UNIVERSITY OF CALIFORNIA SAN DIEGO

Maximum Likelihood Sequence Estimation for a Burst Mode Avalanche Photodiode Receiver

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Electrical Engineering (Photonics)

by

Tyler Jackson

Committee in charge:

Professor George Papen, Chair
Professor Zhaowei Liu
Professor Shayan Mookherjea

2024

The Thesis of Tyler Jackson is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The data processed in this thesis was collected by Sandia National Labs from an apparatus of their design and construction.

Sincerest acknowledgement to Professor George Papen for his support as the chair of my committee. His support was essential to drafting this thesis.

VITA

2016 Bachelor of Science in Physics with a minor in Mathematics, University of California Los Angeles

2017 Associate Systems Engineer, LinQuest Corporation. MILSATCOM

2019 Lead Systems Engineer, Booz Allen Hamilton. NIWC PAC

2020 Principal Systems Engineer, LinQuest Corporation. MILSATCOM

2023 Program Manager, LinQuest Corporation.

2024 Master of Science in Electrical Engineering (Photonics), University of California San Diego

FIELD OF STUDY

Major Field: Electrical Engineering
Studies in Photonics
Professor George Papen

ABSTRACT OF THE THESIS

Maximum Likelihood Sequence Estimation for a
Burst Mode Avalanche Photodiode Receiver

by

Tyler Jackson

Master of Science in Electrical Engineering (Photonics)

University of California San Diego, 2024

Professor George Papen, Chair

This thesis analyzes methods to improve the performance of avalanche photodiodes (APD) in optically switched networks. The optical links in these networks must re-initialize the optical signal after a switching event, which is called burst mode operation. The change in the lightwave signal along with the biasing characteristics of the APD can produce artifacts in the signal that require mitigation. This thesis explores the use of Maximum Likelihood Sequence Estimation to overcome artifacts such as intersymbol interference (ISI).

INTRODUCTION

The backbone of the information age is the datacenters and their enabling communication systems that run the internet. A vast number of resources are dedicated to designing, building, and maintaining these systems. Due to their importance and energy consumption, it behooves designers to create energy-efficient data centers. Within the backbone network optical point-to-point links comprise the long-haul trunk lines of the internet. The switches that enable the internet are still electronic in nature. This means that the internet is a network of electric switching nodes connected by optical links. Because of multiple internet services vendors, it is difficult to make substantial changes to this infrastructure.

In contrast, within a datacenter owned by a single entity, there is an opportunity to replace the electronic switches with optical switches to create an all-optical network. The elimination of the electrical to optical interfaces reduces the number of components in the system, which may reduce the overall power consumption. The risk is that the optical switches introduce loss to the system without compensating with gain. Adding optical amplifiers at the switches reintroduce component count and cost back into the system. So, optical amplifiers may not be a cost-effective solution to overcome the switch loss. In addition, optical switches also break connections so that the optical transceivers must re-initialize after a switching event. In a standard point-to-point, the link is always “on”. Even when no data is transmitted, idles frames are sent to keep the link up. When an optical switch is used, the receiver needs to rapidly re-acquire or lock onto the optical signal after a switching event.

To capitalize the opportunity of optical switching while burning down risk, receivers with greater sensitivity can be implemented utilizing Avalanche Photodiodes (APD). These devices use a high reverse bias to generate an internal gain mechanism that traditional photodiodes do

not have. However, the increased sensitivity provided by the internal gain process has its own complications that make the use of APD for optical switched networks difficult to implement. The internal gain process has it owns form of noise that is seen when the signal is re-acquired after a switching event. APDs introduce signal dependent noise into a link, which can reduce link margin when the link margin is already low because of the optical switch loss. So, the key system attributes of the optical network are realized partly through the implementation of an APD:

- Energy Efficiency
- Extensibility

The main idea is an arbitrarily large information system can be constructed from an arbitrarily large number of components while each component consumes as little power as possible. The structural architecture that realizes those attributes are optical switches, broadband transmitter arrays, receiver arrays, and the pertinent APD. The key performance parameters of the enterprise are:

- Scalability exceeding 250 Pbps
- Energy use less than 2 pJ/bit

To achieve these goals, the system of interest is a receiver for a lightwave guided-mode communication system. The receiver is the waveguide integrated Ge APD designed and fabricated at Sandia National Laboratory. The primary goal of this thesis is to develop advanced detection methods to process the data collected by the prototype APD devices. Successful data processing verifies the functionality of the APD in its contextual domain. All aspects of the physical experiment were performed at Sandia national laboratory by their staff. This thesis only addresses the data processing of data recorded by Sandia.

Chapter 2 BACKGROUND

A communication system transfers information from a sender to a receiver. An input or signal, *s*, (at the sender) is transformed through a channel to an output, *r*, at the receiver. The goal of the receiver is to determine the original input information stream. In this way, information is transmitted through the communication system. One dominant random process called noise impairs the determination of the original data stream. The noise, *n*, is parameterized by one or more random variables. These random variables may or may not be signal dependent. A linear channel is a common model that is deterministic, where the transformation from the input to the output signal in the absence of noise is linear. In general, the linear transformation is characterized by a function *h(t,τ).* A linear change to the input results in a linear change in the output

$$r(t) = \int_{-\infty}^{\infty} h(t,\tau)s(\tau)\,\mathrm{d}\tau. \tag{1}$$

A further assumption is that the channel is time invariant. The time in which the input is made does not affect *h(t)* and that $s(t + \Delta t)$ implies $r(t + \Delta t)$

$$r(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau)\,\mathrm{d}\tau. \tag{2}$$

This formalism enables the powerful linear systems theory to model the communication system using the concepts of the impulse response *h(t)* and the corresponding transfer function *H(ω)*, which is the Fourier transform of *h(t).*

There are many techniques to process data for linear time invariant (LTI) systems. Non-LTI systems are more difficult to process. Maximum Likelihood Sequence estimation (MLSE) is one such technique to process received data through a time-varying channel and will be the primary technique used in this thesis.

Since communication systems transmit information through propagating waves, the analysis is open to the richness of wave physics. While it is possible to communicate over any type of wave such as sound, the electromagnetic spectrum is the primary communications medium because it can support much higher data rates because of the increased carrier frequency and propagate much longer distances. There are many ways to subdivide the electromagnetic spectrum, but the taxonomy useful here is radiofrequency (RF) and optical. RF systems operate below 100 GHz while optical systems operate above 100 THz. The primary advantage of lightwave systems is the available bandwidth. In just the infrared conventional band (1530 nm – 1565 nm), there is 4.38 THz of bandwidth which is inconceivable to radio frequency technologies where the carrier frequency may only be centered at 1 GHz.

INFORMATION CAPACITY OF A COMMUNICATION SYSTEM

The performance requirement on the system is the capacity or how much information can be transported in how much time. For any communication system that is impaired by additive noise, the ultimate channel capacity, *C*, is bounded by Shannon's equation

$$C \leq BW \log_2\left(1 + \frac{S}{N}\right). \tag{3}$$

The channel capacity can always be lower than the limit, but never greater. As presented, Equation (3) is for gaussian noise, but can be modified to account for different and multiple random noise processes. It is difficult for a system to operate near the theoretical Shannon limit, but systems are engineered to get as close as possible under the constraints particular to that system. In general, there are three parameters that can be engineered to improve the performance of the system. This model is valid for noise that is additive.

Bandwidth, *BW*, is the amount of the electromagnetic spectrum used by the channel and is measured in frequency. All channels are bandlimited. Bandwidth is sometimes fixed for the system by licensing government agencies and is the primary constraint such as in wireless radio frequency transmission. For guided mode transmission, the bandwidth of the device is more limited than the channel such as in optical transmission over a fiber. For either system, spectral efficiency is an important consideration in modulation schemes and device design.

The symbol *S* in Equation (3) is not the signal, but the signal power measured in watts. It is the amount of energy per unit time used to transmit the signal over the bandwidth *BW*. The power available to the sender is intuitively constrained. The signal typically goes through an amplifier that generates gain. Wireless transmission can also achieve gain through the directivity of an antenna. The channel then attenuates the signal power. The receiver may also have an amplifier to further aid signal power. Gain and attenuation are measured in decibels (dB) and is defined as $10\ \log_{10}\left({Pout}/{Pin}\right)$ where P_{out} is the power at the channel output and P_{in} is the power at the channel input. The attenuation of a free space mode is called the space loss factor and goes $\sim {1}/{r^2}$, where *r* for this expression is defined as the distance or range. These expressions show that increasing the signal power can increase the channel capacity or increase the range. Amplifiers can be placed in the channel for long haul links to increase range and are

called repeaters. The attenuation of guided mode optical propagation through a fiber at a wavelength that produces the minimum amount of loss typically goes at $\sim -0.2\,{}^{\mathrm{dB}}/_{Km}$. RF waveguides carrying waves in the GHz's only have ranges in the order of meters before the signal is attenuated beyond use. The ability of optical systems to carry frequencies in the 100s of THz many kilometers make fiber optical communication the only practical choice for long-haul guided mode communication.

The symbol *N* in Equation (3) is the noise power. The nature of the noise depends on the system. However, the noise power is subject to the same gain and attenuation mechanisms as the signal. The insidious nature of this point is that while the signal power can be made to be arbitrarily high by chaining an arbitrary number of repeaters, the noise power will be amplified disproportionally to the signal and can eclipse the signal power eventually. This disproportionate gain is quantized through a parameter called the noise figure, NF, of the amplifier, which is a measure of the output SNR (*S/N*) to the input SNR. Every stage of amplification decreases the SNR as quantified by the noise figure. The sender typically uses a high gain amplifier called a boost amplifier, which tends to have a high NF. The receiver typically uses a low NF at the sacrifice of gain called a preamplifier. The noise power is after the channel is at a maximum. These heuristics guide design choices and are not universal. Analysis of a communication system may treat ${}^{S}/_{N}$, as a single parameter called the signal-to-noise ratio (SNR).

The functional requirement on the system is whether it can communicate reliably. The standard metric of the system that transmits bits to reliably communicate is the probability of error of the transmitted bits. The bit error rate (BER) is determined by the statistics of the noise and the manner in which the information was modulated and encoded onto the channel. Therefore, the calculation of the BER is contextual term based on the system.

DETECTION THEORY

A digital communication system encodes the information is a decision between discreet symbols. The measurement of the received signal *r(t)* is used to build a likelihood on *s(t)*. The conditional probability of $p(s|r)$ cannot be known as the receiver does not have a priori knowledge of the sender. The likelihood function is calculated instead

$$\mathcal{L} = p(r|s). \tag{4}$$

A threshold is used to make a binary decision on the most likely transmitted bit. A hypothesis test is performed on the measurement and a discreet decision is made. For a binary communication system with each symbol taking on one of values (0 and 1), the binary hypothesis test to determine if s_1 or s_2 was transmitted is given by.

$$p(r|s_1) > p(r|s_2) \text{ or}$$
$$p(r|s_1) < p(r|s_2). \tag{5}$$

So functionally, a communication system logically conveys information encoded into digital bits (0 or 1) from one location to another.

Digital bits are the fundamental unit of information and are the basis of information and computer systems. The encoded bits are modulated onto a carrier wave for physical transport. The simplest digital communication technique is On Off Shift Keying (OOK) and is common in optical systems. Higher order modulation formats encode more than one bit of information per

transmitted symbol. We do not consider these modulation formats in this thesis.

In many communication systems, the noise statistics are additive and can be accurately modeled using an additive white gaussian noise (AWGN) noise model. The AWGN is a robust assumption due to the central limit theorem. Non-AWGN must be specifically stated in an analysis and characterized. The hypothesis test calculation for an AWGN channel is symmetric and a decision threshold can easily be made. A symmetric channel means that the probability of an error is independent of the transmitted value. In contrast, the probability of an error can depend on the transmitted value for an asymmetric channel. The likelihood functions in Equation (5) for AWGN are

$$f(r|0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-s_0)^2}{2\sigma^2}}$$

$$f(r|1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-s_1)^2}{2\sigma^2}} \tag{6}$$

s_0 and s_1 are the signals corresponding to the two transmitted symbols of 0 and 1.

The exponential form of the probability density functions (pdf) makes it convenient to use the log-likelihood function, $\mathcal{A}$, instead of the linear form $\mathcal{L}$. In fact, the signal and noise powers are expressed in log decibels due to the convenience logarithms provide in calculating gain and attenuation.

The simplest threshold Θ is for a symmetric binary channel. It is halfway between the symbols.

$$\Theta = \frac{s_1 - s_0}{2} \tag{7}$$

The probability p_e of making the incorrect decision or error is symmetrical in this case $p_{1|0} = p_{0|1}$ and $s_1 = -s_0$.

$$p_e = \frac{1}{2}\operatorname{erfc}\sqrt{\frac{s^2}{2\,\sigma^2}} \tag{8}$$

Where erfc() is the complementary error function, which is simply the integral of the gaussian function over a specified domain. The minimum distance d_{min} between the symbols is then defined as $d_{min} = 2\,s$. [1]

SEQUENCE DETECTION

Instead of constructing a hypothesis test using a single symbol, one can also consider testing for the most likely *sequence* of symbols transmitted. This block of bits is treated as a single entity and is discriminated against other blocks of bits. The length of the sequence is called the constraint length *v*. Equation 4 is slightly modified so that the compared likelihoods are the likelihoods of blocks of bits instead of individual bits. However, the method in which the maximum likelihoods of the sequences are evaluated is more complicated than the bit-by-bit demodulation and the number of possible sequences grows exponentially as a function of the length of the sequence.

The appropriate length of the sequence to use depends on the amount of memory in the system. This memory is characterized by the constraint length, which is the number of symbols intervals at the receiver that contain a significant fraction of the energy from a single symbol launched at the transmitter. This technique is called Maximum Likelihood Sequence Estimation (MLSE) and is the principal method of detection used in this thesis.

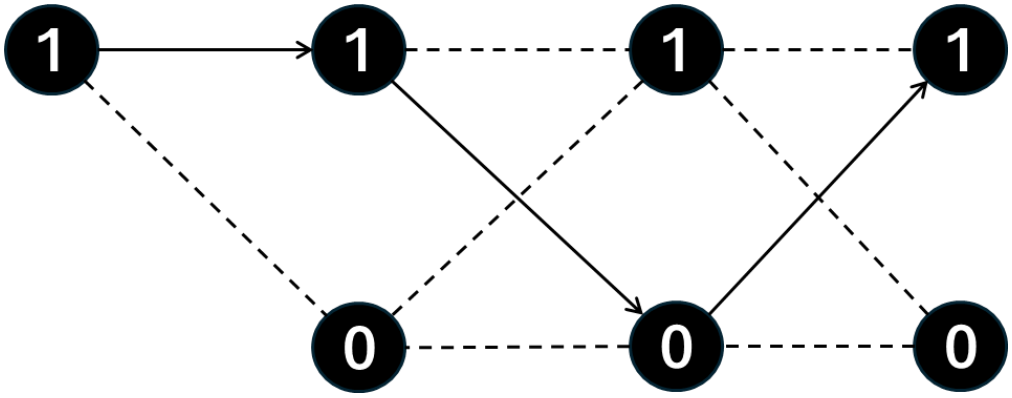

Figure 1: Trellis illustration with a 101 critical path.

An efficient algorithm for sequence estimation is based on mapping the sequences onto a trellis and comparing the likelihoods of sequences on that trellis. [2] The paths through the trellis represent the possible sequences that could be transmitted. Figure 1 is an example trellis where the decoded sequence is '101' with an initial state of '1'. This path is shown using solid lines. At each step of the trellis, likelihoods for possible sequences represented as paths through the trellis are calculated. The pathway with the maximum likelihood is chosen. The concept of BER is then extended to sequences estimation through the Sequence Error Rate (SER). So, for a single sequence error, there can be multiple bit errors within the sequence. However, sequence errors can diverge as a single error can force a wrong decision along the trellis when the subsequent bit is evaluated.

AVALANCHE PHOTODIOES

As stated in the introduction, all-optical switching enables the transmitted information within a datacenter to remain in the optical domain. This eliminates optical to electrical

converters from the system. The downside is the optical switch attenuates the signal. The use of an APD can compensate for this loss in an energy-efficient manner. Therefore, one of the strategies to achieve energy efficiency is to eschew the typical boost amplifier at the sender and to increase the sensitivity at the receiver using an APD. The risk is that all the gain of the system is now at the point where the noise power is at a maximum. The optical switch compounds the problem. The typical optical receiver is the photodiode with P/intrinsic layer/N (PIN) layers. To achieve gain, an APD is used in lieu of the PIN receiver. The signal power, *S*, as defined in the background is constrained by the requirements.

