MULTI-CARRIER SIGNALING FOR HIGH-SPEED ELECTRICAL LINKS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Amir Amirkhany

March 2008

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Mark A. Horowitz) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(John M. Cioffi)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

_____

(Bruce A. Wooley)

Approved for the University Committee on Graduate Studies

Abstract

The demand for ever higher performance at lower power is motivating system designers to re-think their design strategies not just in terms of performance, but in terms of power-performance efficiency. In this new design paradigm, the performance of a communication system is mainly limited by the "total system power" constraint rather than just the "transmit power". Therefore, the choice of the optimum data communication algorithm is a strong function of circuit level power-performance trade-offs. One area where such re-thinking is vital is electrical chip-to-chip communication, where total system power-performance efficiency in terms of mWatts per Gb/s is now the key metric. Multi Gb/s chip-to-chip links find applications in the data interfaces between microprocessors, memories, peripherals, and network processing components in high performance systems.

This thesis covers the design, analysis and implementation of a multi-carrier signaling method, called Analog Multi-Tone (AMT), specifically designed to take advantage of the characteristics of chip-to-chip link systems. In our design approach power-performance efficiency is achieved through a combination of various techniques including better utilization of transmit power, parallelization of the data stream in the frequency domain, channel engineering, in-system characterization of circuits with relevant performance metrics, and eventually, careful circuit design.

We start with a study of the performance and complexity of conventional multi-tone techniques. Based on the insight obtained from this analysis, we propose a novel multi-carrier signaling technique called Analog Multi-Tone, which is customized to the link characteristics. We develop a mathematical analysis of this system including a convex framework, and closed-form jitter modeling. This analysis also provides a

method of comparing the performance of this approach with alternative baseband transmission methods.

The second half of this thesis focuses on circuit design issues involved in the design of a transmitter in a 90-nm CMOS technology supporting Analog Multi-Tone as well as a variety of baseband signaling modes including 4-PAM, 16-PAM and 64-PAM, all at 24-Gb/s. We further propose a Least-Squares based characterization and digital compensation techniques for improving system performance.

Acknowledgments

Of all the parts in this dissertation, the acknowledgement is the part I have reviewed in my mind many times during my past six years at Stanford. Yet, now that I am starting to write, I find it difficult to summarize my gratitude for several years of help and support in only a few paragraphs. Either the Ph.D. program is just too long, or I have been very lucky to have had so many great people around me helping me out with different aspects of my life at Stanford.

I am indebted to my advisor, Professor Horowitz. Mark has a reputation for being a great advisor, and I can only be another testimony to this fact. Mark takes every aspect of his advising role seriously. On the technical side, he is always a great person for analyzing ideas and finding their faults. On non-technical issues, I always find him concerned and available for help. Mark is also a great teacher. What I really like about his approach is his way of training students to break down hard problems to simple fundamental concepts, and I am particularly grateful to him for that. I would only hope to live up to his standards as an advisor and a teacher through the rest of my career.

My associate advisor, Prof. Cioffi, has always been kind to me with any questions I have had and I have also benefited a lot from taking his courses. His approach to teaching communications is very unique in being very practical. A good part of my Ph.D. work relies on what I learned from his courses.

I would also like to thank Professor Wooley, for taking the time to read this manuscript and provide feedback, and Professor Pease, for making it possible for me to study together with my wife on this side of the Bay by offering me a one year fellowship after I was admitted to both Stanford and Berkeley.

I was fortunate to have Rambus support my Ph.D. project. I received a lot of help from many individuals in Rambus during the past four years. None of that would have been possible without Rambus' willingness to invest in my work, and create a team to move forward with implementation aspects of the project.

In the past few years, I have particularly benefited a lot from discussions with two individuals: Dr. Ali Abbasfar of Rambus and Prof. Vladimir Stojanovic of MIT. Ali is a great person for brain storming and discussing different ideas. He has a strong intuition and is a valuable colleague for any research work. Vladimir was the first person who helped me with my research at Stanford even before I got into Mark's group. I am grateful to him for many hours of fascinating discussions and collaborative work, and for both acting as a great mentor and a friend.

The rest of our team at Rambus who worked on the implementation of the transmitter, Bruno Garlepp, our manager, Dr. Jafar Savoj, who designed the digital to analog converter, and Metha Jeeradit, who helped me with the design of the equalizer, have all been great colleagues and friends. I have benefited and learned a lot from our collaboration.

I am also thankful to several other individuals in Rambus: Ravi Kollipara who was my main reference for signal integrity issues, Jared Zerbe who first advocated my project in Rambus, and Joe Louis Chandran and Kumar Gogenini for help with CAD issues.

Many of my colleagues in Mark's group helped me a lot by providing valuable feedback and spending the time to help me out within their areas of expertise. Especially, I would like to first thank my wife, Bita Nezamfar, for being available for technical discussions 24 hours a day, seven days a week! I would also like to especially thank Prof. Elad Alon and Dr. Dinesh Patil for many good suggestions and interesting ideas.

People who need to speak with Mark on a regular basis generally find it very difficult without the help of our kind group admin, Teresa Lynn. Her constant

readiness to make an extra effort to make things smooth and easy deserves a big thank you!

My years at Stanford wouldn't have been as fun without the large community of Persian friends and colleagues. I feel privileged to have so many close friends while being so far away from home.

I have been fortunate to have my wife, Bita, with me all along as a wonderful companion, colleague and friend. It has been really fun living together and studying together, and I can only wish her the best with her own endeavors. I am also very grateful to Bita's family for being extremely kind to me in the past 7 years, for always sharing our stress and happiness, and offering their kindest prayers and support.

It certainly wouldn't have been possible for me to get this far without the sacrifices of my parents, Abbas and Mehri. This dissertation is a result of their 30 years of infinite love and support, and for that, they own all my success. This dissertation is only dedicated to them as a token of thankfulness and gratitude. I am also very thankful to my brother Arash for his constant love, and for filling in for me at home while I have been away.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The increasing demand for higher memory bandwidth, the emergence of powerful game consoles, and the adoption of distributed computation and storage have driven demand for multi gigabits per second data interfaces. Figure 1.1 shows example applications of these types of links, including communication between a linecard and a switch in a network router (Figure 1.1(b)), multi-drop memory bus (Figure 1.1(c)), and between the central and the graphics processing units (CPU and GPU) inside a PC. Modern high-speed electrical links in the heart of such interfaces enable data transfer between integrated circuits.

The electrical signals in all these links travel through a channel that consists of the chip packages and copper traces on printed circuit boards (PCB). These signals are distorted by high-frequency loss of the wires, as well as by reflections from the impedance discontinuities in the signal path. The discontinuities are usually caused by the vias, stubs, and connectors in the signal path that were necessary to route the signal traces across the multiple PCB wire layers and sometimes multiple boards. Signal reflections caused by these discontinuities create resonance frequencies and result in notches in the frequency response of the link channels as shown in Figure 1.1(a). In order to compensate for the dispersive nature of the communication channel, most high-speed links require at least some form of signal processing at the transmitter or at the receiver. The design of the required correction circuits is challenging because of both the extremely high speed requirements and the tight power-consumption constraints of the link system.

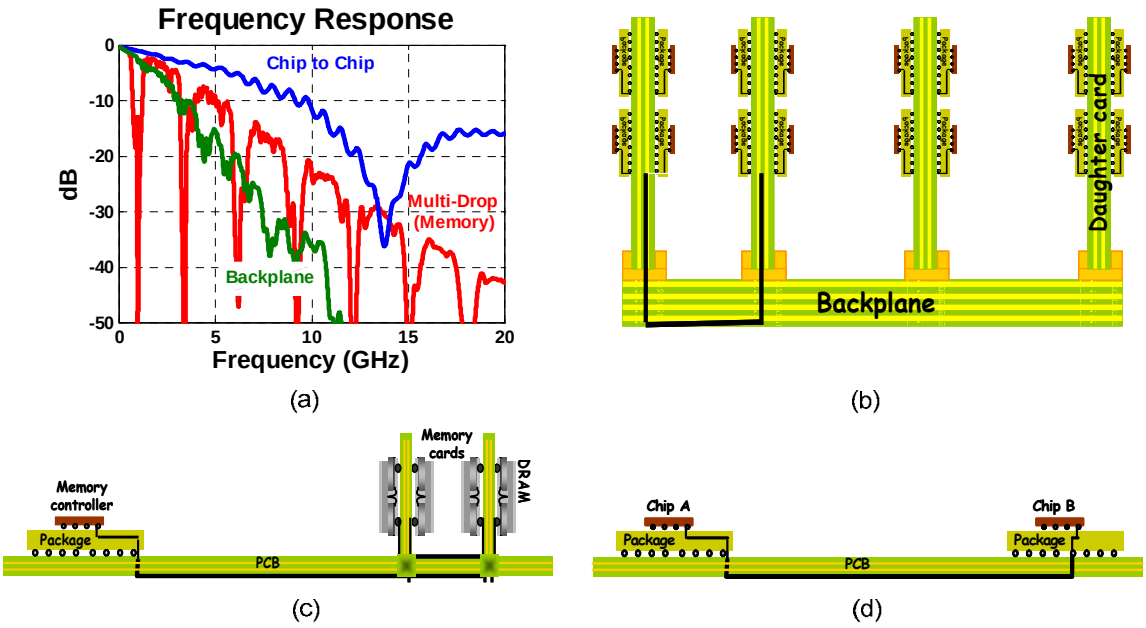
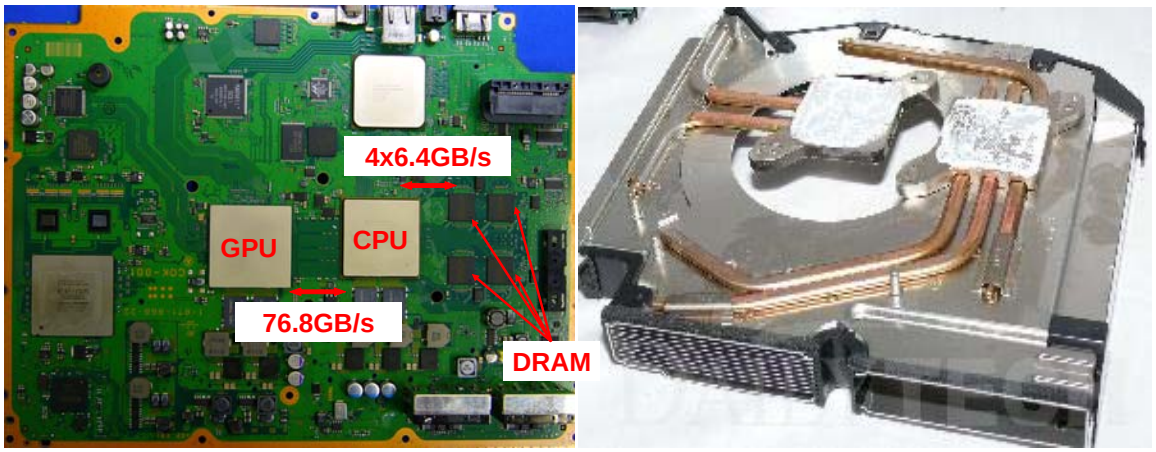


Figure 1.1: (a) Channel characteristics in frequency domain. (b) An electrical link in a network router (c) A multi-drop memory interface (d) A CPU-GPU link

Figure 1.2(a) shows the high-speed links in the main-board of PlayStation 3®. In this application, high-speed data is transferred between the CPU and the GPU and the DRAM memory modules through multiple PCB traces. The signaling rate on the individual PCB traces is 6.4-Gb/s between the CPU and the GPU [1] and 3.2-Gb/s between the CPU and the DRAM [2]. A picture of the heat-sink in PlayStation 3 is also shown in Figure 1.2(b). The elaborate design of the heat-sink is a clear indication of the importance of the system level power-constraints. A simple calculation on the PlayStation 3 system gives a sense of the link requirements. Assuming that 10% of the 100W power budget is allocated for data transfer between the CPU and the two peripherals, split evenly between the two, the DRAM interface can only dissipate 5W to transfer 25.6-GB/s of data. In other words, the system is allowed to dissipate less than 25mW per every Gb/s of data transferred. Compared to an ADSL system, for example, where 12-Mb/s of data is transferred at power consumption of more than 100mW, this is an almost three orders of magnitude tighter constraint. As a consequence of this huge difference, the communication algorithms currently employed in the state-of-the-art high-speed links are much simpler than those used in

DSL systems; the majority of link systems today are limited to simple NRZ signaling with limited linear feed-forward (FF) and feed-back (FB) equalization to compensate for the dispersive nature of the communication channel [3][4].



