford.edu/~jos/filters/Quality_Factor_Q.html, online book, accessed June '07.

[24] R. C. Dorf, J. A. Svoboda, "Introduction to Electric Circuits, 5th Edition." John Wiley & Sons, 2001. Pg. 593.

[25] S. Haykin and B. V. Veen, "Fourier Transform Representations for Periodic Signals," in *Signals and Systems*, John Wiley & Sons: New York, 1999.

[26] Daniel H. Sheingold (1986), "Analog-Digital Conversion Handbook," Analog Devices Inc., pp. 560-568.

[27] Micheal J. Demler, "High Speed Analog-Digital Conversion," San Diego: Academic Press Inc., 1991. pp. 111-126.

[28] Micheal J. Demler, "High Speed Analog-Digital Conversion," San Diego: Academic Press Inc., 1991. pp. 8-10.

[29] C. E. Shannon, "Communication in the presence of noise," *Proc. Inst. Radio Eng.*, vol. **37**, pp. 10-21 (1949)

[30] Micheal J. Demler, "High Speed Analog-Digital Conversion," San Diego: Academic Press Inc., 1991. pp. 143-145.

[31] Analog Devices AD977/977A 16-bit 100kps/200kps BiCMOS A/D Converter Datasheet

[32] E. D. Morris Jr. and H. S. Johnston, "Digital phase sensitive detectors", *Rev. Sci. Instrum.*, vol. **39**, p. 620 (1968)

[33] L. G. Rubin, "The new(?) wave: digital lock-in amplifiers", *Rev. Sci. Instrum.*, vol. **54**, pp. 514-515 (1988)

[34] P. K. Dixon and L. Wu, "Broadband digital lock-in techniques", *Rev. Sci. Instrum.*, vol. **60**, pp. 3329-3336 (1989)

[35] R. W. M. Smith, I. L. Freeston, B. H. Brown and A. M. Sinton, "Design of a phase sensitive detector to maximize the signal-to-noise ratio in the presence of Gaussian wideband noise", *Meas. Sci. Technol.*, vol. **3**, pp. 1054-1062 (1992)

[36] E. Iacopini, B. Smith, G. Stefanini and S. Carusotto, "Digital techniques applied to phase-sensitive detection", *J. Phys. E, Sci. Instrum.*, vol. **16**, pp. 844-847 (1983)

[37] L. A. Barragan, J. I. Artigas, R. Alonso and F. Villuendas, "A modular, low-cost, digital signal processor-based lock-card for measuring optical attenuation", *Rev. Sci. Instrum.*, vol. **72**, pp. 247-251 (2001)

[38] A. A. Dorrington and R. Kunnemeyer, "A simple microcontroller based digital lock-in amplifier for the diction of low-level optical signals", in *Proc. IEEE International Workshop on Design, Test and Appl.*, (2002)

[39] A. Restelli, R. Abbiati and A. Geraci, "Digital field programmable gate array-based lock-in amplifier for high performance photon counting applications", *Rev. Sci. Instrum.*, vol. **76**, pp. 093112-1 to 093112-8 (2005)

[40] M. O. Sonnaillon and F. J. Bonetto, "A low-cost, high-performance, digital signal processor-based lock-in amplifier capable of measuring multiple frequency sweeps simultaneously", *Rev. Sci. Instrum.*, vol. **76**, pp. 024703-1 to 024703-7 (2005)

[41] Electronics Industries Association, "*EIA Standard RS-232-C Interface Between Data Terminal Equipment and Data Communication Equipment Employing Serial Data Interchange*", August 1969, reprinted in Telebyte Technology *Data Communication Library*, Greenlawn NY, 1985, no ISBN

[42] http://www.opencores.org/cvsweb.shtml/MiniUART/doc/

[43] http://www.gnu.org/copyleft/gpl.html

[44] Altera Corp. Application Note, Altera Devices, AN42

[45] Xilinx Virtex-4 User's Guide, "*Block RAM*", p. 109 (March 21st, 2006 Revision).

[46] R. C. Dorf and J. A. Svoboda, "Introduction to Electric Circuits, 5th Edition." John Wiley & Sons, 2001, p. 593.

[47] R. C. Dorf, J. A. Svoboda, "Introduction to Electric Circuits, 5th Edition." John Wiley & Sons, 2001, pp. 742-745

[48] A. V. Oppenheim, R. W. Schafer, "Discrete-Time Signal Processing." Prentice Hall, New Jersey, 1999, p. 468.

[49] R. C. Gonzalez, R. E. Woods, "Digital Image Processing." Prentice Hall, New Jersey, 2002.

[50] A. Bloch, "Modulation theory," *J. I.E.E.* (London), 91, pgs. 31-42 (1944)

[51] L. J. Giacoletto, "Generalized Theory of Multitone Amplitude and Frequency Modulation," *Proc. I.R.E*, July pgs. 680-693 (1944)