There remain two parameters from the capacity equation to engineer: *BW* and *N*. The APDs designed at Sandia were engineered to operate in the O-band but can notionally operate in the C-band. Now, there are no single devices that can utilize such bandwidth. So, the spectrum is subdivided into tractable channels. The data rates for the APDs used in this thesis are 10 Gb/s and 25 Gb/s. The data at 10 Gb/s was successfully processed before the work of this thesis. The APD gain bandwidth product has been demonstrated at 400 GHz in the charge multiplication region [3], but the integrated device does not realize this bandwidth. To extend the bandwidth from 10 GHz, gain has been sacrificed. The APD used in this thesis is bandlimited to about 25 GHz. The goal of the thesis is to process the data generated at 25 Gb/s. It is possible to design a channelizer to capitalize on the large available spectrum through wavelength division multiple access, but that is out of scope for this thesis.

Now, there is one remaining parameter after the *BW* and *S* have been constrained: the noise of the APD. Equalizing and compensating for the noise is a particular interest to the data processing and has been the chief impairment.

To understand the noise encountered in this work, the operation of APD needs to be

understood. The concept of an APD is built upon a PIN diode with a reverse bias applied to generate gain. The PIN diode is constructed using a p-doped layer of semiconductor and an n-doped layer of semiconductor. An optically transparent layer sandwiched in the middle is called the intrinsic layer. When a traveling photon strikes the detector, the energy of the photon $h\nu$ excites an electron across the bandgap of the PN junction where *h* is the Planck's constant and *v* is the photon frequency. The current generated by the flow of excited electrons is then measured by electronics. The electronics do add to the noise. Dark current is detection event that occurs without the presence of light on the detector. The band gap of the PN junction is much larger than any thermal variation that could errantly excite an electron across the gap. The probability of the photon exciting an electron is called the quantum efficiency, η. The responsivity, *R*, of the PIN is the amount of current produced for each watt of energy impinging on the detector.

$$R = \eta \frac{e\lambda}{hc} \tag{9}$$

e is the charge of the electron, λ is the wavelength of the light, and c is the speed of light. The noise statistics of the PIN detector is shot noise. Shot noise is modeled as a poissonian distribution.

The responsivity of the PIN detector is not adequate; there is a need for gain. The APD has the necessary gain over the PIN. The APD does this by replacing the intrinsic layer with two gain layers. [4] The absorption region generates electron-hole pairs when impacted by a photon called impact ionization. The multiplication region generates more electron-hole pairs because of a large applied electric field via a reverse bias voltage on the diode. The actual mechanics of this

are complex, but a phenomenon of note is the internal gain *G* of the APD. G is a random process. Each electron-hole pair has a chance to generate another pair. It is possible for the gain to create a current loop. The randomness of this phenomena produces a long transient response. The long response generates intersymbol interference (ISI) because the received pulses overlap because of pulse spreading. [5]

The internal gain noise needs to be explored further as it is the primary impediment in the data processing. Noise is randomness and there are at least two random processes occurring in the gain mechanism of the APD. The probability of an electron-hole pair creating a secondary pair is the electron-hole ionization ratio, *k*. The statistics of the internal gain noise is the conditional probability $p(m|k)$ of *m* secondary electron-hole pair creation events given some *k*. This conditional probability is transcendental. To make the analysis analytical, random process *G* is replaced with its mean $\langle G \rangle$ to create an approximation on $p(m)$ [6]

$$p(m) = \frac{1}{\sqrt{2\pi}\sigma_m}\big(S(m)\big)^{\frac{-3}{2}} e^{-\left(\frac{(m-\langle m \rangle)}{2\sigma_m^2 S(m)}\right)} \tag{10}$$

This function is another skewed upside-down bell-shaped function. The function $S(m)$ is called the signal-dependent skew factor

$$S(m) = 1 + \frac{(m - \langle \mathrm{m} \rangle)}{\nu\sigma_m}. \tag{11}$$

The function ν is

$$\nu = \frac{\sqrt{W_s F}}{F - 1}. \tag{12}$$

The functions *F* is the excess gain factor

$$F = \frac{\langle G^2 \rangle}{\langle G \rangle^2} = \kappa \langle G \rangle + (1 - \kappa)\left(2 - \frac{1}{\langle G \rangle}\right). \tag{13}$$

For these expressions, W_s is the expected value of the signal impinging on the APD and is given by

$$W_s = \frac{\langle m \rangle}{\langle G \rangle}. \tag{14}$$

This term represents the signal dependent nature of the internal gain noise. The variance σ_m^2 of the secondary electron-hole pairs is

$$\sigma_m^2 = \langle m^2 \rangle - \langle m \rangle^2 = \langle G \rangle^2 w_s (F + w_s) - (w_s \langle G \rangle)^2 = \langle m \rangle \langle \mathrm{G} \rangle F \tag{15}$$

The probability density function $p(m)$ of the generated secondary photoelectrons is a complicated function, but it can simplify back into a gaussian under appropriate conditions. When the skew function *S* is equal to 1, $p(m)$ simplifies into a gaussian. This happens when the gain is unity, $G = 1$. The excess noise factor *F* also becomes 1. This is the term *F* in Figure 2. As is expected, the excess noise factor F for a PIN diode with no gain is unity. Simply, more gain equals more skewing of the pdf because the increased gain produces a longer tail in the pulse response.

The skewing and importantly the signal dependent term affects the results in Equations 7

and 8. Each symbol will have a different likelihood function which makes the threshold calculation asymmetric

$$\Theta = \frac{\sigma_0 s_1 \pm \sigma_1 s_0}{\sigma_1 \pm \sigma_0}. \tag{16}$$

The probability of detection error is.

$$p_e = \frac{1}{2}\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{s_1 - s_0}{\sigma_1 + \sigma_0}\right)\right). \tag{17}$$

The last two equations are variations on the symmetric symbol case that is modified to handle the asymmetric symbol case. The equations hold for binary modulation. However, for multiple symbols or even multiple sequences, each symbol sequence can have its own probability distribution, which will have its own threshold and probability of error. [6]

Sandia simulated the performance of the APD and compared that performance to the baseline case of a PIN. They plotted the BER versus the supplied signal power as shown in Figure 2. The figure also includes a photograph of the APD. The simulation indicates there is a 7 dB gain over the PIN. Notionally, the APD can provide the necessary link margin to enable the energy efficient data center according to the figure. The simulation, however, does make several simplifications, which weakens that claim. The simulation did not include a skew factor but simplified that behavior to gaussians with different variances.

The performance of the communication systems distills to the consideration of the amount of energy used per bit, the wave that carry them, and the frequencies that compose the

waveform. Or as Nikola Tesla succinctly wrote “If you want to find the secrets of the universe, think in terms of energy, frequency and vibration.”

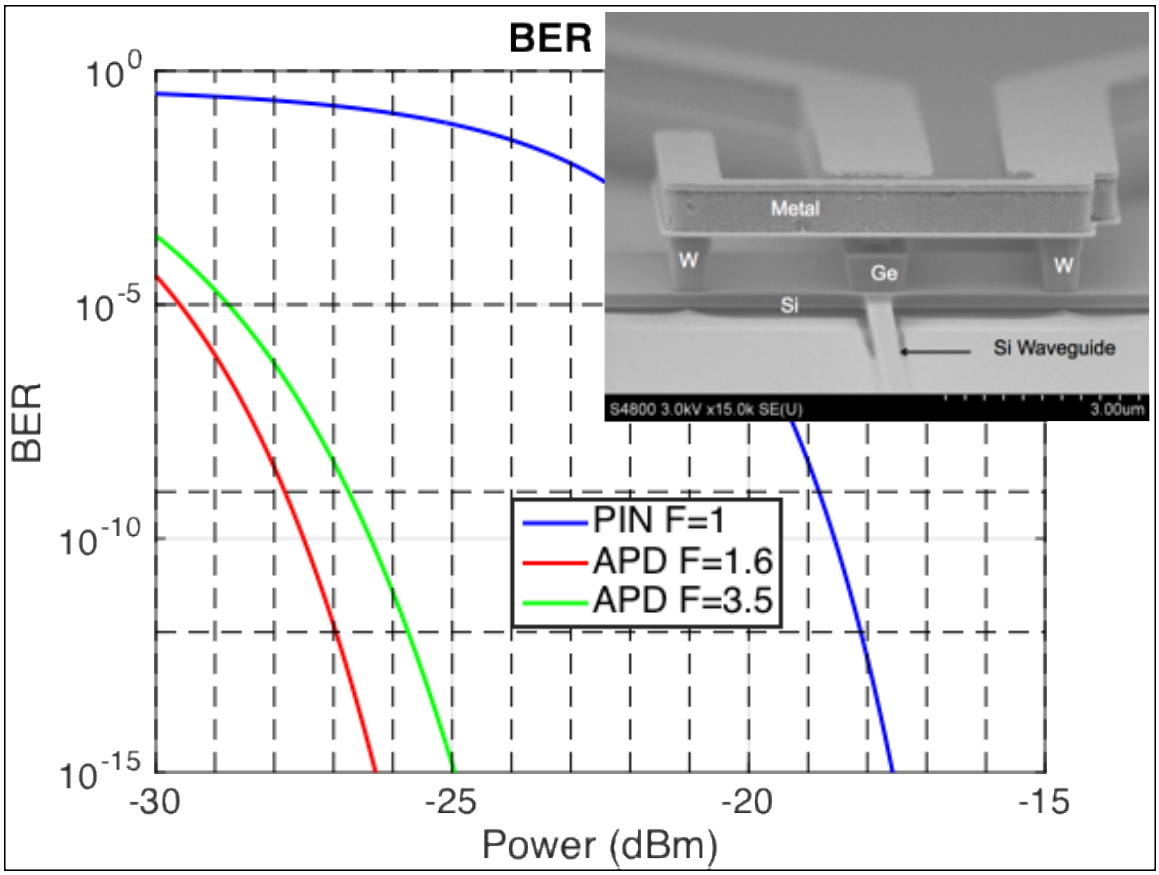

Figure 2: The simulated performance advantage of an APD.

Chapter 3 APPROACH

EXPERIMENTAL SETUP

This section describes the experiment setup used by Sandia to characterize the APDs and will provide context for the data processing. The physical architecture of the experiment is illustrated in Figure 3.

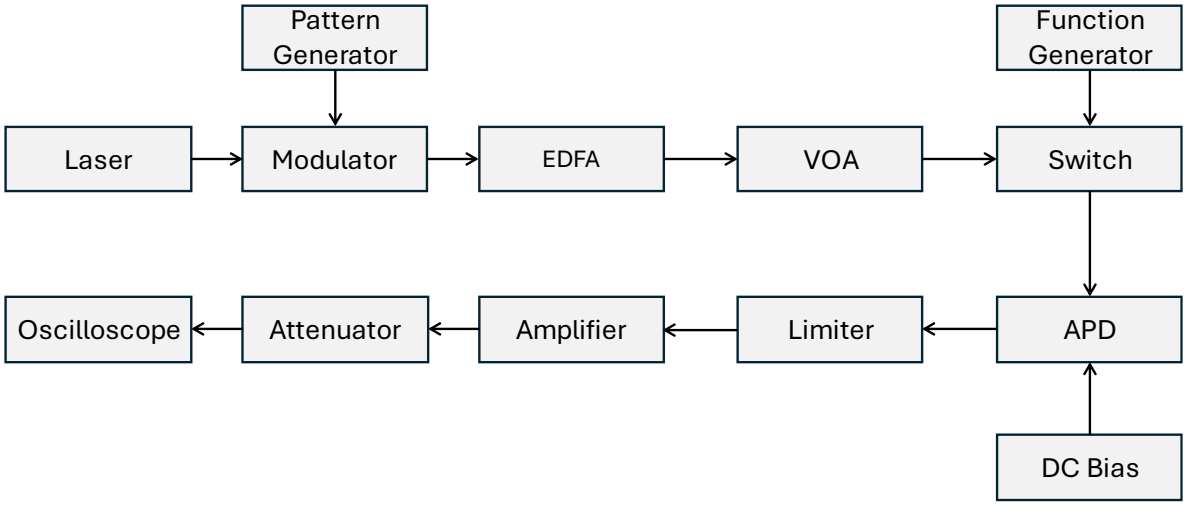

Figure 3: Block diagram illustration of the experimental set up.

Two experiments were performed at Sandia from this set up and the data generated from those experiments was analyzed as part of this thesis. One experiment probed an unpackaged APD that was still on the wafer. Two datasets collected from the probed APD were analyzed. The first had a symbol rate of 10 Gigasymbols per second (referred to as 10G). The second experiment was performed at a symbol rate of 25 Gigasymbols per second (referred to as 25G). The other experiment had a packaged APD. The data collected from the packaged APD was run at a of 23 Gigasymbols per second.

The relevant action of the experiment is a pseudorandom bit sequence (PBRS) that was generated by the pattern generator and then modulated onto the laser light by the modulator driver. The PBRS is $2^{15} - 1$ bits long. As transmitted, the PBRS lasts 3.277 microseconds for the 10G data and 1.310 microseconds for the 25G data. The light then travels through an Erbium Doped Fiber Amplifier (EDFA) and a variable optical amplifier (VOA). These amplifiers are only relevant to the specific experiment and would not be present in the operational system. The amplified laser light is on-off modulated a second time by the switch modulator for 100 microseconds. The light goes through two modulators. The effect is that the APD starts in an off configuration and then transitions to an on configuration. The PBRS is repeated for the duration of the "on" state. For the 10 G data, the pattern is repeated 30 times. For the 25G data, the pattern is repeated 76 times. The intention is to study the burst mode operation of the APD. The repeated patterns are referred to as frames from now on. This invokes the concept that the PBRS pattern is analogous to a frame from a communication protocol in the datalink layer. After the switch modulator, the light is finally fed into the APD, either packaged or probed.

The output of the APD is routed through a limiter, amplifier, and an attenuator before reaching the oscilloscope. The oscilloscope is a real-time oscilloscope capable of 256 Gigasamples per second. The 10G data set was sampled at 25 samples per symbol after the resample. The 25G dataset was sampled at 10 samples per second after the resampling. The data output from this experimental set up is called the real data for the rest of the thesis.

The two APDs were investigated by Sandia. The probed APD was unpackaged and was probed directly on the wafer. The probed APD suffered from impedance mismatch which caused drifts and significant ISI. The second APD was packaged. The packaged APD had a better impedance match. The packaged APD was not operated in burst mode while the probed APD

was. The 23G dataset was collected from the packaged APD. Multiple datasets were collected from each APD for different bias voltages. For the probed APD, it was unclear from the start of the work that it would be possible to process the data collected from it. So, only one of the bias voltage datasets was selected for investigation. The analysis of the packaged APD investigates the receiver performance for different bias voltages. The most relevant bias on the packaged APD was 15 volts. The 23G data presented later in the thesis are with a 15-volt reverse bias unless otherwise noted.

This project lasted three years. The first datasets available came from the probed APD. The research originally focused on the question of whether the data could be processed despite the impairments. All the techniques discussed and developed in this thesis were aimed at the difficult 25G data. 35 months into the project, the packaged APD data became available. The results of the 25G data analysis were then redirected to the 23G data from the probed APD 25G data.

APPROACH

One of the main complications addressed in this thesis is mitigating the effect of burst-mode operation on the APD characteristics. Sandia performed one experiment to measure the effect of turning the light on and off. Figure 4 is some of the data from that experiment. There is an underdamped oscillation that occurs when the light turns off and on as can be seen by simple inspection of the figure. The data plotted in Figure 4 is not the data used in this thesis. The data processed is depicted in Figure 6, though that data set still exhibits the same transients.