(a) (b)

Figure 1.2: (a) PlayStation 3® main-board and (b) heat-sink.

From communication system perspective, another major difference between high-speed link systems and Digital Subscriber Line (DSL) or wireless systems is that system performance can be increased by adding extra channels (pins and traces) to the system, or by improving channel quality with better material or fabrication processes. Even though this approach may lead to increased system cost, improving performance through more sophisticated signal processing over available communication channel capacity also costs added power consumption and silicon area. Given the importance of power, a metric that is used for characterizing high-speed links is energy-efficiency. This is usually defined as the total system power dissipated in the entire link per every Gb/s of data transferred, or mW/Gbps. High-end commercial links today generally operate at data transfer rates of 6-Gb/s to 12.5-Gb/s and energy-efficiencies of less than 30-mW/Gbps, while prototype systems have achieved data rates of 20-Gb/s [5] or power efficiencies of 2.2-mW/Gbps [4].

A study of the limits of signaling in backplane links [6], reveals that there is a large gap between the fundamental limits of signaling, commonly known as the

Shannon Capacity, and the limits achievable with baseband (BB) signaling techniques. In particular, inspection of the channel characteristics shown in Figure 1.1 reveals that, for a class of applications, notches in the frequency domain are part of the frequency response of the link channels. It is well known in communication theory that multi-tone (MT) signaling has the potential to achieve superior performance over such channels compared to BB signaling by allocating transmit energy away from the notch frequencies. MT signaling can therefore potentially be employed to reduce the gap between current link performance and the Shannon capacity in these applications. However, as described earlier, the real challenge for these links is the design of a MT algorithm that is energy efficient. It must optimize the total energy consumption of the system, including both the energy transmitted over the channel and the energy spent to implement the MT processing.

The first MT techniques that were studied in the communication literature were analog in nature. An analog transmission system with overlapping orthogonal filters and offset quadrature amplitude modulation (QAM) was first proposed by Saltzberg [7]. However, the performance of the system was found to be very sensitive to communication channel variations which affected the orthogonality between the system sub-channels. A digital implementation of a MT system based on Discrete Fourier Transform (DFT) was first proposed by Weinstein [8]. A digital implementation of the system proposed by Saltzberg was also proposed by Hirosaki [9], where a combination of poly-phase filtering and DFT was employed to perform the necessary filtering and mixing. Weinstein's approach is known today as Orthogonal Frequency Division Multiplexing (OFDM) or Discrete Multi-Tone (DMT), and Hirosaki's approach became the basis for what is now called Filtered Multi-Tone [10]. Such digital MT techniques leverage the power efficiency of digital circuits to perform substantial signal processing at the transmitter and at the receiver. In addition, almost all conventional MT techniques today rely on power-efficient, fairly high-resolution Analog to Digital Converters (ADCs) for digitizing the received signal at the receiver front-end. However, such digital signal processing blocks and ADC modules, if implemented at multi-GHz operating rates, can substantially add to

the power-consumption of a link system [12][13]. This thesis studies different MT techniques with the goal of finding a MT scheme which can be energy-efficient at the Multi-GHz operating rates necessary for link applications, while providing significant performance improvement over existing BB transmission techniques.

1.1 Organization

We start our analysis in Chapter 2 by studying the requirements for a conventional DMT system in terms of block-size and ADC resolution and speed in a realistic link environment. The goal of this study is to estimate the projected data rates and complexity of a DMT system. The tight link system power constraints severely limit the interference cancellation capabilities of any practical DMT system, generally performed through a long cyclic prefix or a channel shortening filter. As a result, an optimized DMT system for links is interference limited in most cases and the conventional water-filling algorithms and analysis frameworks for wireline DMT systems do not apply. In this chapter we propose a new analysis and integer bit-loading approach based on convex optimization. The analysis framework enables us to obtain an accurate estimate of the complexity of a DMT system. The result of this study also leads to insights into the characteristics of a MT architecture that can potentially be power-efficient in a link environment.

Using the results of our analysis on a DMT system, in Chapter 2 we propose an MT architecture, called Analog Multi-Tone (AMT), for high-speed electrical links, which is customized to the link characteristics, and is therefore power-efficient in a link environment. The AMT system consists of only a small number of wideband sub-channels, and employs linear transmit equalization and receive Decision Feedback Equalization (DFE) to compensate for the effects of the frequency selective sub-channels and the interference among them. AMT will be analyzed from performance and complexity perspectives and will be compared with traditional BB techniques.

With certain assumptions about the communication channel, other variations of AMT are possible which can lead to better performance or lower system complexity.

In Chapter 4 we specifically propose a duobinary AMT architecture useful for partial-response channels, and a sampler-based AMT architecture which does not require mixers and integrators, suitable for complexity-limited applications.

Having investigated the potential of AMT, we describe the architecture of a 24-Gb/s software programmable transmitter which supports both AMT and BB transmission in Chapter 5. The transmitter consists of a digital equalizer and an 8-bit 12-GS/s Digital to Analog Converter (DAC), and the transmission mode is set simply by the way the 64 10-bit equalizer coefficients in the system are programmed. The unified architecture of the transmitter allows for the choice of the best transmission algorithm based on the channel characteristics. In addition, with sufficient flexibility and signal processing capabilities, the transmitter is an ideal platform to enable evaluation of different transmission algorithms in different environments. The digital equalizer in the transmitter was implemented using an automated flow including many ASIC design tools. The flow is described in Appendix C.

A further step towards the implementation of efficient link systems is efficient circuit design, and proper characterization methodologies are important enabling tools in this regard. The 12-GS/s DAC in the prototype transmitter is a wideband circuit, and must be characterized in this context. One such characterization methodology is proposed in Chapter 6. Unlike the standard narrowband approach, our proposed technique employs wideband inputs to characterize the circuit from dc to any target frequency. We extend the technique to cyclically time-variant (CTV) systems which makes it applicable to a wide variety of circuits including time-interleaved data converters and mixers, and consequently to characterizing an entire communication system. We apply our method to the time-interleaved DAC included in the prototype transmitter to capture its time-varying behavior. We further demonstrate how the insights obtained from the characterization enable us to digitally compensate for the time-varying nature of the DAC using hardware already available in the transmitter.

Some additional material related to AMT that are not presented in the thesis are covered in the appendices. In particular, a current-mode implementation of the AMT transmitter, as a low-power alternative to the digital implementation of the prototype

system, is proposed in Appendix A, and a Clock and Recovery Recovery (CDR) algorithm suitable for AMT is suggested in Appendix B.

1.2 Contemporary BB Links

To better understand the constraints under which our MT solution will need to operate, this chapter ends with a quick overview of the techniques used in the state-of-the-art BB links, as shown in Figure 1.3. This link uses 2-PAM signaling with a discrete linear FF equalizer at the transmitter to cancel pre-cursor Inter-Symbol Interference (ISI), a linear peaking amplifier at the receiver front-end to increase sensitivity and compensate for the magnitude loss of the channel, and a DFE at the receiver to cancel post-cursor ISI[1]. This architecture is used since a linear FF transmit equalizer and a DFE do not require additional analog to digital conversion because they take the binary input and output data streams as their inputs. Due to the same power-consumption limits and the sensitivity of some link applications to added data latency, error correction and detection coding is also generally avoided in high-speed links. As a result, systems are generally directly designed for very low Bit Error Rates (BER) of 10^{-12} to 10^{-18}.

From communications systems perspective, in addition to the BER constraint, link systems are constrained by a peak-transmit-voltage constraint. The peak-voltage-constraint is analogous to the average-transmit-energy constraint in classic communications systems. The peak voltage constraint is more relevant constraint because due to the absence of error correction or detection coding in links, clipping of the output signal is avoided. Therefore, transmission algorithms have to be designed such that under worst scenarios, all the circuits at the output stage of the transmitter remain in their intended region of operation. Typical peak transmit voltages in CMOS

[1] Recently, prototype links that additionally utilize a low-resolution Analog-to-Digital Converter (ADC) and a digital linear FF equalizer at the receiver have been demonstrated [14]. Discrete linear equalizers at the receiver are generally avoided in high-speed links due to the power-inefficiency of high-speed ADC's.

technology are below $1.6V_{pp}$ in today's high-speed links utilizing differential signaling.

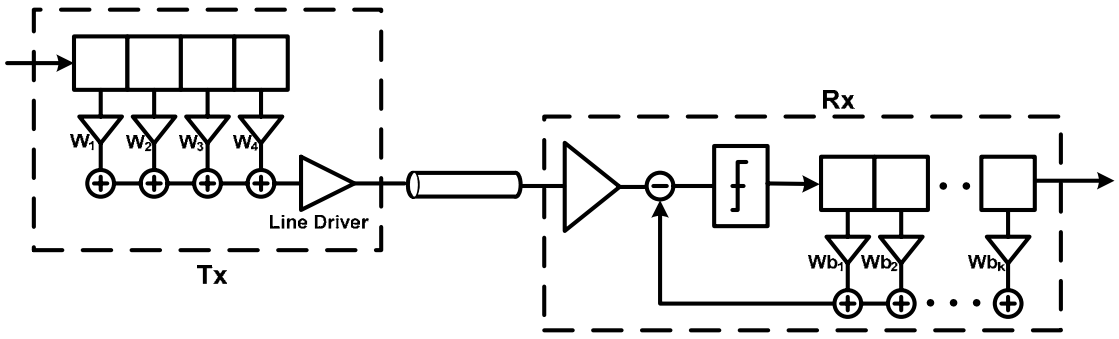


Figure 1.3: A conventional state of the art 2-PAM BB system with a linear FF equalizer at the transmitter, and a linear peaking amplifier and a DFE at the receiver.

A circuit design challenge in increasing 2-PAM BB signaling rates under existing link channels is that increased data rate requires increased signaling rate, and consequently transmission over more attenuated regions of the frequency response of the channel. As a result, the BB receiver needs both to have wider bandwidth and better sensitivity to resolve smaller signals. Such requirements push on both ends of the gain-bandwidth trade-off for circuits, which (to first order) only scale linearly with technology scaling.

Another issue in implementing a 2-PAM BB receiver is closing the feedback loop for the first (few) DFE taps. The DFE has to subtract a weighted sum of the received symbols from the incoming signal. In particular, to close the feedback loop for the first tap, the entire operation of detecting the current symbol, multiplying it with the appropriate weight, and subtracting it from the incoming symbol should be performed in less than a symbol period (Figure 1.4(a)), which is 200ps in a 5 GS/s link. In order to alleviate this constraint, most 2-PAM BB links resort to loop-unrolling techniques [19][20], where two decisions are made for the next incoming bit assuming the current bit is +1 and -1. The correct decision is then chosen once the value of the current bit is known (Figure 1.4(b)). Therefore, a loop-unrolled DFE utilizes two samplers sampling at rate R with thresholds placed at $\pm\alpha$.

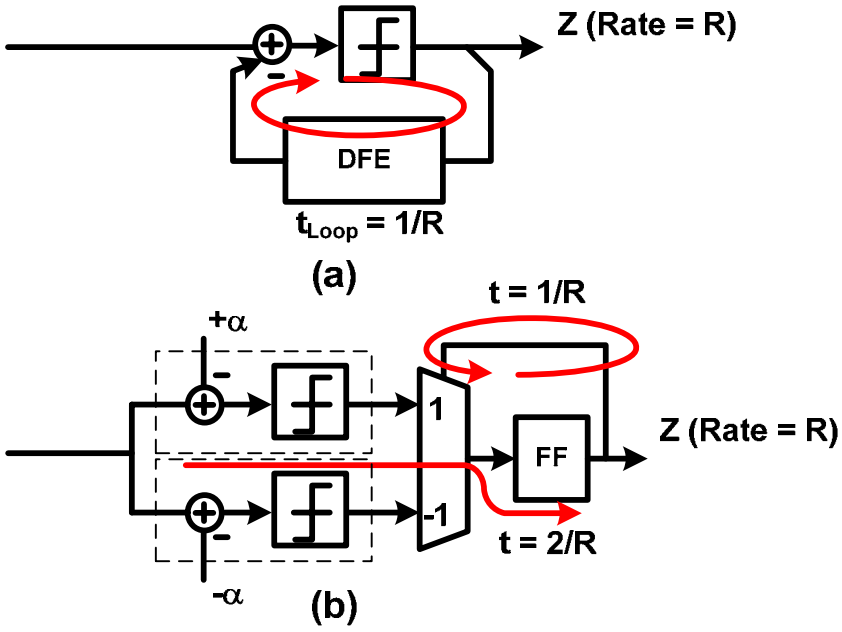


Figure 1.4: (a) DFE loop timing constraint for the first tap (b) A loop-unrolled DFE

The equalization techniques described so far targeted arbitrary forms of ISI. One approach to reduce the challenge with DFE is to shape the channel to have a well-known ISI pattern. This leads to more efficient "partial-response" signal processing techniques which have simpler transmitter and receiver architectures. Let's, for example, assume the ISI pattern in the communication channel is such that the previous bit gets added to the current bit as the signal travels through the channel. In other words, let's assume the received signal at time n (y_n) is given by

$$y_n = x_n + x_{n-1} \tag{1.1}$$

where x_n is the transmitted symbol at time n. Then for a 2-PAM system, the received signal y_n can either be 0 or 2 if x_n and x_{n-1} are equal, or it is 1 otherwise. Therefore, if we know what x_{n-1} is, we can find out what x_n is using regular DFE or loop-unrolled DFE. Alternatively, we may perform the following simple precoding at the transmitter before transmitting the signal:

$$y_n = y_{n-1} \oplus x_n \tag{1.2}$$

where $\oplus$ represents the XOR operation, and transmit y_n instead of x_n. By doing so, it is easy to verify that levels 2 and 0 at the receiver correspond to $x_n = 0$, and level 1 corresponds to $x_n = 1$ independent of the value of x_{n-1}. Therefore, the encoding on the transmitter eliminated the need for DFE at the receiver without increasing the transmitter voltage headroom requirements since the transmitted sequence still consists of 1 and 0. This signaling technique is called duobinary signaling and was first proposed by Lender [7]. If channel characteristics are not exactly as described above, a linear equalizer may precede the channel at the transmitter to equalize the channel to the duobinary ISI pattern. Consequently, in practical systems, duobinary signaling is generally used when channel characteristics are close to a duobinary channel. Duobinary signaling has already been demonstrated at 10-Gbps and beyond over a long FR4 backplane [22][23].

Another way to increase bit-rate is to increase the number of bits transmitted in each symbol. For a BB link, this means using multi-level modulation techniques. 4-PAM prototype systems, for example, have been demonstrated [15][16] to achieve this goal. However, their adoption in commercial systems has been hindered due to their higher complexity and to the limited performance improvement they provide. For a given transmitter voltage swing, the minimum distance between the 4-PAM constellation points (d_{min}) is $1/3^{rd}$ that of a 2-PAM system. Therefore, 4-PAM transmission only improves the receive voltage margin over 2-PAM if the 3x loss in the minimum distance is more than compensated by the reduced channel attenuation gained by lowering the symbol-rate by 2x. Timing margins of a 4-PAM system are also at least 33% smaller compared to a 2-PAM system operating at the same throughput due to the variation in reference cross points. In order to improve the 4-PAM timing margins, codes have been developed [17][18] to eliminate the major transitions from the highest 4-PAM level to the lowest level, at the cost of reduced throughput.

Since all of the baseband techniques become more difficult as bit rates scale, the next section evaluates the potential of a multi-tone approach to improve performance.

Chapter 2

Discrete Multi-Tone Requirements

In this chapter we study the application of Discrete Multi-Tone (DMT) to high-speed links. Our goal is to determine the required ADC resolution and speed, and DMT block-sizes needed to achieve reasonable performance.

Application of Discrete Multi-tone (DMT) signaling to high-speed electrical links requires major modifications to the well-known analysis methods applied to wireline communication systems. Tight power budgets in links impose severe constraints on how well DMT can cancel interference. Consequently, maximum throughput is achieved in a DMT system that is (residual) interference limited and the standard water-filling method is not applicable in its original form. As a result, a different analysis framework is necessary.

In order to address this issue, in this chapter, we cast the DMT system as a Second Order Conic (SOC) problem with peak-transmit-power as the constraint. In addition, we propose a novel incremental integer bit-loading[2] algorithm that complements the convex analysis framework to obtain close to optimum power allocation coefficients and sub-channel bit assignments.

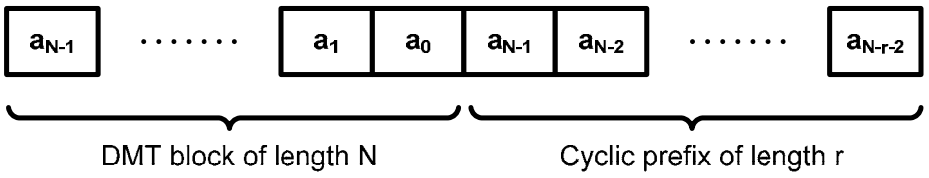
The minimum block-size and ADC requirements for a DMT system are clearly a function of the channel characteristics and the applications. For the analysis in this chapter we mainly focus on backplane links, which can afford less-power efficient

[2] In a DMT system, each sub-channel's constellation size can be different from the others. In other words, different channels can carry different number of information bits. The process of assigning bits to the DMT sub-channels is called bit-loading in the communication literature.

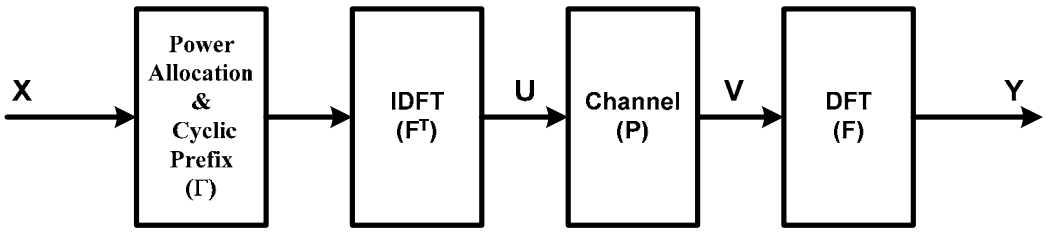
links to achieve a very high data-rate. However, we limit the design-space to the range of system configurations that might fall within our power budget - DMT block sizes of less than 128 and ADC resolutions of less than 8 bits. These assumptions will be justified towards the end of this section where we estimate the complexity of an FFT circuit with a block size of 64 and a 7.5 GS/s 6-bit ADC.

2.1 Discrete Multi-tone System Formulation

In DMT systems, data is broken into blocks and the blocks are transmitted sequentially over the channel. To prevent blocks from interfering with each other, two consecutive blocks are separated by a gap interval which is longer than the channel response (dispersion). Sometimes, no data is transmitted during this interval, and therefore, the received blocks are free of inter-block interference. A more computationally-efficient approach, however, is to transmit the trailing part of the data block during the gap interval, as shown in Figure 2.1(a).



(a)



(b)

Figure 2.1: (a) DMT frame consisting of a data block of size N and cyclic prefix of size r. (b) Block Diagram of a DMT system.

This redundant data sequence is called cyclic-prefix, and will be discarded at the receiver. However, the existence of cyclic prefix in the transmitted data while passing through a dispersive channel leads to a circular interference pattern (as if circular convolution is performed between the data and the channel), which enables exploiting of the simple computational properties of the Fast Fourier Transform (FFT) and its inverse (IFFT) for data down-conversion and up-conversion, respectively. Figure 2.1(b) shows a simplified block diagram of a DMT system.

The overhead of the prefix on system throughput is reduced with shorter channel responses. As a result, some DMT systems utilize a channel-shortening filter either at the transmitter output or at the receiver input to create a less dispersive equivalent channel. However, a simple calculation shows that a simple 4-tap channel-shortening filter has comparable complexity to a 64-point IFFT[3]. Therefore, for small block-size DMT systems, reducing cyclic prefix overhead can be performed almost as efficiently by increasing the IFFT block size. Consequently, we will keep the analysis simple in this chapter and will not consider a channel-shortening filter for our system.

In mathematical terms, addition of the cyclic prefix makes channel's convolution matrix (P) circulant and consequently makes the columns of the DFT matrix (F) its eigenvectors. However, addition of the redundant prefix to the transmitted block incurs a penalty in the overall system throughput as well as the transmitted signal energy. Let

$$P = \begin{bmatrix} p_0 & p_1 & \cdots & p_\nu & 0 & \cdots & 0 \\ 0 & p_0 & p_1 & & & & 0 \\ & & \cdots & & & & \\ 0 & \cdots & 0 & p_0 & p_1 & \cdots & p_\nu \end{bmatrix}_{N\times(N+\nu)} \tag{2.1}$$

[3] A 64-point IFFT implemented as 2 stages of radix-8 IFFT requires roughly 12x64 real additions and 4x64 real multiplications over a 64 sample period. A 4-tap channel shortening filter requires 4 multiplications and 3 additions over 1 sample period.

represent the channel convolution matrix where N is DMT block size (DFT matrix F is NxN) and v+1 is the length of channel dispersion (channel has v+1 taps)[4]. Then, the received signal vector V at the receiver input can be expressed as a linear function of the transmitted vector U at the transmitter output:

$$V_{N\times 1} = PU_{(N+\nu)\times 1} \tag{2.2}$$

For small DMT block sizes, and consequently small prefix lengths (r), it is very likely that channel dispersion exceeds the DMT frame size (v+1>N+l), causing inter-block interference. A DMT frame is defined as a DMT block of size N plus the cyclic prefix. In order to model all interference in the system correctly, we rewrite Equation 2.2 as follows:

$$V_{N\times 1} = P_{Post}U_{Post} + P_{Main}U_{Main} + P_{\mathrm{Pr}e}U_{\mathrm{Pr}e} \tag{2.3}$$

where

$$P = \begin{bmatrix} P_{Post_{Nxl}} & P_{Main_{Nx(N+r)}} & P_{\mathrm{Pr}e_{Nxs}} \end{bmatrix}$$

$$U^T = \begin{bmatrix} U^T_{Post_{Nxl}} & U^T_{Main_{Nx(N+r)}} & U^T_{\mathrm{Pr}e_{Nxs}} \end{bmatrix}$$

Parameters "l" and "s" are equal to the number of taps that cause interference from the following and preceding blocks respectively (v = s+r+l). Without loss of generality and in order to simplify the notation we can assume that the channel is padded with enough zeros such that both s and l are equal to (r+N). If inter-block interference extends to multiple DMT frames, Equation 2.3 can be easily generalized to have multiple $P_{Post}U_{post}$ and $P_{Pre}U_{pre}$ terms; however, we will skip this generality to keep

[4] We are following the notation used in [22] in this paper, which assumes the format of $[x_n\ x_{n-1}\ \ldots\ x_1]$ for a vector X as opposed to the more customary format of $[x_1\ x_2\ \ldots\ x_n]$.

equations simple. We further assume that the receiver locks to the portion of the channel response with the highest energy. The received vector V can be expressed as[5]:

$$V = P_{Post} F_{ext}^{T} \Gamma X_{Post} + P_{Main} F_{ext}^{T} \Gamma X_{Main} + P_{\Pr e} F_{ext}^{T} \Gamma X_{\Pr e} \tag{2.4}$$

where $F_{ext}^{T} = \begin{bmatrix} F^{T} \\ F_{,}^{T}(1:r) \end{bmatrix}$ is the IDFT matrix augmented with its first r rows to create the cyclic prefix, Γ is the diagonal N by N power allocation matrix, and X_{Post} and X_{Pre} are the following and preceding interfering blocks of symbols respectively. Finally the received signal Y is given by:

$$Y_{N\times 1} = FV = FP_{Post} F_{ext}^{T} \Gamma X_{Post} + FP_{Main} F_{ext}^{T} \Gamma X_{Main} + FP_{\Pr e} F_{ext}^{T} \Gamma X_{\Pr e} \tag{2.5}$$

The first and third terms on the right hand side of the equation basically model interference from other transmitted signal blocks at the receiver.