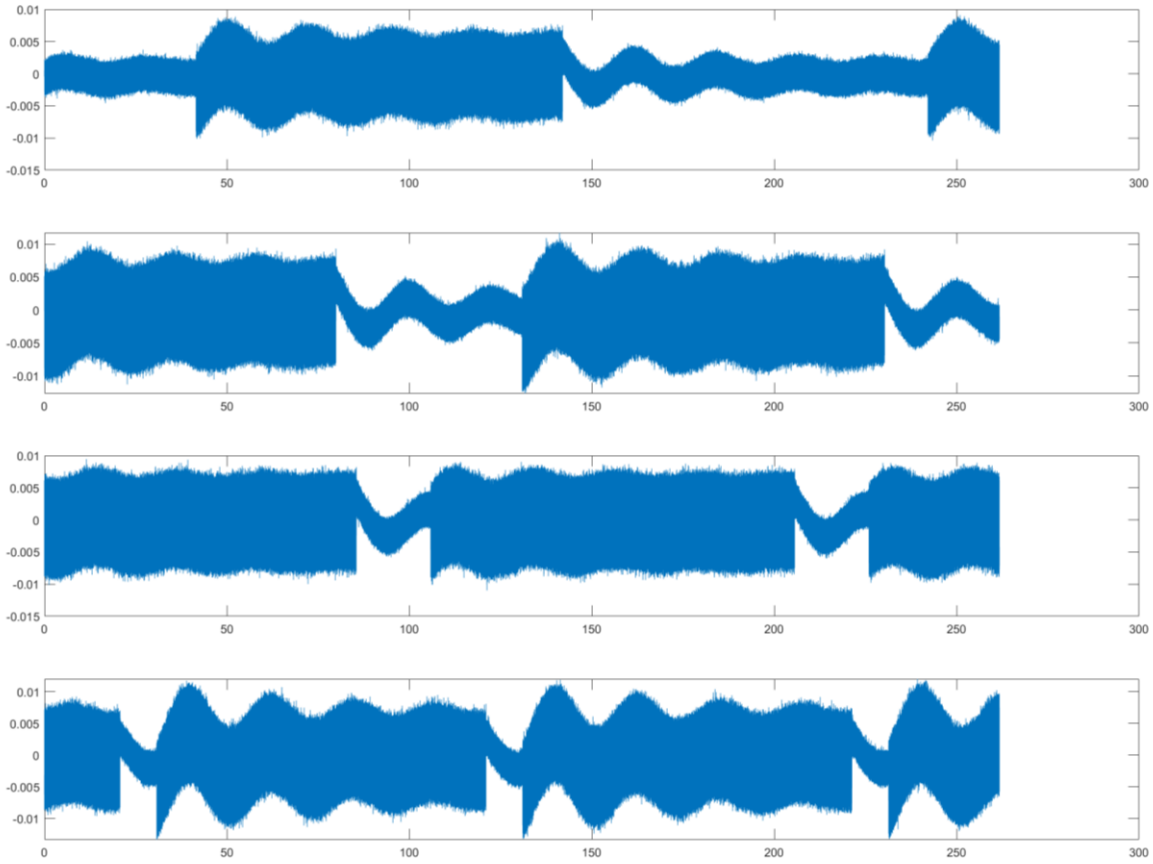

Figure 4: Example traces of the unpackaged APD operating in burst mode.

Our approach for the analysis was to construct a synthetic dataset from the same PBRS used to modulate the signal in the real experiment. Then, this synthetic dataset was made to be incrementally complex to model the various impairments over the course of the research effort. As the complexity was increased, the goal is to have the synthetic data mimic the real data. The base synthetic data set is rectangular pulses that nearly match the same PBRS fed into the experimental set up. The 10G synthetic data matches the 25 samples per symbol of the real data, and 10 samples per symbol likewise for the base 25G synthetic data.

We then developed a data processing method for each type of impairment. After the impairments are added to the synthetic data set, the corresponding method was tried on the real

data. The ultimate method was maximum likelihood sequence estimation (MLSE) with bespoke known sequences or codewords taken directly from the data. The 10G data requires less processing and fewer methods were used.

Table 1: Methods used on the datasets.

	Dataset		
	Unpackage 10G	**Unpackaged 25G**	**Packaged 23G**
Method	Integrate and Dump	Integrate and Dump	Integrate and Dump
	Matched Filter	Matched Filter	
	FIR Filter	FIR Filter	
	8 State MLSE Trellis	8 State trellis	8 State trellis
		32 State trellis	32 State trellis
		8 State MLSE with state-dependent metric	
		32 State MLSE with state-dependent metric	
		8 State MLSE “Full Sequence”	
		32 State MLSE “Full Sequence”	

Chapter 4 DATA PROCESSING

PROBED APD

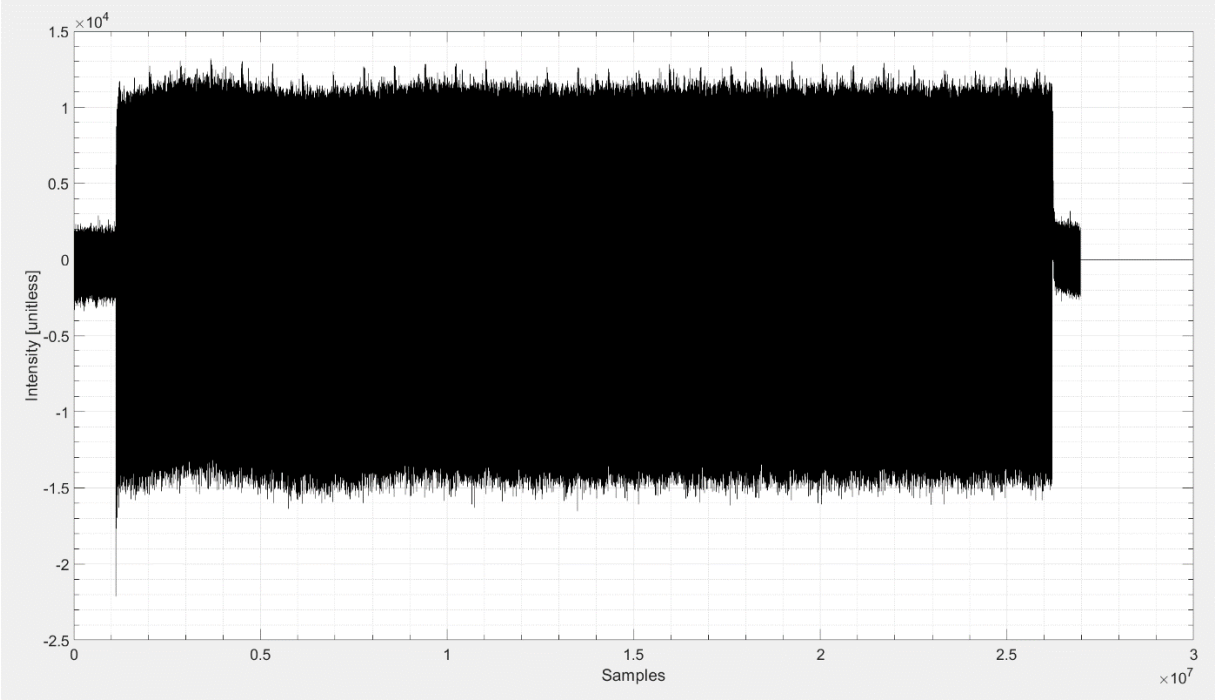

Figure 5: Raw output from the probed APD for the 10G dataset.

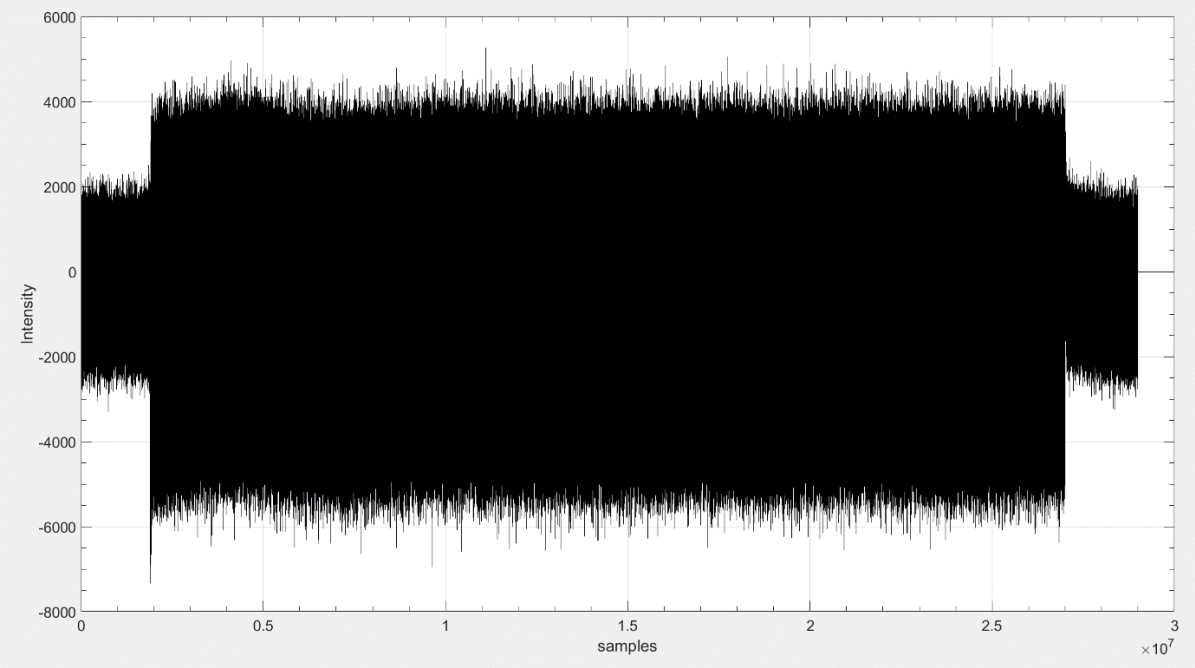

Figure 6: Raw output from the probed APD for the 25G dataset.

Two of the data sets generated from probe measurements of unpackaged devices taken at Sandia are shown in Figure 5 and Figure 6. These figures represent all the probed measurements used in this thesis. The figures show unitless intensity across samples taken. Each sample corresponds to 4 picoseconds, but the x-axis is unitless as well. So, each 25G symbol has a temporal duration of 40 picoseconds and the 10G symbol has a temporal duration of 100 picoseconds before the channel. The oscilloscope used at Sandia had a sample rate of 256 Gigasamples per second.

The first operation on the dataset is to resample it using MATLAB. The default resampling algorithm in MATLAB uses interpolation. [7] The algorithm resampled the data to create 250 samples from the original 256 samples for 25 samples per bit for 10G and 10 samples per bit for 25G.

We observe two regions in the data collected by the real time oscilloscope. The first is

when the light is turned on and is referred to as the “signal plus noise region”. The second region is when the light is turned off and is referred to as the “noise only” region. The separation of the waveform into these two regions allows the burst-mode behavior of the APD to be analyzed during an optical switching event. The switch in the test setup shown in Figure 3 was specifically added to evaluate the burst mode behavior of the APD.

The first notable artifact from the data is the spike at the beginning of the “on light” part of the waveform. The most probable cause of this spike is a transient caused by the burst mode behavior of the APD and possibly electrical reflections caused by an impedance mismatch between the probe and the device. The transient resembles the response an undamped driven oscillator to an impulse. It does not settle down until halfway through the signal region. It is for this reason that the data processing is only done in the last half of the signal region after the 10s of microseconds that the transient lasts.

Another artifact is the presence of harmonics at the output of the APD and its electronics. These harmonics are seen as the spikes in Figure 5 and Figure 6. The spectrum of these signals, generated by using a fast Fourier transform (FFT), of both the signal region and the noise region are displayed as log-log plots in Figure 7 and Figure 8. Both are log-log plots. The Y axis is the amplitude, and the X axis is the frequency in Hertz. It is important to note that the harmonics are present in data sets irrespective of the presence of light. This suggests that the harmonics are not intrinsic to the lightwave signal but are caused by some other artifact of the probing of the unpackaged devices. For the processing of the waveforms from the unpackaged devices, the spikes are removed with notch filters centered at the exact locations of the spikes in the spectrum. The inverse FFT is taken to produce a signal that is free from the harmonics. These harmonics appear not to be statistical in nature and seem to be a form of systematic error in the

experimental set up. A better solution is to remove the harmonics with a hardware solution instead of a software solution. Indeed, the 23G from the packaged APD is the hardware solution, but it was only available later on in the project.

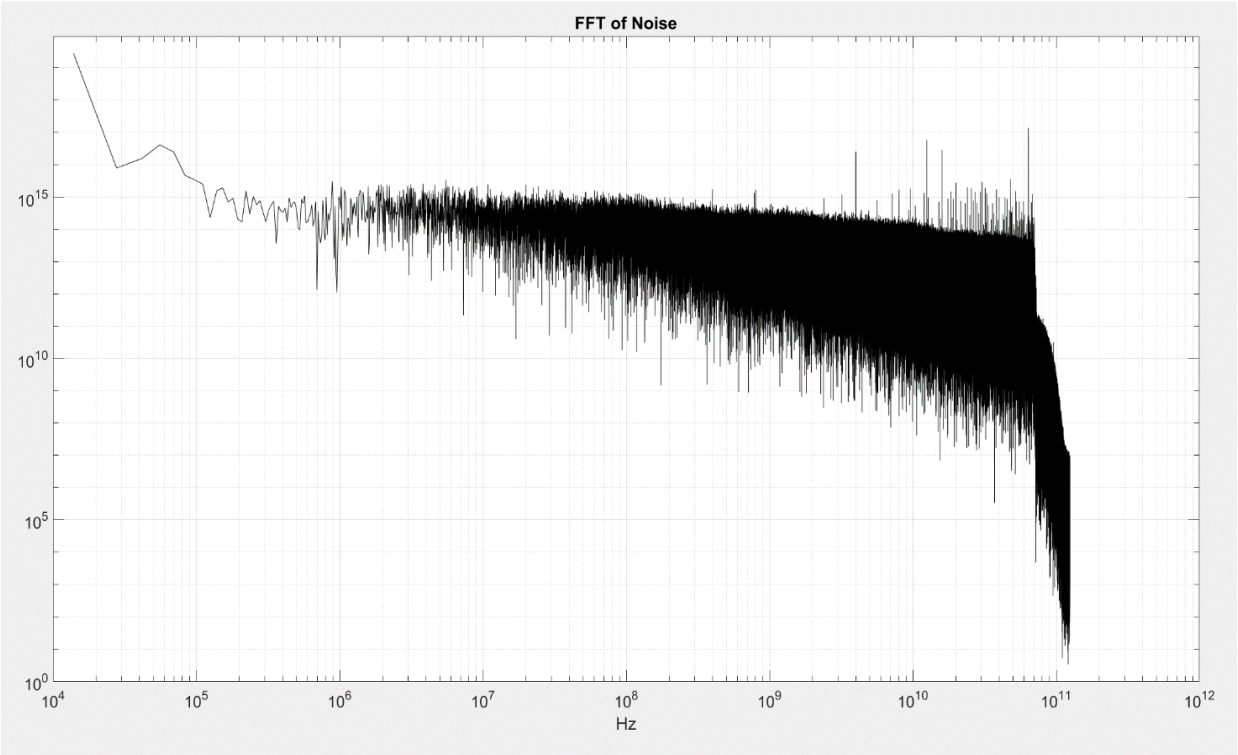

Figure 7: FFT of the “noise only” region for the data from the unpackaged device.

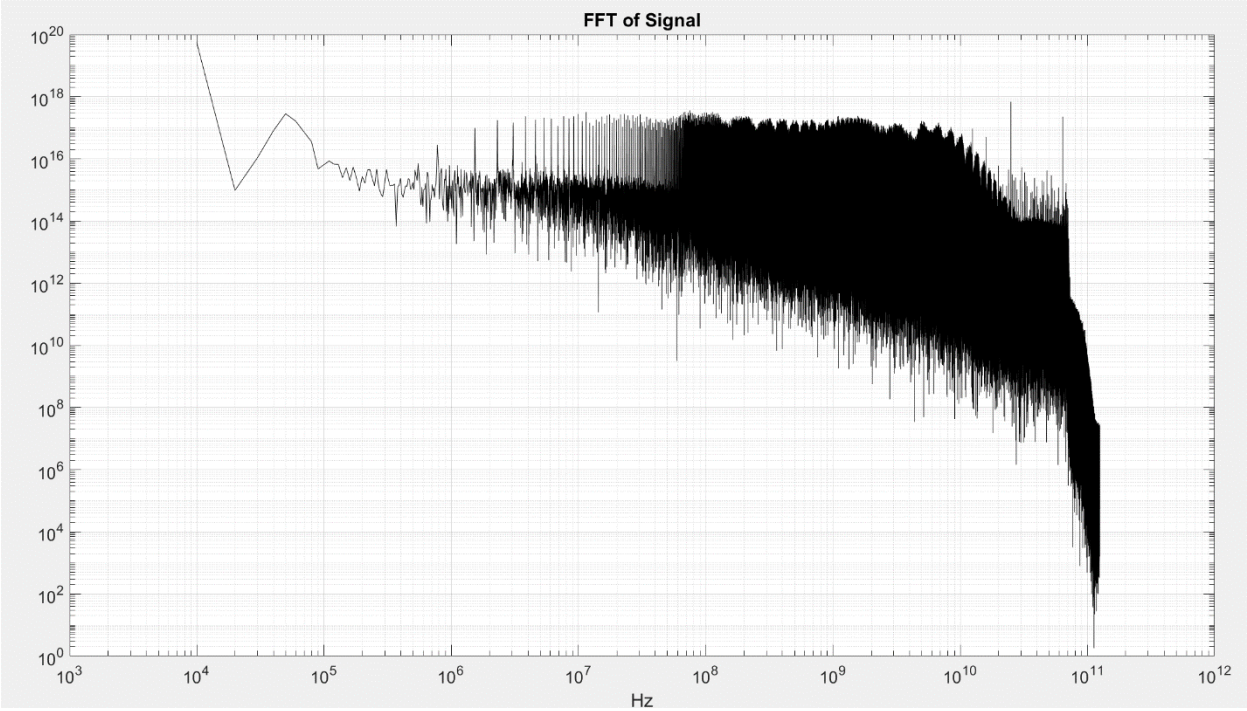

Figure 8: FFT of the "signal + noise" region without filtering
region for the data from the unpackaged device.