The DMT system described above is bound by two types of constraints: per sub-channel BER constraint and transmit peak voltage constraint. Using Equation 2.4, for a given bit-loading $b = [b_1 \quad b_2 \quad \cdots \quad b_N]$, we can express both constraints as convex inequalities as functions of the power allocation variables Γj.

Using standard BER expression for a PAM constellation [22], the BER constraint for the k^{th} sub-channel is [6]:

$$2\left(1 - 2^{-b_k}\right) Q\left(\frac{d_{\min\_k}}{2\sigma_k}\right) < BER \tag{2.6}$$

Here d_{min_k} is the minimum distance between the constellation points for the signal y_k received at the k^{th} sub-channel, and σ_k is the standard deviation of noise at the

[5] Conjugate symmetry should be assumed for the input vector X so that transmitter output U is real. A conjugate symmetric vector still has N independent real variables and the vector of N real variable can be mapped to the conjugate symmetric vector X through a linear transformation. Therefore, for simplicity, in all the equations we assume that X represents the vector of the N real variables and the linear transform that maps X to a conjugate symmetric vector is absorbed into the IDFT matrix F^T.

receiver. A Gaussian distribution is assumed for the noise, which consists of thermal noise and (residual) interference. This approximation is justified for white input data stream since for reasonable performance (residual) ISI is caused by a large tail of reflection taps with small magnitudes. Equation 2.6 can equivalently be written as:

$$Q^{-1}\left(\frac{BER}{2\left(1-2^{-b_k}\right)}\right)\sigma_k - d_{\min\_k} < 0 \tag{2.7}$$

Using Equation 2.5, we know that symbol x_k transmitted from the k^{th} sub-channel would appear at the receiver as:

$$Y_k = F_k P_{Main} F_{ext}^T \Gamma\left(x_k 1_k\right) \tag{2.8}$$

Where 1_k is a vector that is one at the k^{th} position and zero otherwise. F_k is the k^{th} row of the DFT matrix. Also assuming that energy (per dimension) (E_x) for all transmitted symbols is unity, using standard expressions for the energy of a PAM constellation with minimum distance of d_{min} [22], the minimum distance between the constellation points of x_k is:

$$d_{\min\_k\_TX} = \sqrt{\frac{12}{2^{2b_k}-1}} \tag{2.9}$$

Consequently, the minimum distance between the constellation points at the receiver is:

$$d_{\min\_k} = F_k P_{Main} F_{ext}^T \Gamma\left(\sqrt{\frac{12}{2^{2b_k}-1}} 1_k\right) = \left(\sqrt{\frac{12}{2^{2b_k}-1}} \Gamma_k\right) F_k P_{Main} F_{ext}^T 1_k \tag{2.10}$$

which is a linear function of Γ_k. Furthermore, using Equation 2.5 again:

$$\sigma_k^2 = F P_{Post} F_{ext}^T \Gamma E_X \Gamma^T F_{ext} P_{Post}^T F^T + F P_{Pre} F_{ext}^T \Gamma E_X \Gamma^T F_{ext} P_{Pre}^T F^T + \sigma_{Thermal}^2 \tag{2.11}$$

which is a quadratic function of power allocation coefficients Γj (j = 1,2,…,N). Therefore the BER constraint is a convex, or more precisely a Second Order Conic (SOC) inequality in Γj. The middle term in Equation 2.5 does not contribute to (2.11) due to the cyclic prefix.

The transmitter peak voltage constraint exists because transmitter output has a maximum swing Vp, and consequently, all output samples should be less than or equal to Vp to keep the line driver circuit in intended operation region. Starting from the expression for transmitter output vector U, the peak voltage constraint can be derived as:

$$U = F_{ext}^{T}\Gamma X_{Main} \le \left|F_{ext}^{T}\right|\Gamma Max\left(X_{Main}\right) = \left|F_{ext}^{T}\right|\Gamma Diag\left(\sqrt{\frac{3\left(2^{b_k}-1\right)}{2^{b_k}+1}}\right)1 < V_p \qquad (2.12)$$

which is equivalent to a set of N independent linear inequalities in Γj and the single-ended swing Vp. $\left|F_{ext}^{T}\right|$ is the element-wise absolute of $\left|F_{ext}^{T}\right|$ and 1 is a vector of all ones. The first inequality in (2.12) is based on the fact that maximum swing at the output is created when the signs of the (real) input symbols match the signs of the IDFT matrix coefficients, and the input symbols assume their maximum values. With the assumption of unit energy (per dimension) for transmit signals, $\max(x_k)$ is equal to:

$$Max(x_k) = \left(2^{b_k}-1\right)\frac{d_{\min\_k\_TX}}{2} = \sqrt{\frac{3\left(2^{b_k}-1\right)}{2^{b_k}+1}} \qquad (2.13)$$

For large FFT sizes, the above formulation would be a little pessimistic, because the probability of getting a peak would go even below the target BER of 10-15. A better approximation in such cases would be through clipping probability. The signal variance vector at the transmitter output is given by:

$$\sigma_U^2 = \left|F_{ext}^{T}\right|^2 \Gamma^2 1 \qquad (2.14)$$

where $\left|F_{ext}^{T}\right|^2$ is the element-wise norm-2 square of F_{ext}^{T}. In order to achieve certain clipping probability P_{clip}, we should have the following, which is also an SOC constraint.

$$Q^{-1}\left(P_{Clip}\right)\sigma_U < V_P \tag{2.15}$$

Finally, for a given sub-channel bit allocation, the optimal power allocation can be obtained through solving the following convex optimization problem [21][25]:

$$\begin{aligned}
&Minimize\ V_P \\
&Subject\ to: \\
&Q^{-1}\left(\frac{BER}{2\left(1-2^{-b_k}\right)}\right)\sigma_k - d_{\min_k} + eye_k < 0 \quad k = 1,2,...,N \\
&\left|F_{ext}^{T}\right|\Gamma Diag\left(\sqrt{\frac{3\left(2^{b_k}-1\right)}{2^{b_k}+1}}\right)1 < V_p
\end{aligned} \tag{2.16}$$

where eye_k is the minimum resolvable signal level (or eye opening) or receiver sensitivity at the k^{th} sub-channel output. It is therefore a positive number by definition[6]. If V_P is set by implementation constraints, then the following feasibility problem can be solved instead.

$$\begin{aligned}
&Is\ Feasible: \\
&Q^{-1}\left(\frac{BER}{2\left(1-2^{-b_k}\right)}\right)\sigma_k - d_{\min_k} + eye_k < 0 \quad k = 1,2,...,N \\
&\left|F_{ext}^{T}\right|\Gamma Diag\left(\sqrt{\frac{3\left(2^{b_k}-1\right)}{2^{b_k}+1}}\right)1 < V_p
\end{aligned} \tag{2.17}$$

[6] eye_k can be set to zero if infinite resolution is assumed for the decision devices at the receiver. The reason for the introduction of this new parameter will become clearer shortly as we discuss the bit-loading algorithm.

The above SOC problems can be efficiently solved using conic solvers like Mosek[7]. The sub-channel bit loading, however, is in integer space, and therefore, obtaining the optimal bit-loading requires a little bit more work.

2.2 Incremental Bit-loading and Power Allocation

A well known integer bit-loading method for DMT systems is the Levin-Campello (LC) algorithm, which is based on greedy optimization [26]. Starting from an "efficient" bit-loading $b = [b_1 \quad b_2 \quad \cdots \quad b_N]$, the bit-loading that leads to the least transmit power (peak voltage[8]) among all bit distributions with same total number of bits, the algorithm assigns the next incremental bit to the sub-channel that leads to the least increase in the overall transmit energy (peak voltage). This algorithm is therefore efficient when the increase in total transmit energy due to an incremental change in the bit distribution can be tabulated or obtained easily. In systems that are constrained by average transmit power, this is achieved through the well known gap approximation [22]. However, in a system driven by optimization problem (2.17), obtaining such data requires solving the convex optimization of (2.16) N times for all possible incremental bit-assignments to the sub-channels. This process can be very slow. Fortunately, the structure of the problem offers a faster alternative.

Looking at the feasibility problem (2.17) we can verify that the first term on the left side of the BER constraint and the peak voltage constraint are both relatively insensitive to weak functions of the bit-loading while the second term (d_{min}) is a strong function. Therefore, assuming the problem is feasible for a given eye-opening and certain bit-loading, power allocation and V_P, it is possible to trade a larger constellation size (smaller d_{min}) with smaller eye-opening (higher slicer sensitivity) as

[7] www.mosek.com
[8] Any function that increases with adding a bit can be substituted here.

long as the eye remains positive[9]. More exactly, in order to increase b_k to b_{k+1} while satisfying the BER constraint, eye_k should be decreased by

$$-\Delta d_{\min\_k}(b_k) = \sqrt{3}\Gamma_k F_k P_{Main} F_{ext}^T 1_k \left(\frac{1}{\sqrt{2^{2b_k}-1}} - \frac{1}{\sqrt{2^{2b_k+2}-1}} \right) \tag{2.18}$$

Therefore, we can conclude that at a given problem setting (given eye opening, power allocation and bit-loading), every $\frac{eye_k}{-\Delta d_{\min\_k}(b_k)}$ of the eye opening is worth a bit for the k^{th} sub-channel. Based on this heuristic, the incremental bit-loading algorithm can be expressed as follows:

- Step 1: Find a feasible bit-loading
- Step 2: Starting from a feasible point, maximize total incremental bit assignment for the system by solving the convex problem (2.19) and obtain the optimum set of eye-openings.

$$\begin{aligned} &Maximize\ \Delta b = \sum_k \Delta b_k = \sum_k \frac{eye_k}{-\Delta d_{\min\_k}(b_k)} \\ &Subject\ to: \\ &Q^{-1}\left(\frac{BER}{2\left(1-2^{-b_k}\right)} \right)\sigma_k - d_{\min\_k} + eye_k < 0 \quad k = 1,2,...,N \\ &\left| F_{ext}^T \right| \Gamma Diag\left(\sqrt{\frac{3\left(2^{b_k}-1\right)}{2^{b_k}+1}} \right) 1 < V_p \end{aligned} \tag{2.19}$$

- Step 3: Sort the sub-channels based on their eye-opening. Start with the sub-channel with maximum eye-opening.
- Step 4: Assign one bit to the sub-channel.

[9] We are treating eye_k as a variable here.

- Step 5: Run feasibility problem (2.17). If feasible, go to step 2. If not feasible, revert previous bit assignment, choose the next sub-channel with the largest eye-opening and go to step 4[10].

The algorithm terminates when no bits can be added to any of the sub-channels. Since throughout the algorithm, power allocation is performed to maximize incremental bit-assignment, the final bit-assignment would be very close to the optimal. Step 1 of the algorithm can be performed as follows:

Initialization Phase: Finding a starting feasible point:

1. Assign one bit to all sub-channels.
2. Check for feasibility (2.17).
3. If feasible, done; otherwise, assign zero bits to the channel with the highest attenuation among active channels. Repeat step 2.

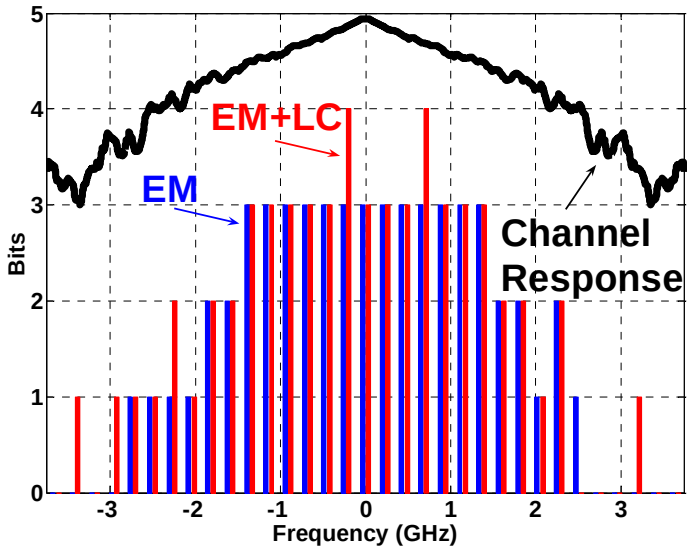
We will refer to this algorithm as Eye Maximization (EM) algorithm in the remaining of the paper. The algorithm may also be combined with the "efficentizing"[11] [26] phase of the LC algorithm to correct any deviations from the optimum bit-assignment. For example after one run of EM, the final bit-loading can be "efficentized" and then a new run of EM starting from the new feasible point leads to the global optimum. Our simulations show that about 5-10% improvement in overall data rate may be possible using the combined scheme at the expense of 5-10 times longer simulation time. The gains diminish with larger FFT block size.

---

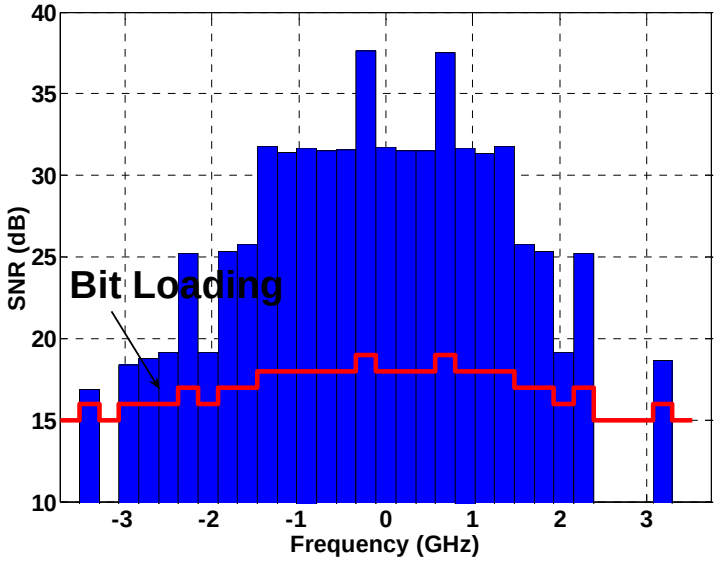
[10] It would be more logical if in the 3rd step, channels are sorted based on the (real) valued incremental bits (Δb_k) that can be assigned to them by solving the following equation; however, simulation results don't show any difference in the final results.

$$\sqrt{3}\Gamma_k F_k P_{Main} F_{ext}^T 1_k \left(\frac{1}{\sqrt{2^{2b_k}-1}} - \frac{1}{\sqrt{2^{2b_k+2\Delta b_k}-1}} \right) = eye_k$$

[11] "Efficentizing" is a stage in the LC algorithm where starting from a given bit distribution, the bit distribution is obtained that with same total number of bits leads to the minimum total transmit energy (or voltage in our case).



(a)



(b)

Figure 2.2: (a) Bit-loading over a 20” backplane FR4 channel for EM and EM+LC. (b) SNR at the receiver for EM+LC. Both figures for a DMT system with block size of 32, cyclic prefix length of 8, and sample rate of 7GHz. Peak to peak transmit voltage is 1.6V and Noise Figure is set to 10dB.

To demonstrate the result of the bit-loading algorithm, Figure 2.2(a) shows the bit-loading results obtained over a 20” FR4 backplane channel in one example setting where FFT size is set to 32, cyclic prefix is 8, and Nyquist frequency is 3.5GHz. The results are shown for the EM bit-loading as well as for the EM+LC. Bits from the left to the middle of the graph represent the Q rail and from middle to the right represent the I rail of the QAM constellations. The channel response is also imposed on the figure with an arbitrary scale. Signal to Noise Ratio (SNR) at the receiver output is shown in Figure 2.2(b). It can be seen that every extra bit is equivalent to about 6dB of SNR.

2.3 DMT System Requirements

Using the analysis framework developed in the previous section, we can find the DMT system requirements in terms of block size and ADC resolution and speed. We will perform this analysis on two standard 20” backplane link channels. The frequency response of the two channels is shown in Figure 2.3.

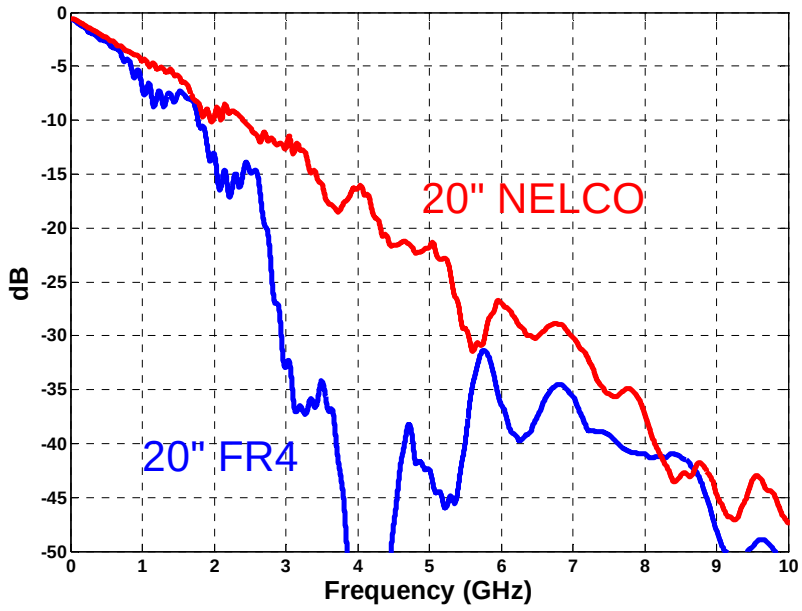


Figure 2.3: Frequency Response of 20” FR4 and NELCO backplane channels

The FR4 channel (same as the one used in the previous section), is a PCB trace which connects a switch on a line-card to another switch on a different line-card over the top layer of the backplane. Due to the existence of vias in the signal path, notches exist in the frequency response of the channel. A cross-section of a backplane system is shown in Figure 2.4. The NELCO channel is similar to the FR4 channel except that vias in the signal path are removed through a manufacturing process, and the trace is routed on the bottom layer of the backplane board. As a result the NELCO channel has a smoother frequency response.

We will first look at the effect of DMT block size on the performance of the DMT system, and then include the effect of ADC quantization noise into the results.

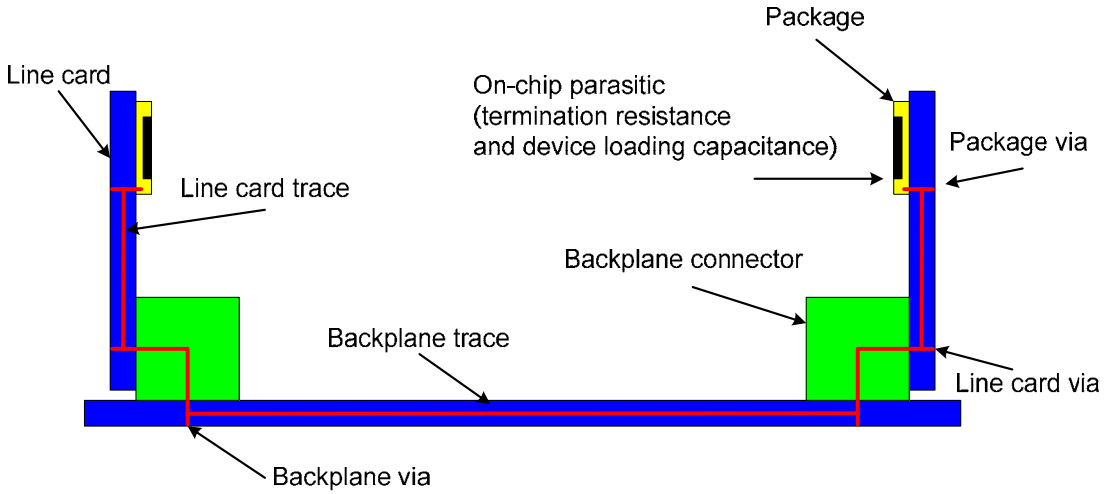


Figure 2.4: Cross-section of a backplane system with two line-cards.

2.3.1 **Effect of DMT Block Size**

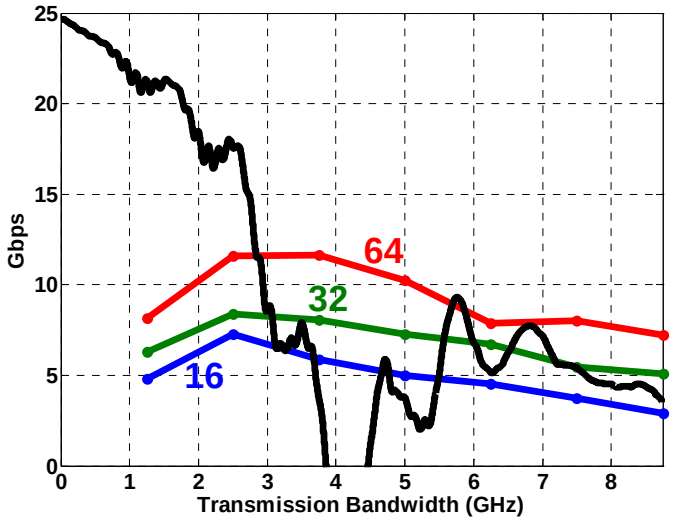
Figure 2.5 shows the simulation results over 20" FR4 and NELCO backplane channels for DMT block sizes of 16, 32, and 64. The transmitter voltage is constrained to 1.0V peak-to-peak and a Noise Figure (NF) of 10 dB with respect to a 50Ω termination was considered for the receiver. No other non-idealities were considered, including ADC quantization noise. For every channel, the transmission bandwidth (half the DMT transmitter output sample rate) was varied from 2.5-GHz to 8.75-GHz

in steps of 1.25-GHz and optimum sub-channel bit assignment and power allocation were obtained in each scenario using the EM algorithm. For each data point in Figure 2.5 the length of the cyclic prefix was varied from 8 to 20 in steps of 4, and the maximum data rate was selected. Figure 2.5 demonstrates a trade off between DMT sample rate and the length of the prefix. As sampling rate increases, system can exploit a larger portion of the channel; however, since there will be more interference taps, cyclic length increases as well. There is a point where the cyclic penalty cancels out the effect of the increased data rate and throughput starts to fall. Based on the results, optimum settings for data transmission over the FR4 channel are a DMT block size of 64, a cyclic prefix length of 8, and DMT sample rate of 5-GHz (transmission bandwidth of 2.5-GHz) to achieve 11.5-Gb/s of data-rate. Over the NELCO channel the optimum configuration is obtained with a DMT block size of 64, a cyclic prefix length of 16, and DMT sample rate of 10-GHz to achieve 23-Gb/s of data-rate. If we exclude the cyclic prefix penalty from rate calculations, we may conclude that data rates in excess of 15-Gb/s and 30-Gb/s over FR4 and NELCO are achievable respectively in DMT system with very large block sizes.

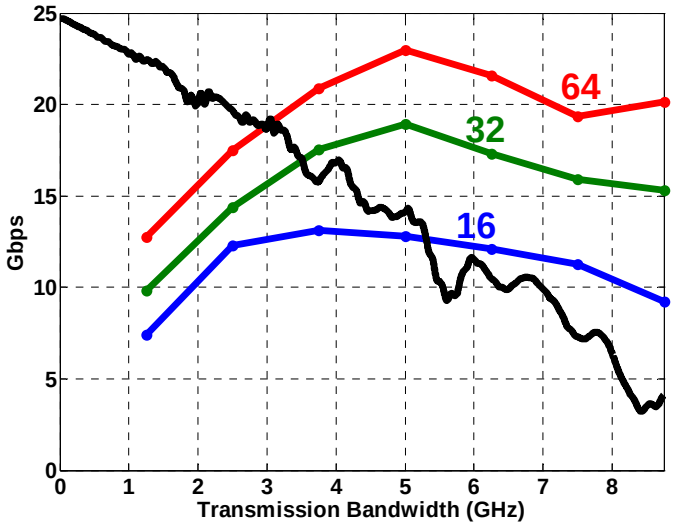
Figure 2.6 shows the optimum bit-loadings used to achieve the above maximum data rates. An interesting observation in both cases is that due to the gradual roll-off of the channel, number of bits assigned to the sub-channels varies very slowly as the sub-channels move away from DC. In fact it is possible to group every few of adjacent sub-channels to larger macro sub-channels which have same number of bits per symbol. This is done in Figure 2.6 with solid red lines connecting the bars together. The new bit-loadings are still feasible and total throughput is about 6% and 12% less than the original configuration for FR4 and NELCO respectively. This observation suggests if it wasn't for the prefix penalty, the size of the DMT block could have been very small. We will exploit this property in next chapter to design an MT transmission algorithm which is feasible within the link complexity and power constraints.