After the harmonics are removed, the next step in the data processing is analyzing the "noise only" region. The statistics of this region can be determined directly from the data. Figure 9 shows a histogram of the "noise only" region. It is clearly a gaussian showing no signal dependence.

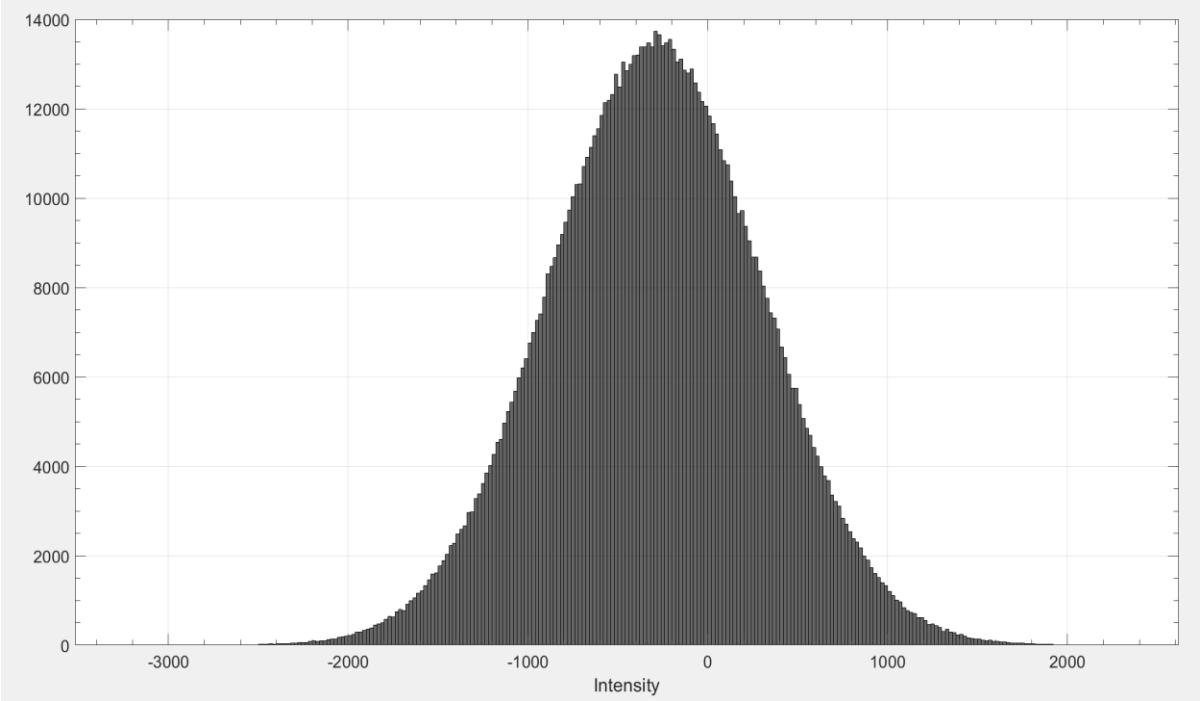

Figure 9: Histogram of the no light region.

The determination of the noise statistics for the “signal+noise” region is difficult to analyze because it requires *a priori* knowledge of the signal.

One simple qualitative method to determine the performance of the APD is generating and analyzing an eye diagram. Eye diagrams superimpose traces from many symbol transitions. Any ISI is apparent from inspection of the diagram. The communication system works well when the eye is said to be “open.” The height of the eye is proportional to the OMA or optical modulation amplitude, which is related to the SNR. Figure 10 shows the eye diagram of the 10G data and is comprised of 2000 traces. For this data rate, the eye is “open” qualitatively indicating that it is possible to successfully process the data to determine the bit error rate (BER). The eye has an approximately trapezoidal shape. This is caused by ISI as the symbol energy spreads into its neighbors. It is reasonable to conclude from Figure 10 that there is some ISI present in the datasets.

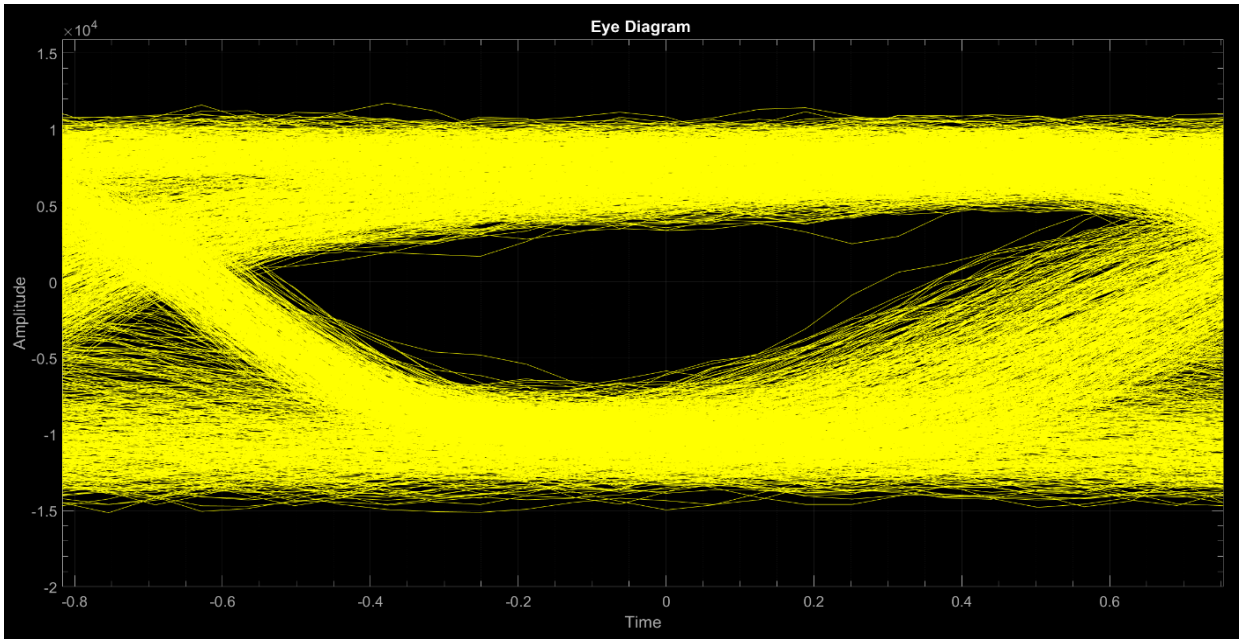

Figure 10: Eye diagram of the 10G data.

Figure 11 is the eye diagram of the 25G data. From inspection, the eye is said to be “closed”. This is evidence that the communication system will not perform well at 25G using simple threshold-based techniques. The 25G dataset is of particular interest to this thesis due to the difficulty of processing. Most of the effort in this thesis has been applied to determine if data can be extracted from the 25G waveforms and this the dataset was the default dataset used in this thesis.

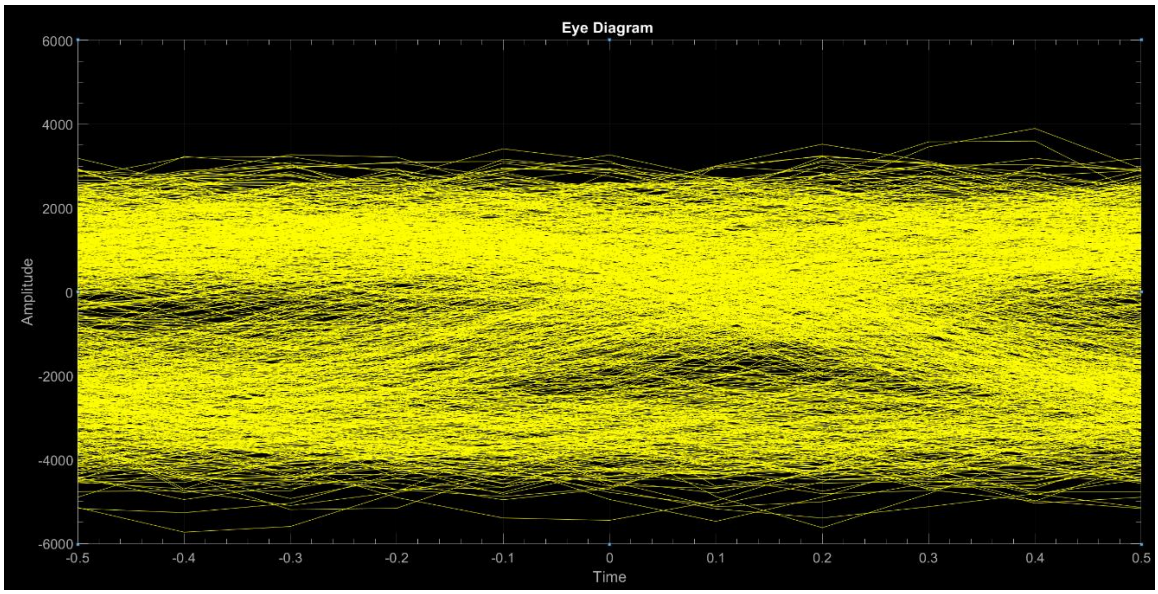

Figure 11: Eye diagram of the 25G data.

The quantitative processing of the data sets begins with the resampled data after the removal of the harmonics. The data is then normalized for further processing. Figure 5 and Figure 6 show that the data as collected has both negative and positive electrical amplitudes. The APD is a specific type of photodiode that measures intensity instead of electric field. Because intensity is inherently positive, the negative-valued raw data is an artifact of the electronics used to bias the unpackaged APD. Level shifting the data set is another parameter that can be added in software or hardware for the convenience of processing. Another artifact is a long-term time varying drift in the data, which is on the order of several symbol intervals. We conjecture that there are some uncharacterized, signal-dependent parasitics for the unpackaged APD that cause this long-term drift. This drift will affect the data detection process and needs to be precisely compensated for sequence detection to be effective.

To address the normalization and drift issues, the first step is to find the global mean of the data and subtract that value to the entire set. This removes the global drift. Then the global root mean square amplitude is used to estimate the global energy. The dataset is divided by the

estimate of the global energy to make generate a normalized unit energy over a single symbol interval. In addition to the global normalization, different local normalization windows were tried to compensate for the time-varying drifts seen in the data. Normalization windows were chosen from five to hundred symbol lengths.

Sophisticated signal processing methods such as sequence estimation require the channel response. This channel response may be time invariant and therefore characterized by an impulse response $h(t)$, or it may be time varying $h(t,\tau)$ depending on both the absolute input time τ as well as the output time t instead of the difference $t-\tau$ as would be the case for a time invariant system. Compensating for a time varying channel posed significant challenges.

Once the data was normalized, the determination of the channel response requires finding a portion of the measured dataset that matches some known sequence that was transmitted. To do so, a cross correlation is taken between the desired sequence used to determine the pulse response and raw data. The desired sequence logical sequence of data is then transformed into a rectangular waveform using the appropriate number of samples per symbol. The target sequence is then level shifted so that a "0" symbol corresponds to a unitless intensity of -1 and the "1" symbol has a unitless intensity of +1. The level shift for this part of the processing is chosen to create an antipodal data set that is nearly zero mean.

This resulting data set will produce a strong cross-correlation peak when the raw data sequence is aligned with the target response. A manual check was done to verify that the signal corresponds to the target sequence. A feature of the repeated PBRS is that the target sequence is repeated at regular intervals of 2^{15}=32,767 symbols. Therefore, for the 25G data set, the n+2^{15} sample corresponds to the same sequence. This fact provides the opportunity to investigate the mean signal sequence by averaging across all the PRBS frames to mitigate the effect of noise and

produce and estimate a “noiseless” sequence needed for sequence estimation. Figure 12 and Figure 13 show an example of two target sequences from the signal. The top halves of the two figures are the stem plots of the sample-by-sample average. The bottom half is the histogram of the aggregate of all the samples over all the PBRS frames. From the stem plot, intersymbol interference (ISI) can easily be seen. The symbol width is 10 samples; therefore, the rise and fall time leak into the adjacent symbol intervals. Figure 13 has a ‘101’ subsequence. In that subsequence, the “0” symbol does not return to the “DC” value and does not respond quickly to the subsequent “1”. For this specific sequence, using bit-by-bit detection would fail to detect this “0” symbol.

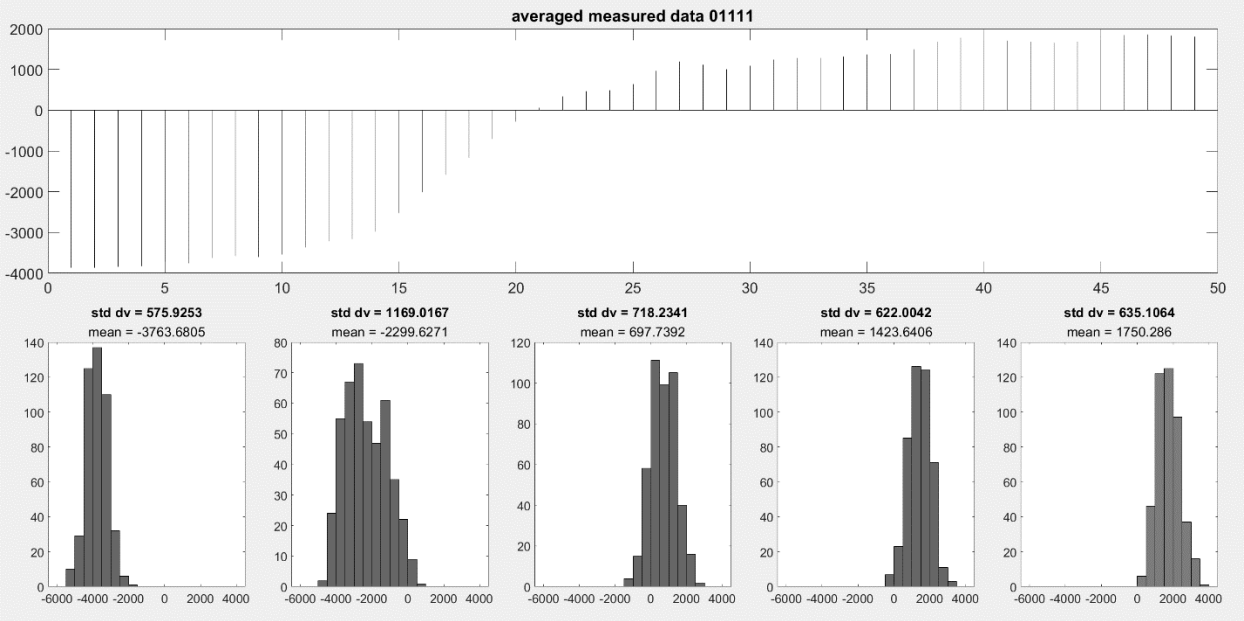

Figure 12: The 01111 measured sequence extracted from the full sequence.