Figure 2.7 shows the power of the residual interference at the receiver output for both channels. Even for an FFT block size of 64 the system is completely (residual) interference dominated: the interference power is about 12 dB above thermal noise for

FR4. This means that the system would clearly benefit from shortening of the reflection tail of the channel.

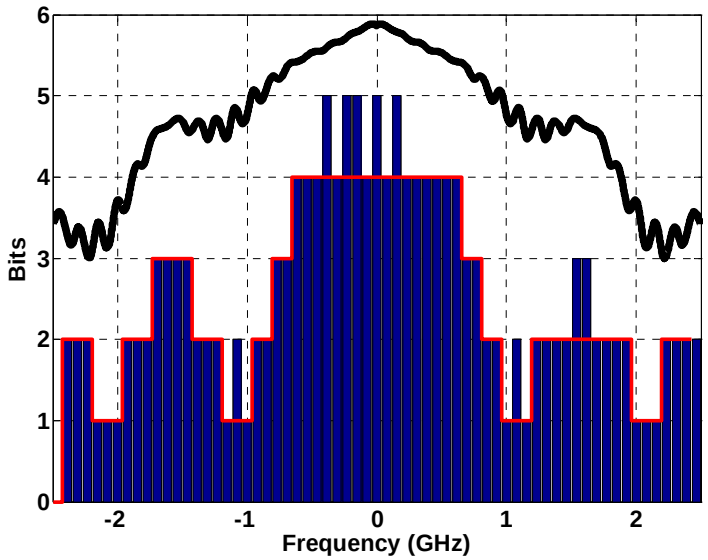


(a)

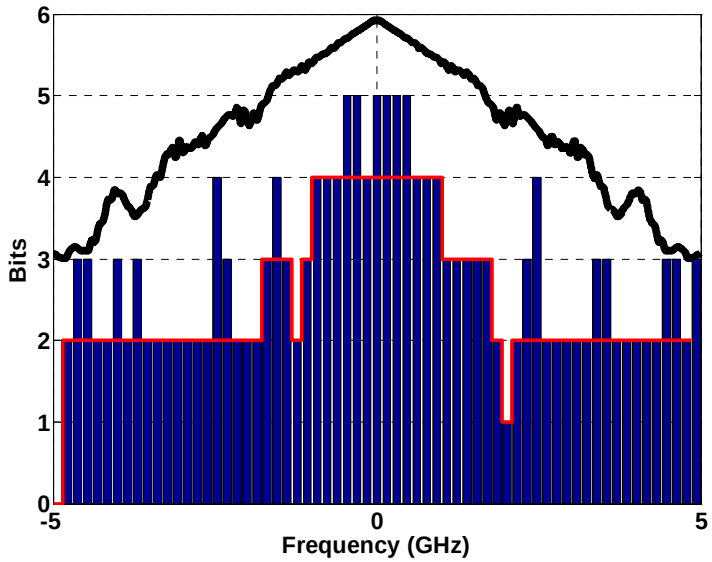


(b)

Figure 2.5: Maximum achievable data rate (with optimum prefix length) over 20" FR4 (left) and 20" NELCO (right) channels for DMT block-sizes of 16, 32, 64 and transmission bandwidths (half the FFT rate) of 1.25-GHz to 8.75GHz. Channel frequency response (in dB) super-imposed on the figures with an arbitrary scale. Noise Figure is assumed to be 10dB.

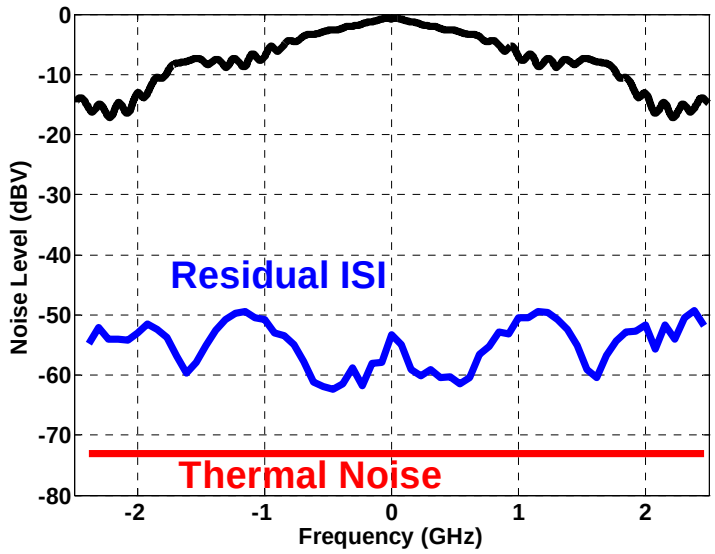


(a)

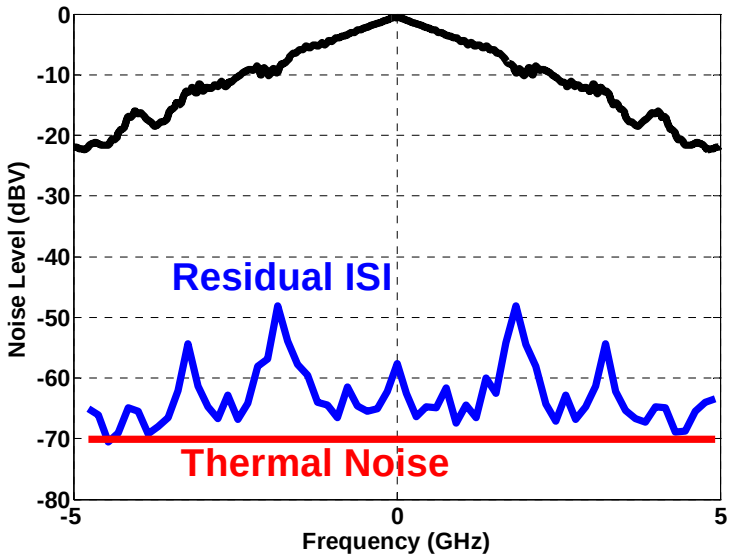


(b)

Figure 2.6: Optimum bit-loading obtained with the EM method to achieve maximum throughput over the FR4 (left) and NELCO (right) channels. The Solid red line in both figures represents an alternative feasible bit-loading, in which adjacent sub-channels with similar bit assignment are grouped together.

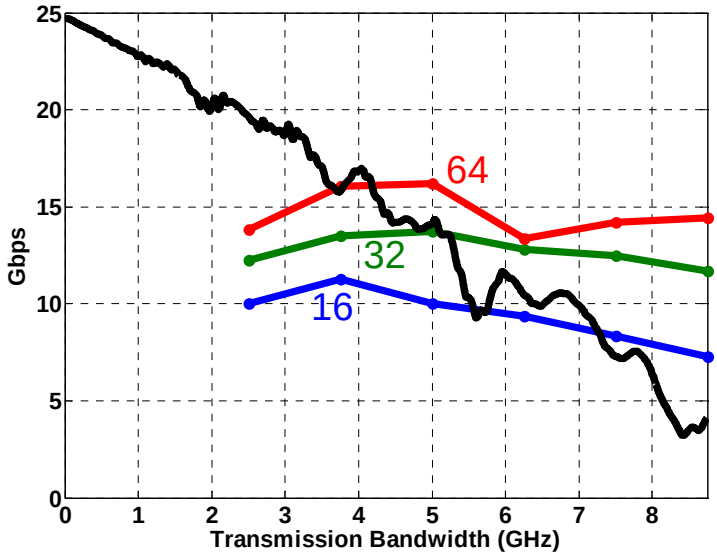


(a)

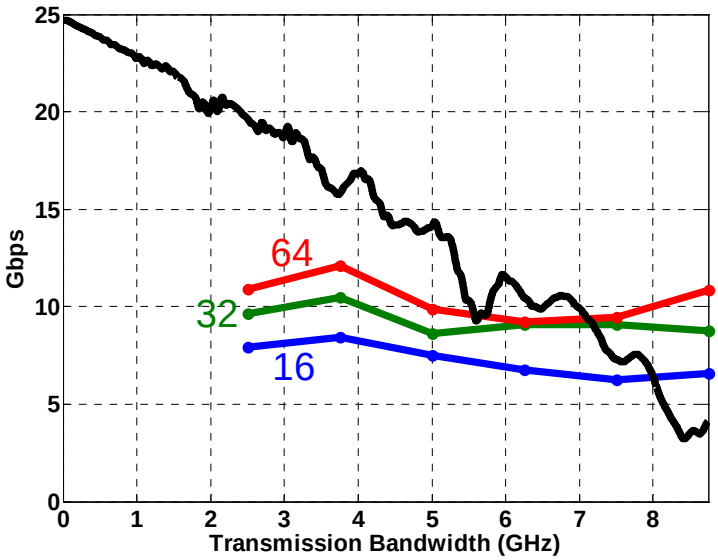


(b)

Figure 2.7: Residual interference at the Receiver for the FR4 (a) and the NELCO (b) channels. Noise Figure of 10dB over system bandwidth is also shown in the figures.



(a)



(b)

Figure 2.8: Maximum achievable data rate (with optimum prefix length) over the NELCO channel for DMT block-sizes of 16, 32, 64 and ADC resolution of 7-bits (a) and 6-bits (b). Other simulation settings are similar to Figure 2.5.

2.3.2 Effect of ADC Quantization Noise

The results in the previous section neglected the effect ADC quantization noise. Assuming a white uniform distribution for the quantization noise, Figure 2.8 shows the simulation results for a 7-bit ADC at the receiver for the NELCO channel. The rest of the simulation settings and parameters are similar to Figure 2.5. The results indicate that the maximum achievable data rate with 7-bits of quantization is 16-Gb/s, a 30% reduction from 23-Gb/s with an infinite-resolution ADC. The maximum with a 6-bit ADC drops to 12-Gb/s.

2.3.3 System Requirements

The results in Figure 2.5 and 2.8 indicate that 3.75-GHz of transmission bandwidth is close to optimum for different quantization noise levels. For this specific transmission bandwidth, Figure 2.9 shows the maximum achievable data rate with DMT block sizes of 16, 32, and 64, and ADC resolutions of 5, 6, and 7 bits.

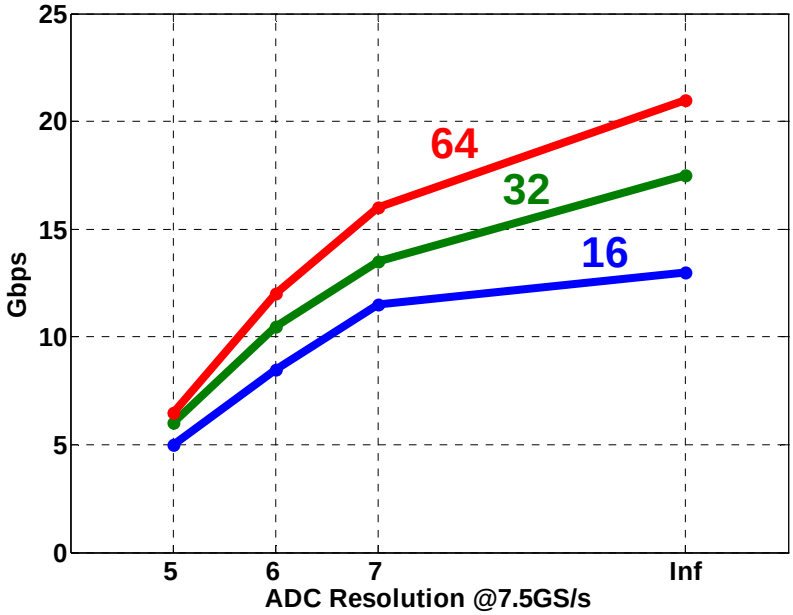


Figure 2.9: Maximum achievable data rate (with optimum prefix length) over the NELCO channel for DMT block-sizes of 16, 32, 64 and ADC sampling rate of 7.5-GHz, and resolution of 5, 6, 7, and infinite bits.

A state of the art BB system achieves a data rate of 12-Gb/s over the NELCO channel used as a reference in our analysis in this section. One way to characterize the DMT system's complexity is to find the system requirements to match the performance of the BB system. The data provided in Figure 2.9 indicates that the DMT system requires either a 6-bit 7.5-GS/s ADC and a block size of 64, or a 7-bit 7.5-GS/s ADC and block size of 32. Let's consider the second configuration to obtain a simple lower-bound on the complexity of the DMT system.

64-point FFT is employed in wireless LAN systems compatible with IEEE's 802.11a standard and a few journal publications are available on its power and area [28][29]. However, the operating rate of such circuits is almost three orders of magnitude lower than our target rate of 7.5 GS/s. For example, the architecture in [28], implemented in a 1.8V 0.25-μm technology supporting 16-bit inputs, runs at 20-MS/s, dissipates 41-mW of power, and occupies 6.8-mm^2. Clearly it is not possible to take the exact same architecture and employ it in a 7.5-GS/s system. Nonetheless, by making the assumption that it is possible and applying standard CMOS scaling rules we can obtain a lower-bound for the power and area of a 64-point FFT circuit at our target rate. Assuming a 1.0V 90-nm CMOS technology, the estimated power for the architecture in [28], running at 7.5-GS/s with 6-bit inputs is 650-mW. The corresponding area is 0.33-mm^2. We should further discount this number to account for the fact that the FFT circuit required for a high-speed link has real inputs. The FFT of 64 real numbers can be obtained by computing the FFT of 32 complex numbers [30]. Consequently, the lower bound is around 270-mW for an FFT, or 540-mW total for the transmitter and the receiver combined. An ADC with 3.5 effective number of bits consumes 1.2W at 24-GS/s in a 90-nm CMOS technology [31]. Same ADC has 5-bit effective resolution when running at 7.5-GS/s. Assuming power scales linearly with the operating frequency, the estimated power-consumption is 375mW.

Using these estimates, a lower-bound for the power consumption of a 12-Gb/s DMT link to match the performance of a BB link is around 915-mW, equivalent to an energy-efficiency 76-mW/Gbps. This number is almost twice the energy efficiency of

the state of the art BB links. The actual power consumption of the DMT system would certainly exceed this number.

2.4 Summary

Our analysis of DMT over typical link channels indicates that theoretically MT has the potential for achieving data rates as high as 15-Gb/s to 30-Gb/s. DMT with medium block size and cyclic prefix in particular can achieve 11.5-Gb/s to 23-Gb/s over the channels of Figure 2.3. Unfortunately even in these simplified setups, the power consumption of the system will likely be too high to compete with baseband links.

One promising observation about the characteristics of the link channels is that close to optimum bit-loading (transmit power allocation) can be achieved with very small number of sub-channels, and the large block size, with the high computation power, is only needed to reduce the penalty of the prefix overhead. However, the essence of MT is its efficiency in achieving the optimum transmit power spectrum; cyclic prefix is just a clever way to simplify the implementation. Therefore, thinking outside the framework of Discrete Multi-tone and drawing insight from our analysis in this section leads one to conclude that the most efficient MT architecture for backplane links is one in which dispersion is treated independently of the number of tones. Furthermore, our analysis indicates that ADC resolution and speed requirements can significantly add to the complexity of a link system. Conventional BB systems get around this problem by performing the FF signal processing in the receiver in analog domain. A similar idea, if applied to an MT architecture can lead to a more efficient system. The analog multi-tone system described in the next chapter leverages both of these insights to create a power efficient multi-tone system.

Chapter 3

Analog Multi-Tone

In this chapter we propose an MT architecture, called Analog Multi-Tone (AMT), which is customized to the characteristics of link systems, and is therefore power-efficient. The AMT system consists of only a small number of wideband sub-channels, and employs linear transmit equalization and receive DFE to compensate for the effects of the frequency selective sub-channels and the interference among them.

We start the chapter by describing the architecture of the AMT system and compare its equalization complexity to an equivalent baseband system. In order to obtain the equalizer taps, we model the MT system as a Multi-Input Multi-Output (MIMO) system where equalizers perform both equalization and power allocation at the same time. We will describe the mathematical system model and show that the BER and transmit (peak) power constraints can be cast as Second Order Conic (SOC) constraints [21], and therefore, optimal equalizer taps can be obtained through convex optimization. We also show that with appropriate definition of the cost function and some approximations, a BER-constrained zero-forcing solution can be obtained for the equalizer taps. In this approach, equalizer taps (up to a scale factor) are obtained through conventional ZF techniques, and then joint power allocation is performed through simple matrix inversion. Such a solution lends itself well to MMSE adaptation.

The AMT system can achieve superior performance compared to BB systems through better allocation of the transmit power over-channels with frequency selective characteristics, and also through fundamentally parallelizing the receiver architecture,

leading to better receiver sensitivity. Using the developed analysis framework, we compare the performance of the AMT system with conventional BB system to verify its potential.

3.1 Analog Multi-Tone Architecture

In order to have an energy-efficient MT architecture at the extremely high operation rates necessary for high-speed links, a BB link can be extended into a bank of parallel links operating at different carrier frequencies. Figure 3.1 shows a conceptual MT system based on this idea.

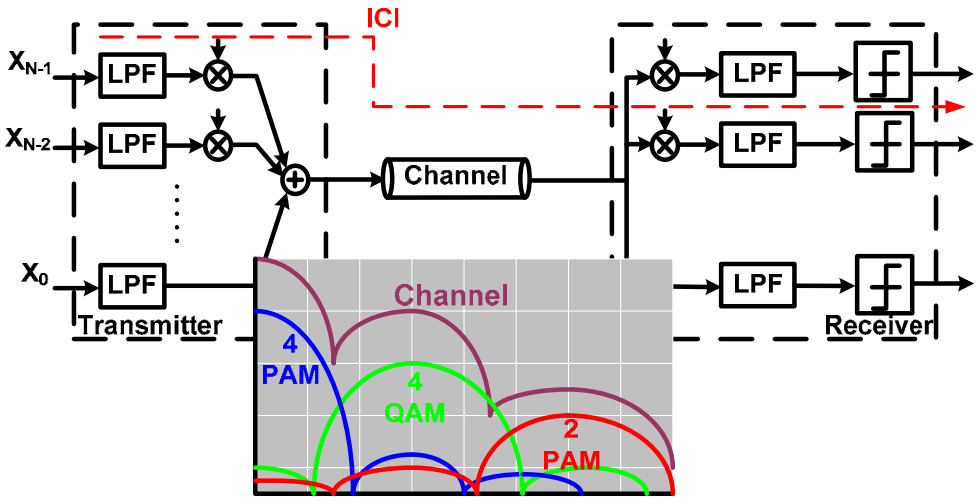


Figure 3.1: A conceptual multi-tone system. X_0, …, X_N are input sequences

Each sub-channel (the path from an input at the transmitter to the corresponding output at the receiver) in this figure can potentially have a bandwidth of a few GHz and only a small number of sub-channels may exist. This architecture, however, suffers from the unavailability of fully integrated filters with sharp roll-off. Therefore, energy from one sub-channel will inevitably spill over to the neighboring sub-channels, causing inter-channel interference (ICI). ICI is similar to ISI in nature, with the exception that it represents interference from symbols from another sub-channel. Therefore, as long as the transfer function from the input of the interfering sub-channel to the output of the target sub-channel does not change from sample to sample

– i.e. as long as the system is linear time invariant (LTI) in the discrete domain – ICI can be cancelled in the same way as ISI; through equalization. For the system to be considered LTI in discrete domain, these two constraints should be fulfilled:

1. All sub-channel symbol rates are the same.
2. All carrier frequencies are integer multiples of the sub-channel symbol rate.

These two constraints will cause all carrier frequencies to complete full cycles from one symbol to another. This means that from the input sequence's perspective the system does not change from one cycle to another, and is therefore time-invariant in the discrete domain. In particular, the ICI pattern does not change from one symbol to another, and consequently, both ICI and per-sub-channel ISI can be canceled through equalization. However, ISI and ICI cancellation in this multi-input multi-output (MIMO) system requires MIMO linear transmit equalization and MIMO receive DFE [32]. Alternatively, as a consequence of the two constraints imposed on the system, N-times[12] over-sampled equalizers per sub-channel can be utilized at the transmitter to substitute the MIMO equalizer, the low-pass filters, and the mixers (Figure 3.2). An appropriate choice for the low-pass filter in the receiver that leads to superior performance and can be implemented reliably on chip is an integrate-and-dump circuit that integrates over one sub-channel symbol period [33][13]. Figure 3.2 shows the finalized AMT architecture.

In the architecture of Figure 3.2, each N-times over-sampled equalizer can shape the transmission bandwidth (from dc to the Nyquist frequency) of the entire system. Therefore, when all the sub-channel equalizers are optimized simultaneously, they work together to cancel both ISI and ICI at the same time.

---

[12] N-times with respect to sub-channel symbol rate. Assuming all sub-channels are 2PAM, each sub-channel equalizer is operating at the throughput rate.

[13] In general any low-pass filter that forms a perfect reconstruction set together with its up-converted versions can replace the integrators. Filers with sharp frequency-domain roll-off require more transmit equalizer taps and less cross DFE taps.

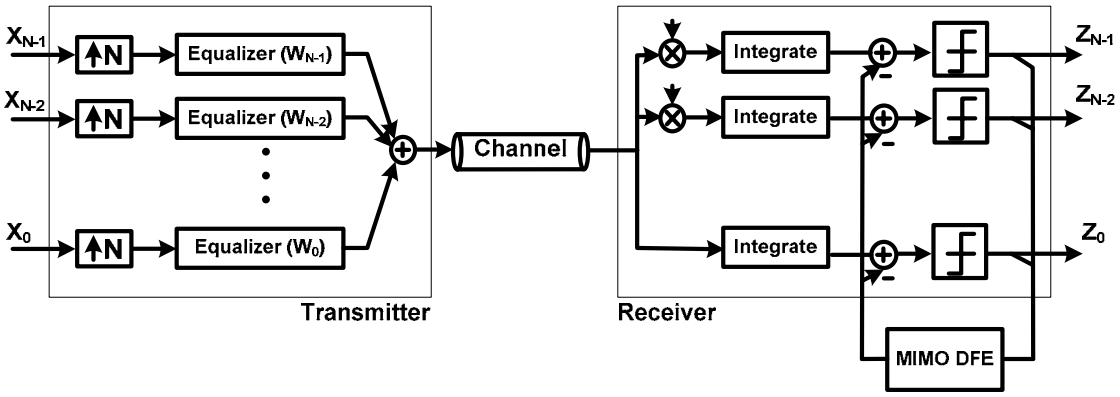


Figure 3.2: Analog Multi-Tone architecture

From a time-domain perspective, the AMT system is a trans-multiplexer operating based on the principle of perfect reconstruction [34]. The time-domain waveforms of an example 2-channel AMT system are plotted in Figure 3.3 to illustrate this point further.