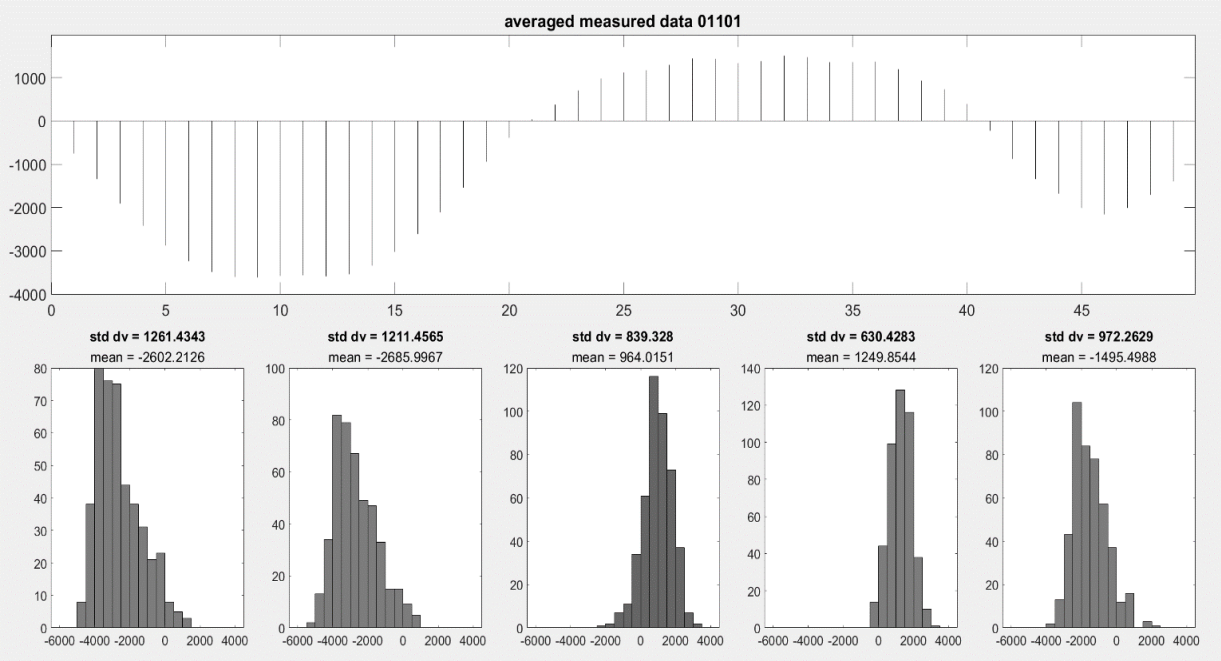

Figure 13: The 01101 measured sequence extracted from the full sequence.

The method to find a target sequence within the larger sequence can be used to determine the pulse response by using a target sequence that has an isolated ‘1’ such as the sequence ‘000010000’. The measured pulse can then be averaged across all the frames. Figure 14 and Figure 15 show the average of the measured pulses. This estimate of the “noise-free” pulse is then used in the sequence detection algorithm as described below.

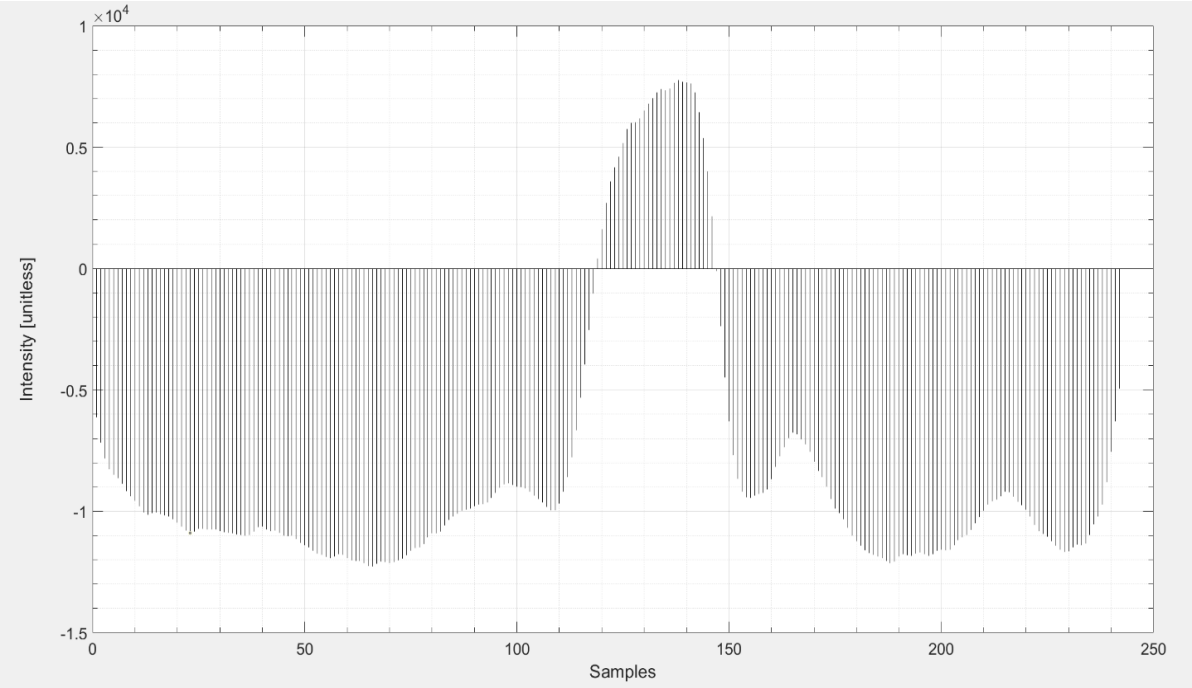

Figure 14: Pulse Shape of 10G data.

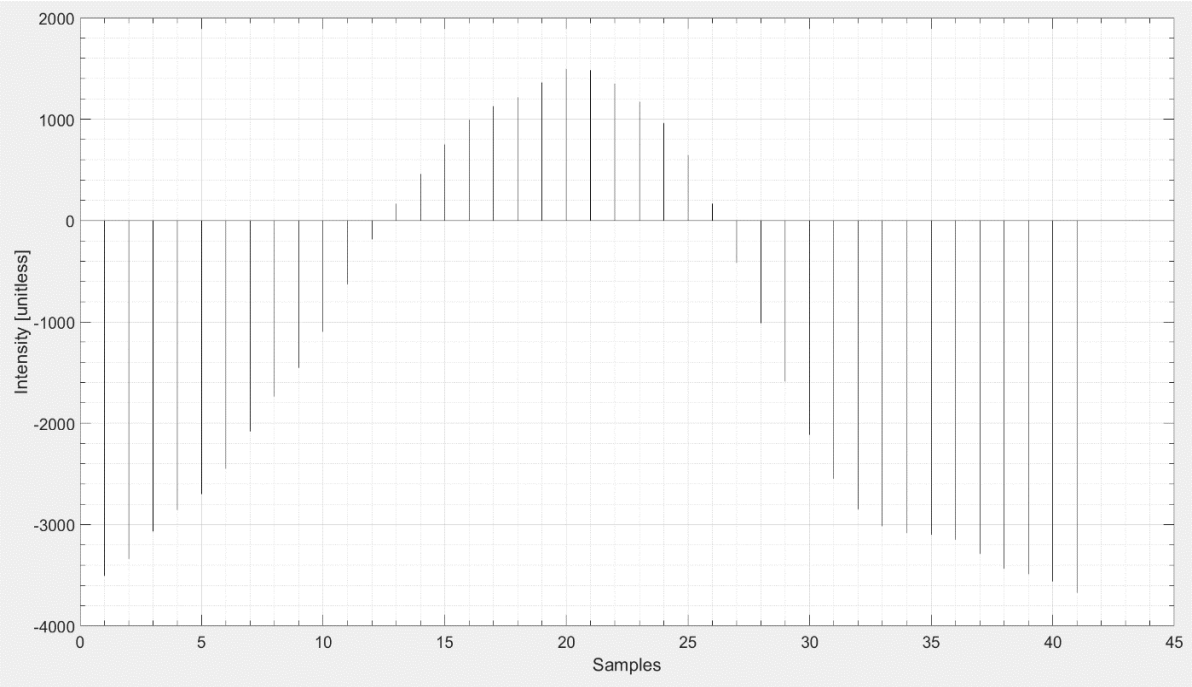

Figure 15: Pulse shape of 25G data.

The most prominent feature of the measured pulse is the spreading of the pulse energy into adjacent symbol intervals. This band limiting effect resembles a low pass filter. For the 10 G data, the band limiting produces significant signal energy out to the 3rd subsequent symbol, which is 75 samples. For the 25 G data, the band limiting process produces energy out to the 5th subsequent symbol, which is about 50 samples. As can be seen in the data, the 25G pulse is subject to much worse ISI because of the larger spectral content of a 25 G pulse compared to the 10 G pulse. Specifically, the 25G pulse takes more than a one entire symbol interval to reach its peak intensity.

A simple mitigation method for this effect is to use a filter. In the absence of ISI, the typical filter of choice is the matched filter. The matched filter is the optimal detector for a received signal that only has additive white gaussian noise (AWGN) but is not optimal when ISI is present. The measured pulse was cross correlated with the data set to find clean peaks that could be discriminated into ones. Because of the large amount of ISI, the matched filter was ineffective, as expected.

Given that we estimated the channel response *r* for a single pulse, the impulse response *h* can be estimated from the matrix equation

$$\underline{r} = \mathbf{T}h + \underline{n} \tag{18}$$

where **T** is a matrix representing the known training sequence ‘000010000’ used to determine the pulse response and *n* is the noise. The solution for the impulse response *h* is plotted in Figure 16.

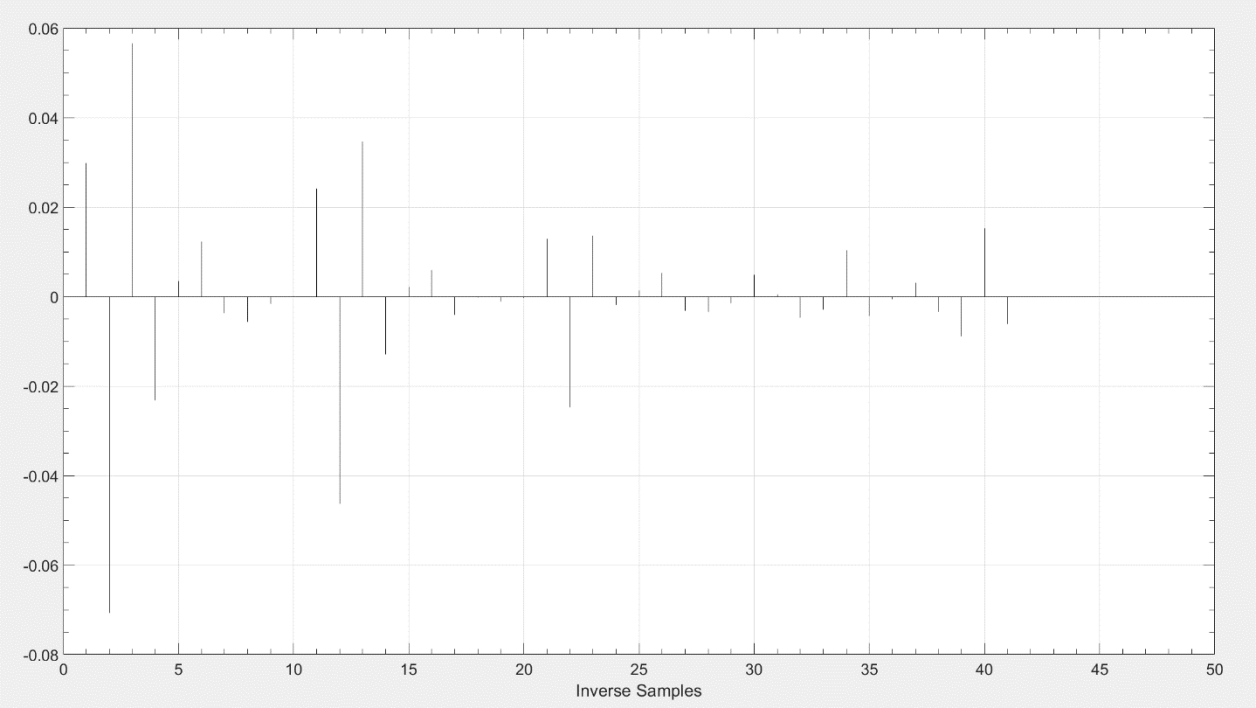

Figure 16: Estimated impulse response *h(t)* of the channel and the APD. The duration of the impulse response is over five symbols long for the 25G symbol rate.

Once the impulse response *h(t)* is estimated, its Fourier transform is computed to find the frequency response $H(\omega)$. A filter is constructed by taking a raised cosine function (the raised cosine function is parameterized by a number called β and dividing by $H(\omega)$ in frequency space. The last step inverts the Fourier transform to create the final filter in the time domain. The filter is convolved with the dataset. The core idea was to attempt to redistribute the pulse energy into a smaller temporal duration thereby reducing the effect of ISI. The parameter β determines the roll-off of the filter and different values of beta were used to optimize the filter. The simple raised cosine filter was not adequate to compensate for the ISI. There are other kinds of FIR filters that are commonly used, but they were not attempted and would not be adequate.

PACKAGED APD

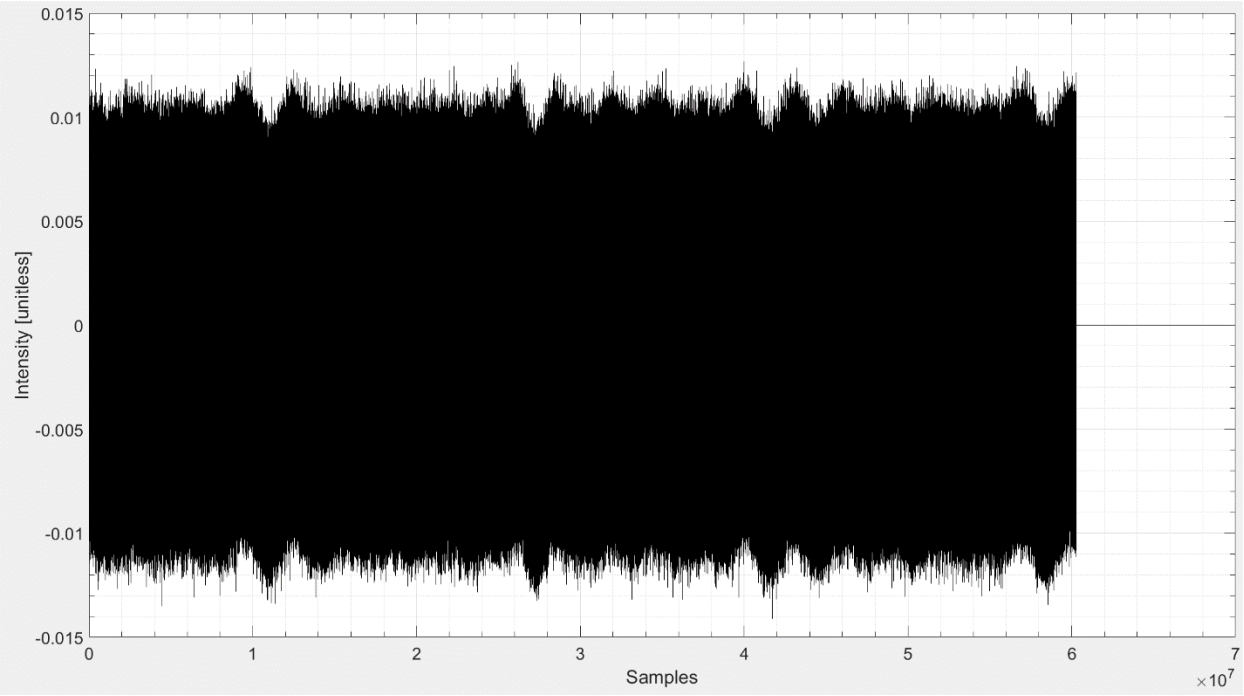

Figure 17: 23G dataset from the packaged APD using a 15-volt bias.

Figure 17 shows one half of the complete data set from the oscilloscope for the 23G dataset from the packaged APD used later in this project. The bias voltage was set to 15 volts. As presented, it has been resampled to 230 samples from 256 samples by the MATLAB resample algorithm. Each symbol is 40 samples long after resampling. The 23G dataset is not modulated by the switch, which does not allow investigation of the burst mode operation of the APD. It only has a "signal + noise" region. After resampling, the data was normalized with a global normalization factor. Even though the experiment is not run in a burst-mode, there is some drift in the normalization. The figure of the 23G contains 25 million samples and the drift can be seen clearly by eye. This time-varying drift encumbers the sequence estimation, but the packaged 23G

data drifts less than the unpackaged 25G data. The drift is quantized in the histogram in Figure 18. The energy is calculated by the root mean square intensity values. The root mean square is used because the mean of the data is zero.

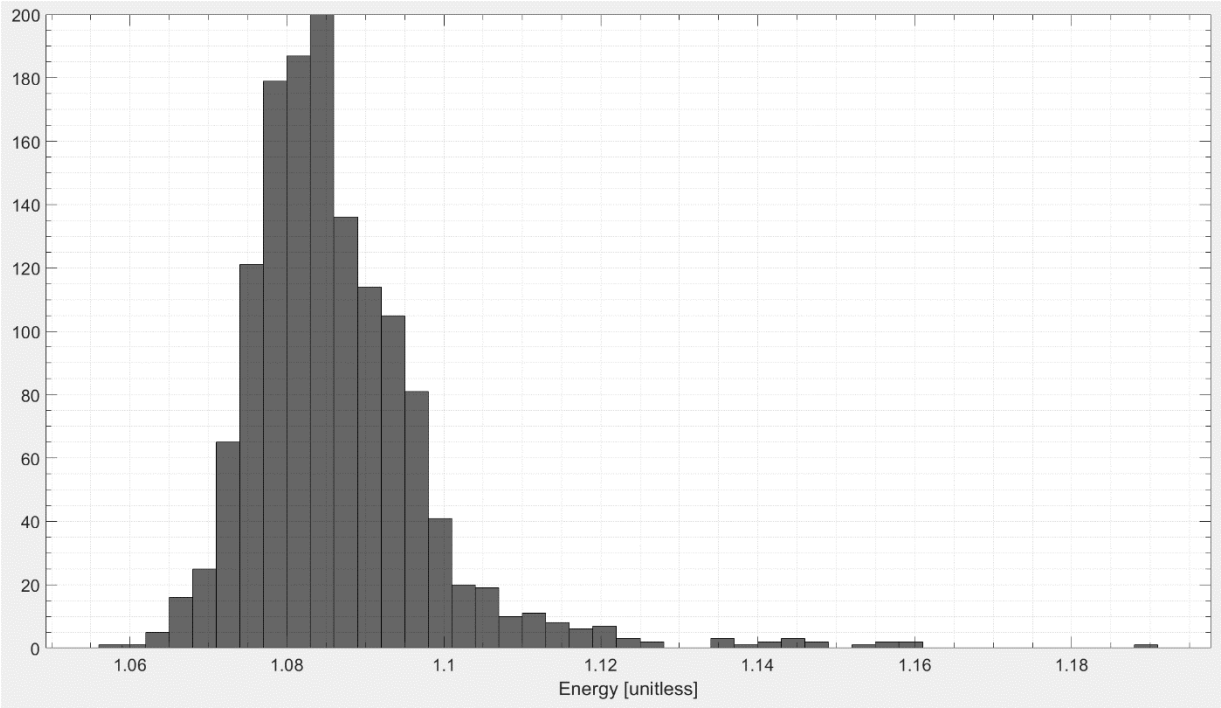

Figure 18: Distribution of energy per 1000 bits over the 23G data.

The drift energy per bit of the 23G data is a small impairment. The primary concern is ISI. The eye diagram of the raw data is shown in Figure 19. The eye of the packaged APD at 23 G is “open” and is substantial improvement over the unpackaged 25G dataset shown in Figure 11.

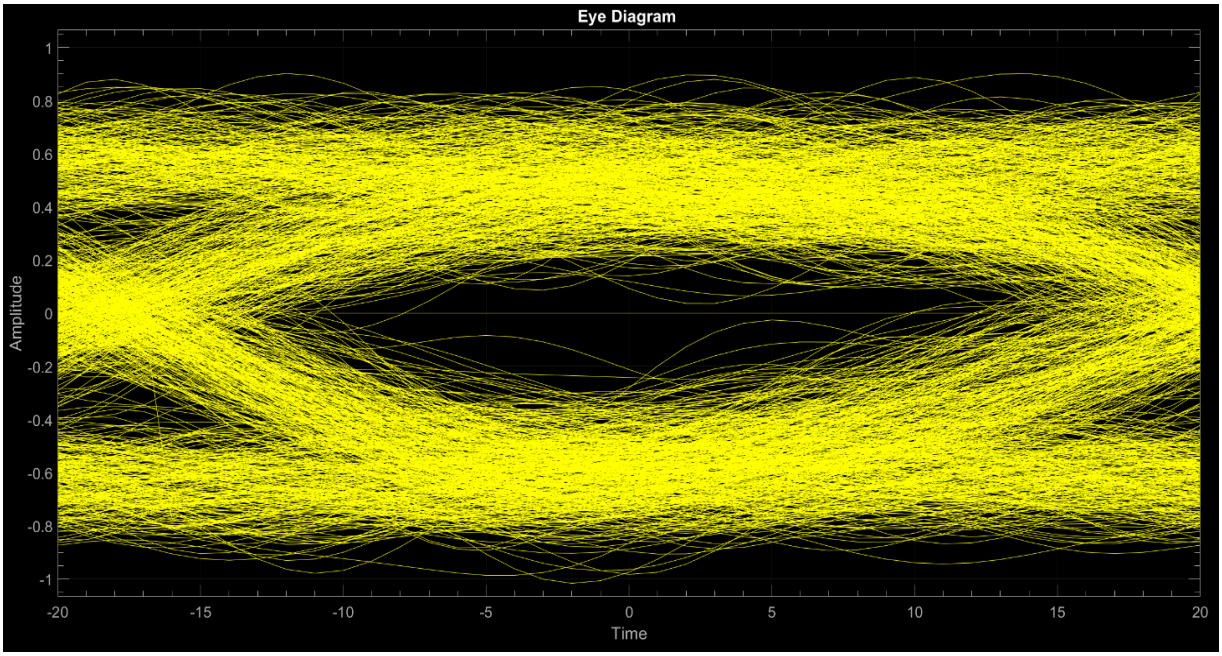

Figure 19: Eye diagram for the 23G dataset.

Since the 23G data was generated from a packaged device, there is less impedance mismatch and harmonics in the data compared to the probed device shown in Figure 8. Indeed, the harmonics seen in Figure 20 are significantly smaller compared to the unpackaged device shown in Figure 7.

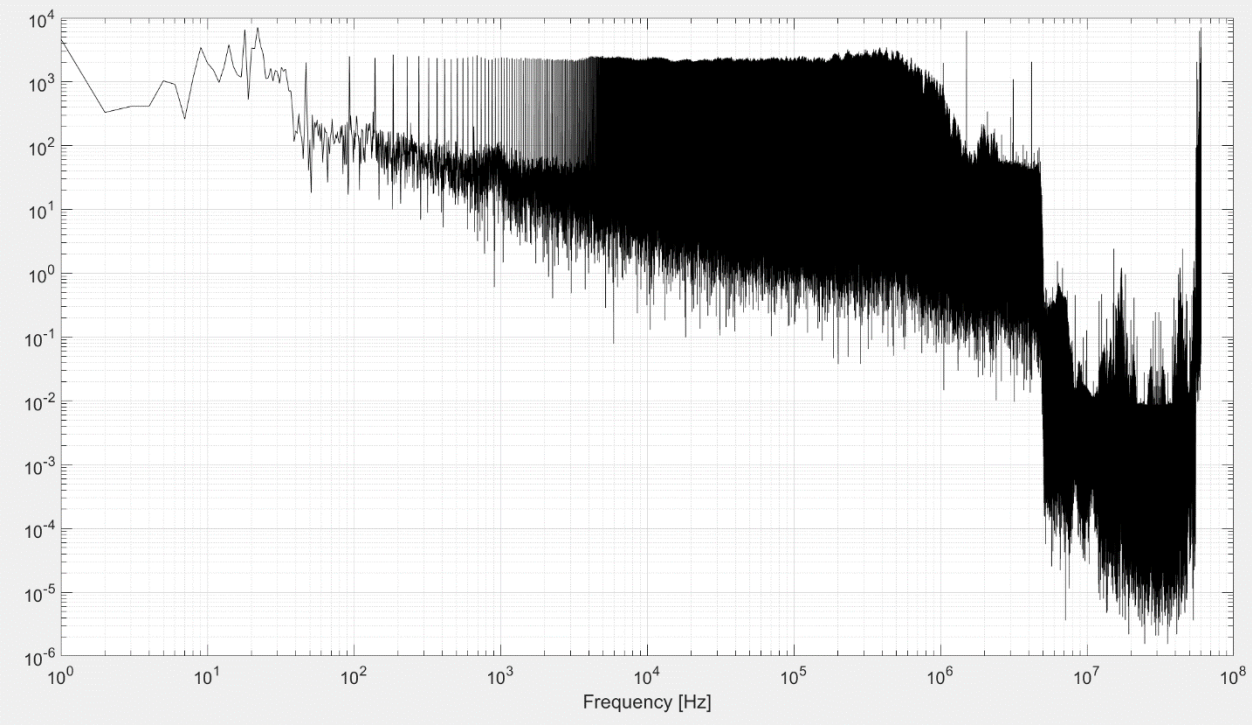

Figure 20: FFT of the 23G dataset for the data from the packaged device.

Figure 21 is the pulse shape for the 23G data at a 15-volt bias. The transmitted signal has a pulse width of 40 samples. The received pulse width from the figure has a full width at half maximum of approximately 50 samples. ISI is present, but much less so than the probed APD. Using these values, the constraint length used for sequence estimation, which is amount of memory in the system, is less than two symbol intervals.

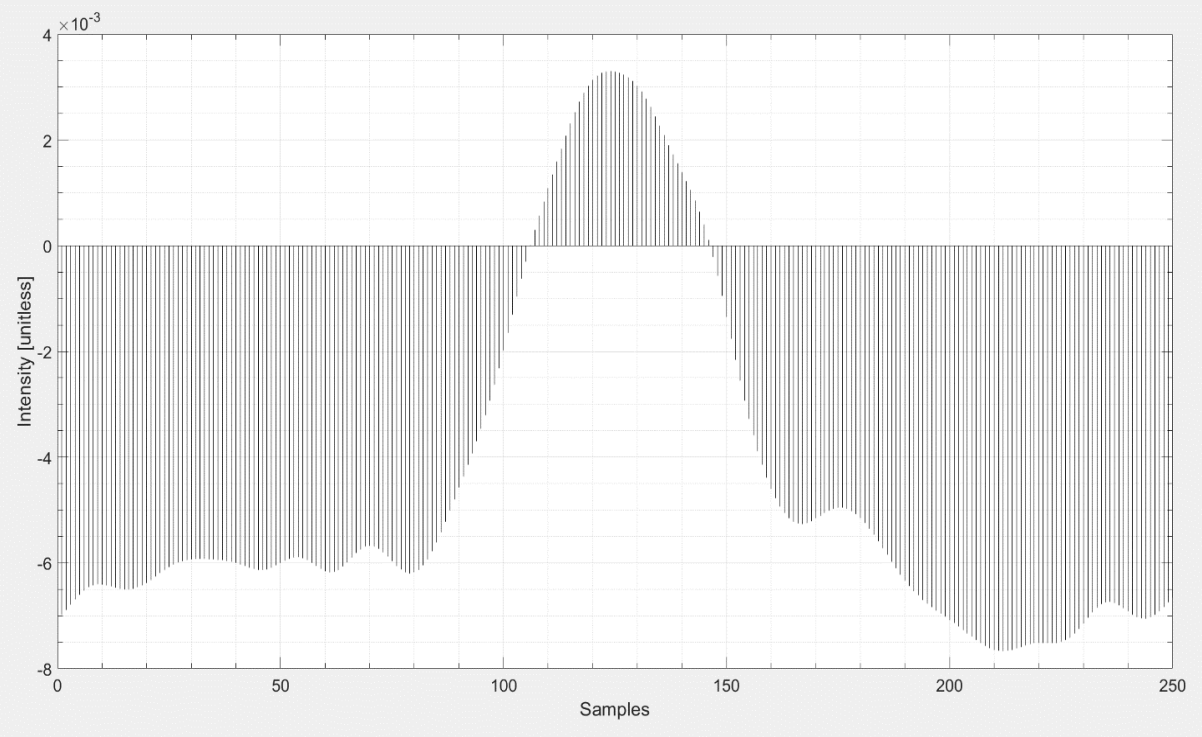

Figure 21: Pulse shape for the 15-volt 23G data from the packaged APD.

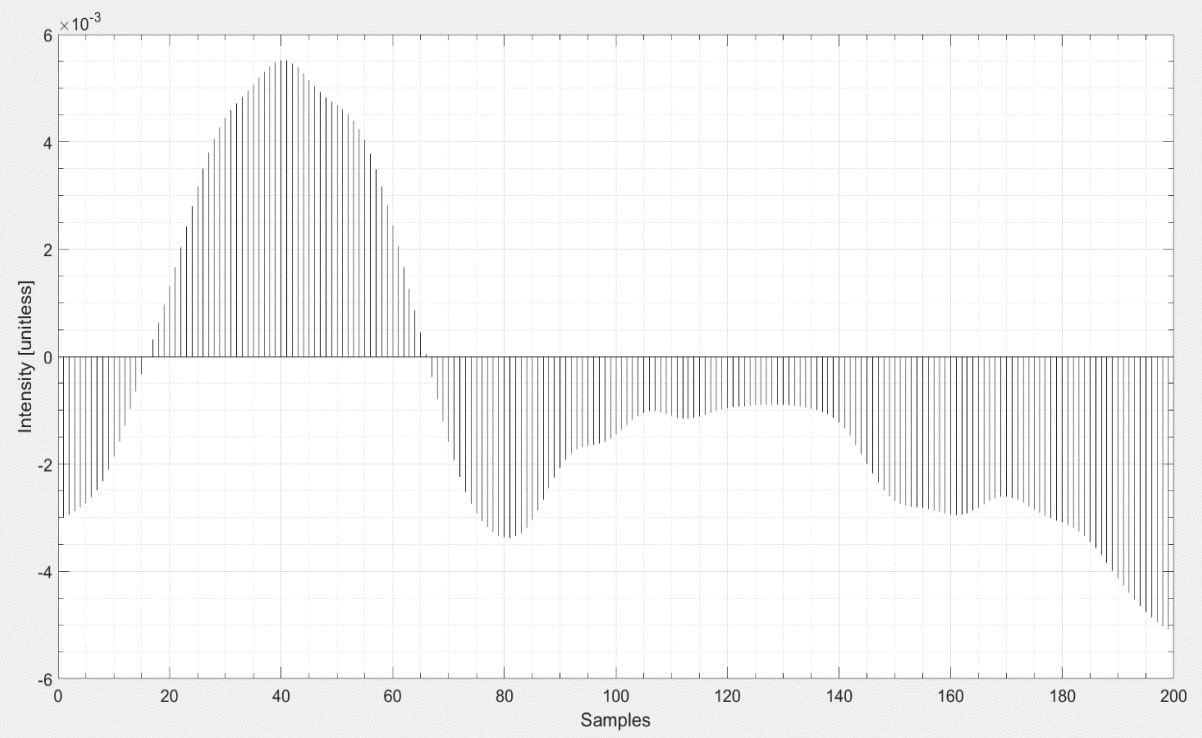

Figure 22: Pulse shape for the 12-volt 23G data.

Chapter 5 MAXIMUM LIKELIHOOD SEQUENCE ESTIMATION

The simplest form of demodulation decides each bit using a single symbol interval. In this bit-by-bit detection method, the energy used to determine the most likely transmitted symbol is accumulated over a single symbol interval. No information outside each symbol interval is used. After a fixed duration determined by estimating the clock transition, a sample is constructed from the total energy in a symbol interval and that sample is compared to a threshold. At its most basic form, this simple integrate and dump process does not compensate for the presence of ISI and does not use any filtering.

The ISI for the 25G dataset affects the five symbols for each pulse and neither bit-by-bit detection or simple filtering is sufficient to recover the data. For this case of severe ISI, accurate detection can be accomplished by comparing blocks (or sequences) of bits using a process called sequence detection, which was briefly discussed earlier in the thesis. If the ISI affects five symbol intervals, it is possible to compare sequences of five symbols long. The length of the sequence that is used for this comparison is defined as the constraint length ν. Looking at Figure 14 and Figure 15 for the unpackaged devices, the constraint length for the 10G data is three and the constraint length is five for the 25G data. A constraint length of five for binary symbols means that there are $2^5 = 32$ sequences to compare for each bit discriminated. This is computationally expensive. Instead, a trellis algorithm compares two partial sequences called branches.

Referring to the illustration shown in Figure 1, the two branches are continuations of the last $\nu - 1$ bits where it is determined if next bit of the sequence is either a “0” or “1”.
The most likely path through the trellis is determined using a *branch metric*, $d(r|s)$ for a received sequence *r* given a known noiseless sequence *s*. The metric calculates the likelihood for

each branch in a trellis. The branch chosen is the one that has the maximum likelihood.

When the conditional probability distributions are modeled as gaussians with signal independent noise (cf. Equation 6), it is convenient to work with a log likelihood to remove the exponential from the expression. Noting that the variance is the same for both distributions when the noise is signal independent, we can then define a Euclidean distance metric using the following expression

$$d(r|s) = \Lambda(r|s) = (r - s)^2 = s^2 - r \cdot s + r^2 \tag{19}$$

where *r* is the measured received sequence and *s* is the expected sequence. This metric is a direct result from Equation 6. The exponential is dropped as the log-likelihood is computed. Each branch carries the last term r^2. The term can be dropped from the calculation since each branch is compared. The term s^2 may also be dropped in the case where every symbol has the same energy.

The expected signal sequence *s* is sometimes called the codeword and it must be calculated prior to the sequence estimation. Each codeword was constructed from the convolution of the measured pulse shape and the target sequence. The use of the convolution to construct the codeword assumes a linear time invariant channel. Figure 23 and Figure 24 are two examples computed for the 25G dataset from the pulse in Figure 15. The two figures are each five symbols long and assume the constraint length is five. For a constraint length of five, there are 32 codewords calculated. This means that the branch metrics for 32 possible 5-bit sequences need to be evaluated. The path(s) with the largest branch metric are saved and the other branches are discarded. When all surviving paths originate from a single symbol, that symbol is the most

likely symbol for that state of the trellis.

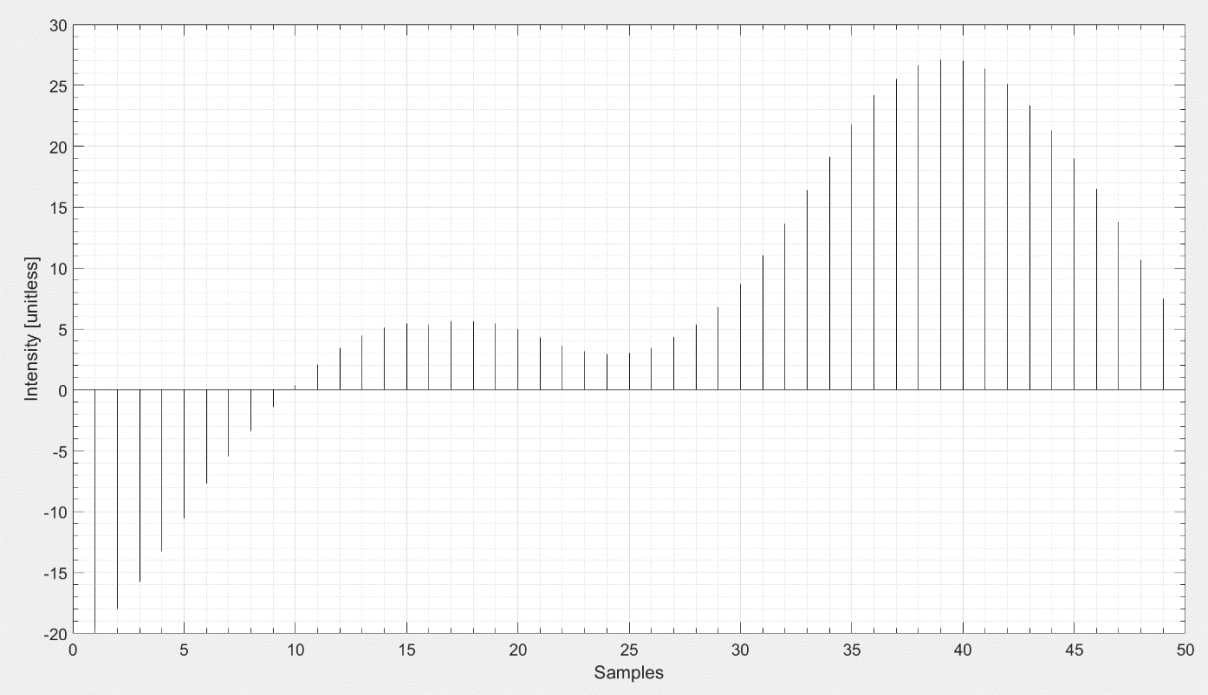

Figure 23: The 01011 codeword for the unpackaged APD.

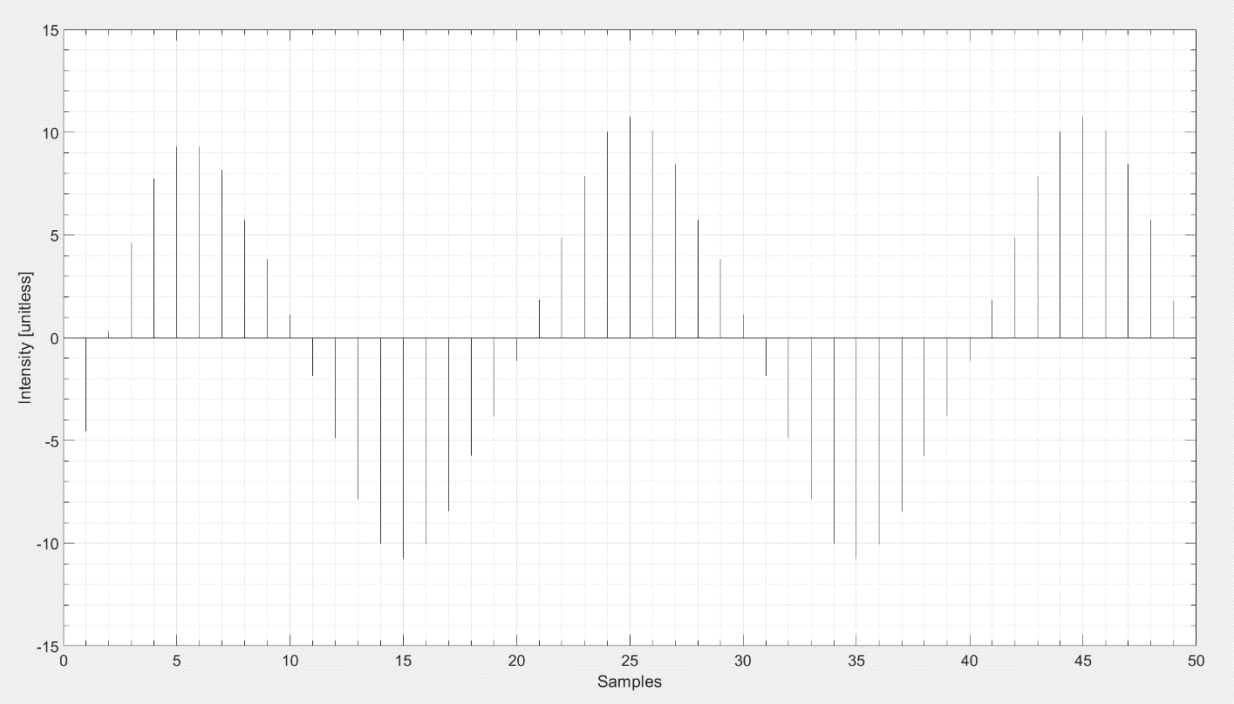

Figure 24: The 10101 codeword for the unpackaged APD.

Longer constraint lengths increase the computational cost to run the sequence estimator. The memory cost also increases as the set of codewords increases. It is possible to run the sequence estimator with arbitrarily long sequences. However, both computation and memory costs increase as the power of number of symbols. For simple binary systems, the complexity goes as 2^{ν}. It is engineering practice to run the sequence estimator with the shortest constraint length that achieves desired BER. Analyzing Figure 15 again reveals that it may be possible to run the sequence estimator with a constraint length of three as most of the pulse's energy was constrained within 30 samples or three symbol lengths. As such, the simple Euclidean distance metric was adequate to process the 10G data, but it was not adequate to process the 25G dataset. The 10G data was processed with a constraint length of three which had eight codewords (or states).

The metric used for the initial analysis assumptions that all sequences have the same energy. However, the probability distribution for the APD is modeled as a skewed gaussian as discussed in the background. Ideally, the skewness is sufficiently small so that the response was approximately gaussian. This assumption failed for the 25G data, but not the 10G data.

For the 25 G data, we adopted a state-dependent metric that included the effect of different energy distributions for different sequences.

$$d(r|s) = \Lambda(r|s) = \frac{(r-s)^2}{2\,\sigma^2} + \ln \sigma \tag{20}$$

where σ is the standard deviation of the target sequence extracted from the data and can vary from sequence to sequence. The state-dependent value of σ is calculated for each codeword and

can be calculated from the data shown in Figure 12 and Figure 13. However, in contrast to Equation 19, where the constant value of σ can be dropped, the state-dependent noise term σ is not a constant and must be included in the metric. The codewords are the same as the previous case and are the same as seen in Figure 23 and Figure 24. The extra step is to generate the state-dependent standard deviation from the data. The result is that each codeword is now weighted by a value representing the noise in that sequence instead of have the same noise irrespective of the sequence. Two statistical models were used to model the noise statistics: gaussian and poissonian. The mean of the gaussian was zero. The variance for both distributions was determined from histograms shown in Figure 12, Figure 13, and the other 30 codewords.

The state-dependent branch metric is a more complicated model because it is non-linear, but it still relies on a time-invariant channel. The unpackaged APD has parasitics and reflections as the probe was not well matched. Accordingly, the time-invariant assumption does not hold.

Our analysis of the unpackaged 25 G data showed that the test set-up used for the APD was *both* time-variant and non-linear. For this severe kind of impairment, instead of constructing the noiseless time-invariant sequence response by convolving the pulse response with the noiseless sequence, we directly measured the sequences from the data itself. While this method is memory and computationally intensive, it captures both the time-varying nature of the system as well as the nonlinearity. Codewords are generated directly from the dataset. Each sequence was determined using a cross correlation between the target sequence and the data. The target sequence is averaged over several frames. Figure 25 and Figure 26 are two examples out of 32 of this process. They have the same sequences as the calculated codewords in Figure 23 and Figure 24. The “full” sequence codewords are fed into the same algorithm as before. Comparing Figure 23 and Figure 26, the calculated codeword using the pulse response is symmetrical. The directly

measured codeword is not. The channel and unpackaged APD are clearly not time-invariant. The measured codeword may look clean, but it is averaged and has the additive noise removed.

All MLSE methods were sensitive to normalization. Naturally, the cross terms in the metric calculation need to be on the same order of magnitude lest they are drowned by noise. Normalization is typically a simple step. However, when processing the burst mode data from the unpackaged devices, the unwieldy time-varying drift, which was likely caused by the unmatched probe, was difficult to compensate for in practice.

A dynamic normalization was attempted in the metric calculation. Each branch was made to have the same energy per bit based on the projected bits in the codeword. Each branch then had different normalization constants. Including this time varying normalization in addition to all the other methods attempted on the data, the best BER achieved on the 25G data was 0.41, which is only slightly better than randomly guessing the bits. Our results demonstrate the difficulty of compensating a set of severe impairments.

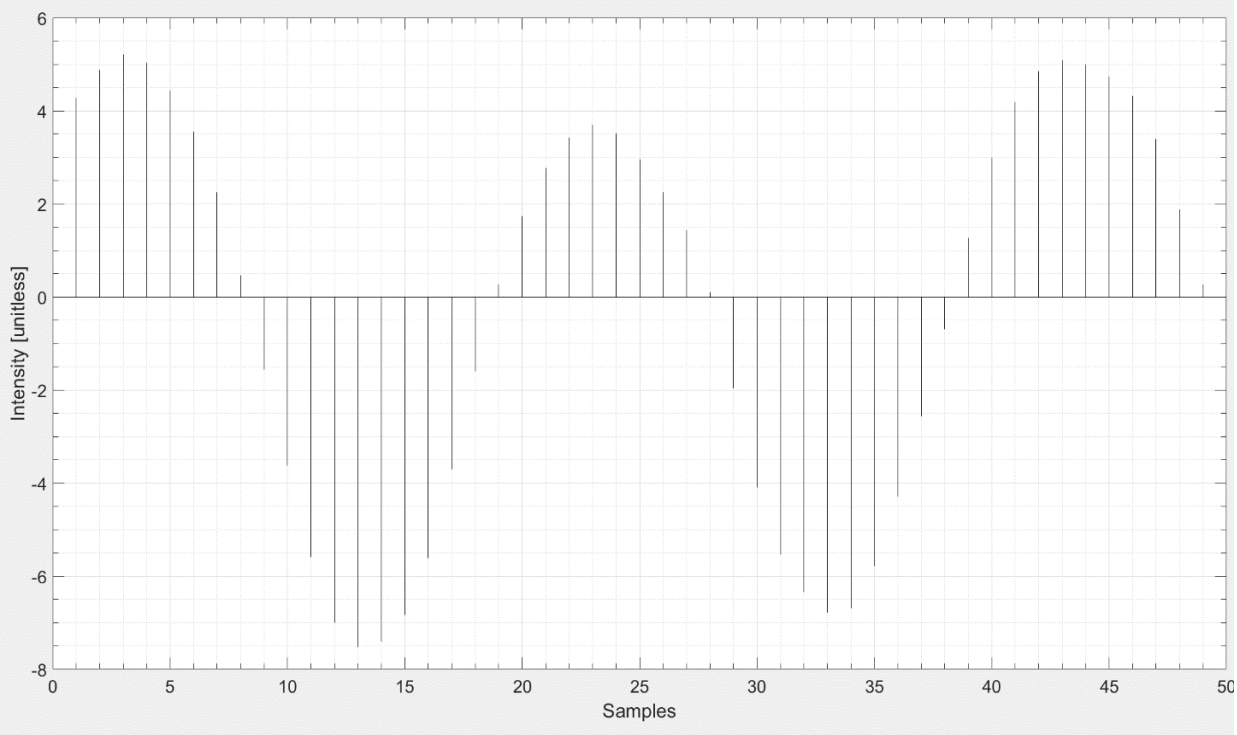

Figure 25: The 10101-sequence determined from the 25G dataset.

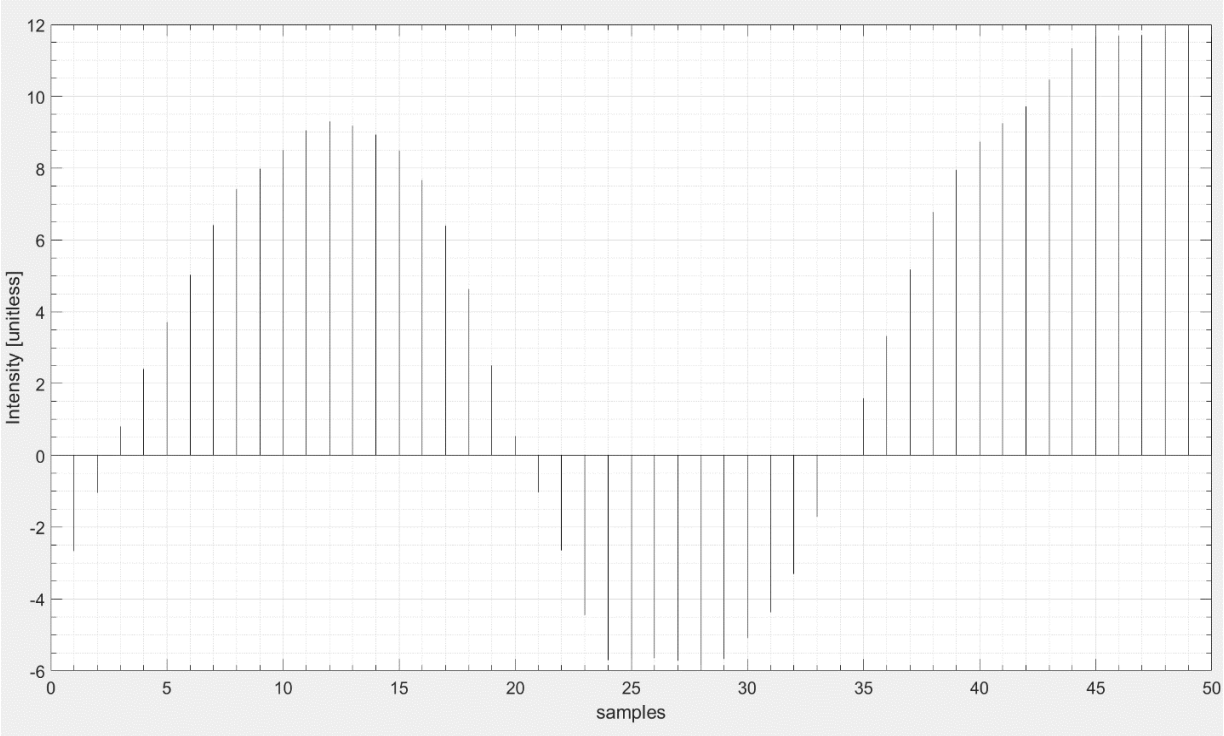

Figure 26: The 01011-sequence pulled from the 25G dataset.

PACKAGED APD

The 23G data from the packaged APD was much easier to process than the 25G data from the probed APD. In fact, bit-by-bit detection is adequate for processing the 23G data for high bias voltages for a BER of better than 10^{-5}. Accordingly, our research interest in this data set was to determine if sequence estimation can improve the performance over bit-by-bit detection since the techniques have already been developed when the device is not optimally biased.

The results of using MLSE on the 23G data are in Table 2 and Table 3. The baseline for comparison is bit-by-bit detection facilitated by the integrate and dump algorithm. Both tables include the integrate and dump performance denoted "intDump". The first attempt at MLSE uses codewords constructed from rectangular pulses for simplicity. The first column is the number of possible states which is the number of symbols to the constraint length 2^{ν}. The bit-by-bit BER performance achieved 10^{-5} before any more sophisticated analysis was used. If MLSE could do better, more than 10^{6} bits were used in the experiment. The dataset, as seen in Figure 17, was a $N = 2^{15} - 1$ PBRS sent over the channel that was repeated 40 times. The "numBits" column in the table represents this fact. As evident from the table, MLSE did not improve the performance of the system when using codewords generated from rectangular pulses.

Table 2: BER results from MLSE on the 23G data using square pulse shape codewords.

square codewords			
Method	Errors	numBits	BER
IntDump	695	40*N	5.30E-04
2 state	694	40*N	5.30E-04
4 state	695	40*N	5.30E-04
8 state	695	40*N	5.30E-04
16 state	695	40*N	5.30E-04
32 state	695	40*N	5.30E-04

Table 3: BER results from MLSE on the 23G data using pulse convolved shape codewords.

convolved pulse shape codewords			
Method	Errors	numBits	BER
IntDump	12	5*N	7.32E-05
2 state	4	5*N	2.44E-05
4 state	96	5*N	5.86E-04
8 state	45	5*N	2.75E-04
16 state	189	5*N	1.20E-03
32 state	67	5*N	4.09E-04

The use of rectangular pulses for the codewords did not improve the performance. So, the next step is to increase the complexity as was done with the 25G data. The convolved pulse shape codewords are generated by convolving the measured pulse shape from Figure 21 with the target sequence. Again, the number of target sequence or states is the first column of the table. Some of the codewords are presented in Figure 27, Figure 28, and Figure 29 as examples.

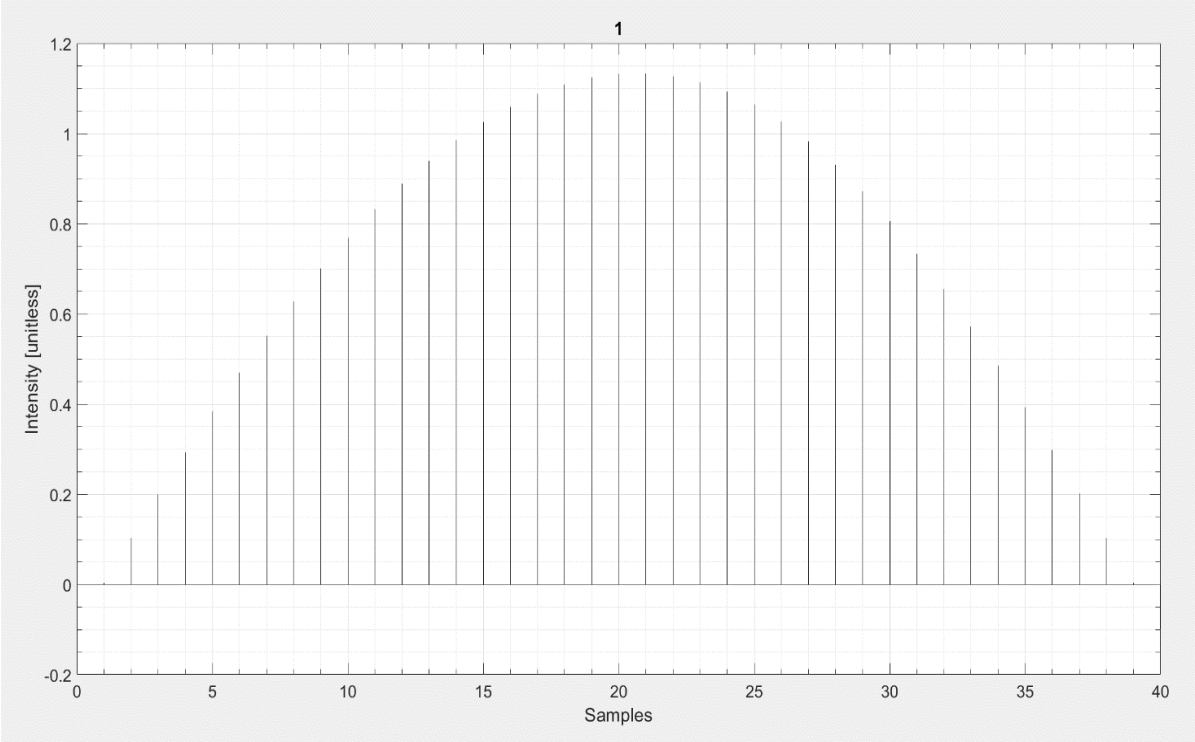

Figure 27: "1" codeword for the 2 state MLSE.

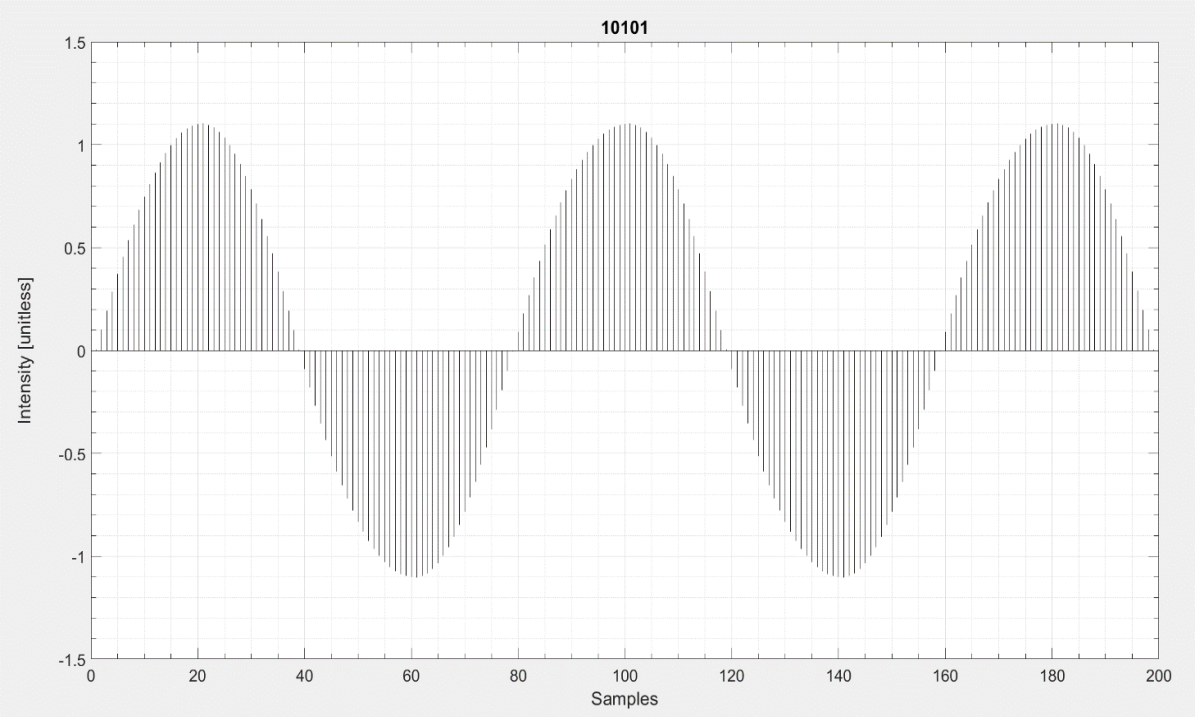

Figure 28: "10101" Codeword for the 32 MLSE.

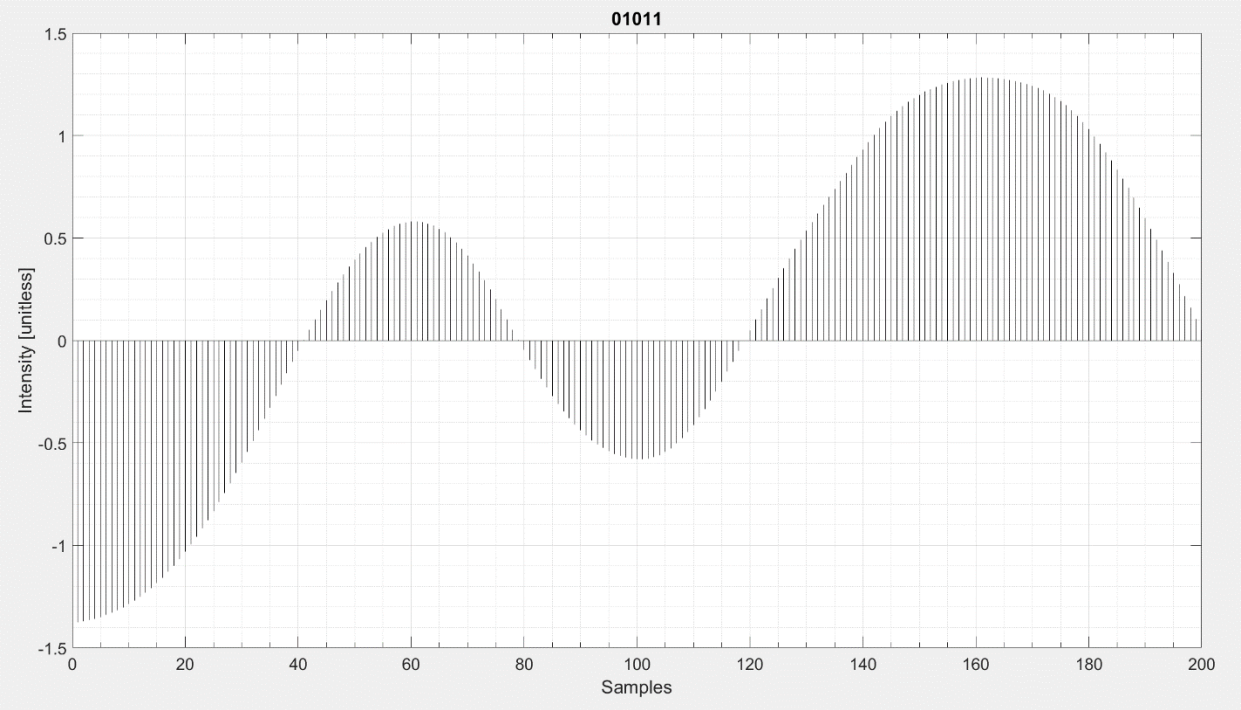

Figure 29: "01011" codeword for the 32 state MLSE.

Now that the codewords are computed, the resulting BER performance in Table 3 does not show an improvement over bit-by-bit. In fact, it is worse in most cases. The rest of the analysis will focus on an explanation for this result.

The results of Table 3 also run over a smaller data set compared to the previous table. There are two reasons for this. MLSE is computer resource intensive. The bit-by-bit MATLAB code requires seconds to process while the MLSE took an hour to process 40*N bits. The second is if the MLSE does not perform better than bit-by-bit then less bits are required to build the statistics needed for the algorithm comparisons.

The first conclusion from the table is that bit-by-bit detection performs better over a smaller number of bits. Table 4 investigates this further. The first and last rows are repeated from Table 2 and Table 3.

Table 4: Effect of number of bits on the BER

Bit-by-bit	
Num bit processed	BER
5*N	7.32E-05
10*N	1.13E-04
20*N	1.13E-04
40*N	5.30E-04

The shorter sets of data produce less errors. This effect is caused from some uncompensated drift which was already hinted at in Figure 18. The normalization procedure fails for longer data sets, and this is why the shorter length produces fewer errors. An attempt to correct the drift was made by normalizing across smaller windows. The data presented was for a window size of 128 bits. A thousand-bit window does increase the performance over a window that is as long as the entire dataset, but any shorter length has no effect. The normalization was done in simple blocks. A more sophisticated normalization procedure could perhaps correct for the drift or a more careful analysis of the physical hardware could be made to extract more systematic errors from the experiment.

Looking again at Figure 21, the full width at half maximum pulse width is 50 samples. The transmitted pulse was a square pulse 40 samples wide. The ISI of the packaged only increased the pulse width by one quarter of a symbol interval. The memory of the packaged APD has a constraint length less than two, while the probed APD had memory that lasted for five symbol intervals. The ISI was small enough where the MLSE does not provide significant performance benefit for the packaged device.

There was one instance in which the sequence estimation did improve performance. The 2 state MLSE had fewer errors than the bit-by-bit detection. In the case with the square pulse

codewords, the 2 state MLSE reduces to bit-by-bit detection. This is not the case for the pulse convolved codewords as that technique accounts for the pulse response. The “1” codeword is displayed in Figure 27.

There are two reasons for this improvement. The memory of the channel is less than two symbol intervals. Therefore, if there was an increase in performance, it would be for a constraint length of two and more states would not increase performance. The second reason is that the codeword in Figure 27 acts as a time-domain filter.

Chapter 6 CONCLUSION

The overall goal of the APD project was to increase link margin so that an optical switch could be used without the need for an optical amplifier. This, in turn, would decrease the energy costs for data centers. The gain provided by the APD comes at the cost of ISI and additional complexities in processing burst mode data.

The APD experiment occurred in two configurations. The unpackaged, probed APD experiment was run at 10 Gbps and 25 Gbps. The packaged APD was the same device but with special capacitors and packaging to eliminate the impedance mismatch of the probed APD. [8]

The unpackaged APD presented a challenging dataset to process at 25 Gbps but worked quite well at 10 Gb/s. The powerful method of MLSE was able to process the data from the unpackaged APD at 10 Gbps but not at 25 Gbps. The pulse shape was much longer than a symbol interval at 25G. In addition to the ISI, there was a time-invariant response in the unpackaged APD. Many techniques including dynamic normalization were attempted to compensate for this. However, the overall time-varying and nonlinear impairments in the unpackaged data were too difficult to overcome by the methods used in this thesis.

REFERENCES

[1] G. C. Papen and R. E. Blahut, Lightwave Communications, New York: Cambridge University Press, 2019.

[2] B. Sklar, "How I learned to love the trellis," *IEEE Signal Processing Magazine,* vol. 20, no. 3, pp. 87-102, 2003.

[3] T. Lentine, *Private Communication,* Albuquerque: Sandia National Lab, 2021.

[4] J. Campbell, "Recent Advances in Avalanche Photodiodes," *Journal of Lightwave Technology,* vol. 34, no. 2, pp. 278-285, 2016.

[5] J. Campbell, "Evolution of Low-Noise Avalanche Photodetectors," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 28, no. 2, pp. 1-11, 2022.

[6] J. S. C. J. P. R. D. a. M. M. H. Erum Jamil, "On the analytical formulation of excess noise in avalanche photodiodes with dead space," *Optics Express,* vol. 24, no. 19, pp. 21597-21608, 2016.

[7] "resample documentation," The MathWorks, Inc, 24 06 2024. [Online]. Available: https://www.mathworks.com/help/signal/ref/resample.html. [Accessed 24 06 2024].

[8] M. Geln, *Private Communication,* Allbuquerque: Sandia National Labs, 2024.

[9] R. McIntyre, "The distribution of gains in uniformly multiplying avalanche photodiodes: Theory," *IEEE Transactions on Electron Devices,* vol. 19, no. 6, pp. 703-713, 1972.