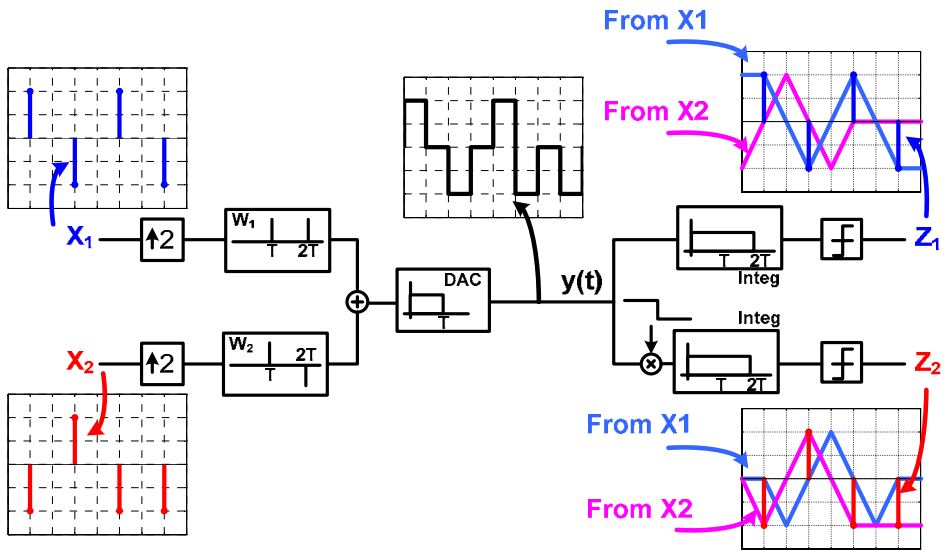


Figure 3.3: Signal waveforms in an example 2-channel AMT system. Two-tap equalizers per sub-channel at the transmitter, ideal channel, and no DFE at the receiver. Continuous-time ISI and ICI patterns are shown at the sampler inputs

In this example, the input sequences X_1 and X_2 are binary, transmit equalizers are 2-times over-sampled 2-tap filters programmed to create a low-pass and a high-pass spectrum for X_1 and X_2 respectively, a zero-order hold filter performs discrete-to-

analog conversion, the channel is ideal with no dispersion, and there is no MIMO DFE in the receiver. Continuous time ISI and ICI patterns at the input to the samplers are also shown along with the sampled sequences, Z_1 and Z_2. It can be seen that even though the transmit equalizers are very short, and consequently significant energy overlap exists between the two sub-channels at the transmitter output, both the ISI and the ICI are forced to zero at the sampling points at the receiver. As a result, X_1 and X_2 are fully recovered at the sampler outputs without any interference. However, the ISI and the ICI are not necessarily zero at points other than the sampling point. This is very similar to the operation of an equalized BB system where a linear equalizer forces the ISI to zero (only) at the sampling points.

From a different perspective, ignoring the MIMO DFE, for large number of sub-channels, the AMT system shown in Figure 3.2 is a very inefficient implementation of a DMT system in which the IFFT operation at the transmitter is performed using N-times over-sampled equalizers [27]. However, for a small number of sub-channels, the AMT system is highly efficient and its performance is not limited by cyclic prefix overhead - a limiting factor in small block size DMT systems. In fact, it can be shown that the equalization complexity (and consequently equalization power) of the AMT system is comparable to a BB system operating at the same overall throughput. To demonstrate this point further, Figure 3.4 shows a 2-way parallelized BB transmit equalizer, and a 4-tap-per-channel 2-channel AMT equalizer.

The BB and AMT equalizers look structurally identical except that the taps in the lower branch of the 2-way parallelized BB equalizer are constrained to be a shifted version of the upper branch. The taps in the two branches of the AMT equalizer, on the other hand, have more degrees of freedom and can take values independently. In other words, the transmit equalizer of the AMT system is a general form of a BB transmit equalizer[14]. The additional degrees of freedom in the transmitter enable the AMT system to shape the transmit spectrum better than a BB system. As a result, an

[14] The same complexity (power) argument also applies to the MIMO DFE at the receiver of the AMT system.

AMT system performs considerably better than a BB system over channels with notches in their frequency response, where optimal shaping of the transmit spectrum is crucial. Over smooth channels, however, where both AMT and BB can achieve close to optimum transmit power allocation, the performance of the AMT system is affected by its higher Peak-to-Average Power Ratio (PAPR) compared to the BB system[15].

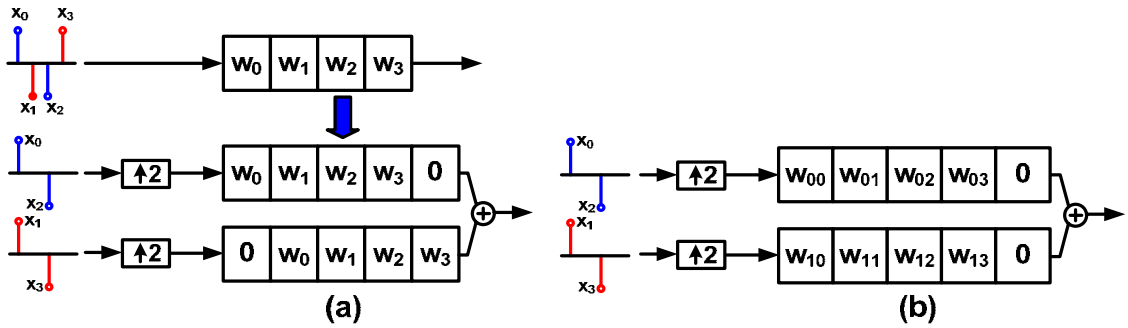


Figure 3.4: (a) A 4-tap linear equalizer represented as a 2-way parallelized filter (b) A 2-channel 4-tap per channel AMT equalizer

3.2 System Modeling and Analysis

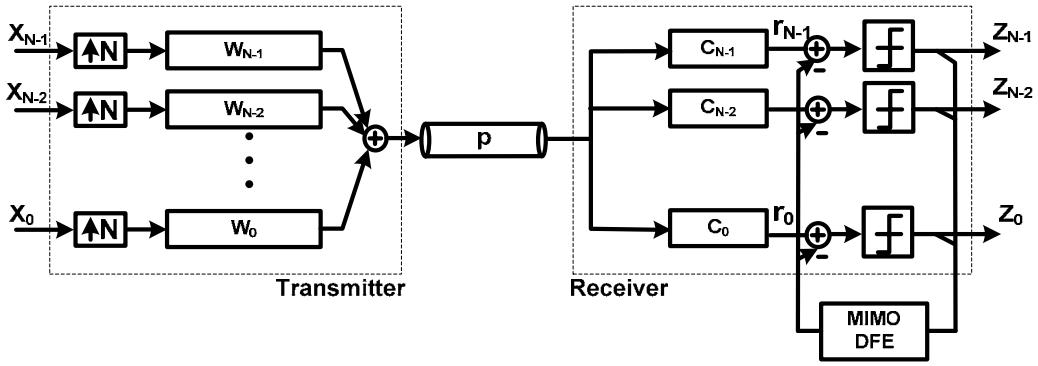


Figure 3.5: AMT System Model

[15] PAPR is the ratio of the average transmit power to the peak transmit power for any modulation scheme. For example, PAPR for 2-PAM and 4-PAM modulations are 1 and 9/5, respectively. PAPR for 2-channel AMT with 2-PAM sub-channels is 2. For a given transmit peak voltage, PAPR determines the maximum average transmit energy that can be delivered to the channel (or equivalently, the transmit signal to noise ratio).

A simplified block diagram of the system is shown in Figure 3.5. In this figure w_0, w_1, …, $w_{(N-1)}$ represent the linear equalizer taps at the transmitter, p is the channel pulse response, and c_0, c_1, …, c_{N-1} are filters matched to the mixers and integrators at the receiver.[16] p may be assumed to be time-invariant and known.[17] Let:

$$\mathbf{p}_k = \mathbf{p} * \mathbf{c}_k = \left[p_{k0}\ \ p_{k1} \cdots p_{kNv}\right]^T{}_{(Nv+1\times 1)}$$

$$\mathbf{w}_m = \left[w_{m_0}\ \ w_{m_1} \cdots w_{m(Nn_f-1)}\right]^T{}_{(Nnf\times 1)}$$

$$\mathbf{x}_m(n) = \left[x_m(n-n_f-v+1) \cdots x_m(n-1)\ \ x_m(n)\right]^T{}_{(nf+v\times 1)}$$

$$\mathbf{w}_{b_{km}} = \left[0\ 0 \cdots 0\ w_{b_{km0}}\ \ w_{b_{km1}} \cdots w_{b_{km(n_b-1)}}\ \ 0 \cdots 0\right]^T{}_{(N(nf+v)\times 1)} \quad k,m = 0,1,\ldots,N-1 \tag{3.1}$$

Where * is the convolution operator. In the above expressions w_{b_km} represents the feedback taps from the m^{th} slicer output to the k^{th} slicer input and the zeros in w_{b_km} correspond to indices 0 to Δ+1 and Δ+nb+2 to nf+ν assuming a delay of Δ for the system. Received signal at the k^{th} slicer input may be described as:

$$\begin{aligned} z_k(n) &= \sum_{m=0}^{N-1} \mathbf{x}_m^T(n)\mathbf{P}_k\ \mathbf{w}_m - \sum_{m=0}^{N-1} \mathbf{x}_m^T(n)\mathbf{w}_{b_{km}} \\ &= \sum_{m=0}^{N-1} \mathbf{x}_m^T(n)\left[\mathbf{P}_k \quad -\mathbf{I}_{(n_f+v)\times(n_f+v)}\right]\begin{bmatrix}\mathbf{w}_m \\ \mathbf{w}_{b_{km}}\end{bmatrix} \end{aligned} \tag{3.2}$$

where P_k is the convolution matrix corresponding to p_k. The above formulation is equivalent to removing the rows Δ+nb+2 to nf+ν of the matrix P_k and eliminating feedback taps w_{b_km} from the formulation. We will define Q_k to be the resulting matrix

---

[16] If $r_k(t)$ represents the mixer waveform for the k^{th} sub-channel, where $r_k(t+kT) = r_k(t)$ (k = 1, 2, … and T is the sub-channel period), h(t) represents the integrator impulse response, and the sampling points at the receiver are given by $KT+\Delta$ ($0 \le \Delta < T$), $c_k(t) = r_k(\Delta - t)h(t)$. The equivalency only holds at the sampling points.

[17] In practice link channels are very slowly time-varying, with variations caused by changes in the temperature, humidity and other environmental effects. Therefore, either the system requires adaptation at a very slow rate, or interference caused by channel variations should be budgeted into the receiver's voltage margin.

after the elimination of the rows from P_k and use the following simple formulation in the remaining sections:

$$z_k(n) = \sum_{m=0}^{N-1} \mathbf{x}_m^T(n) \mathbf{Q}_k \mathbf{w}_m \tag{3.3}$$

3.2.1 **Convex Formulation**

For convex optimization we will assume that all gain factors in the transmitter are absorbed in the equalizer taps and input symbols $x_m(n)$ have unit average energy. With this assumption, the minimum distance between the constellation points at the receiver for the k^{th} sub-channel is given by:

$$d_{\min_k} = \sqrt{\frac{12}{2^{2b_k}-1}} 1_\Delta^T \mathbf{Q}_k \mathbf{w}_k \tag{3.4}$$

where b_k is the number of bits transmitted on the k^{th} sub-channel. Interference power is also given by:

$$\sigma_{I_k}^2 = \sum_{m=0}^{N-1} \mathbf{w}_m^T \mathbf{Q}_k^T \mathbf{Q}_k \mathbf{w}_m - \left\| 1_\Delta^T \mathbf{Q}_k \mathbf{w}_k \right\|^2 \tag{3.5}$$

where 1_Δ is a column vector that is 1 at the $\Delta+1^{th}$ position and is zero otherwise. Let $offset_k$ represent the minimum resolvable voltage swing by the k^{th} slicer at the receiver and P_e represent the target BER. Following the formulation in [6] the BER constraint can be expressed as:

$$2\left(1-2^{-b_k}\right) Q\left(\frac{0.5 d_{\min_k} - offset_k}{\sigma_{noise}}\right) \le P_e \tag{3.6}$$

or equivalently:

$$Q^{-1}\left(\frac{P_e}{2\left(1-2^{-b_k}\right)}\right) \sqrt{\sigma_{I_k}^2 + \sigma_{Thermal}^2} - \sqrt{\frac{3}{2^{2b_k}-1}} 1_\Delta^T \mathbf{Q}_k \mathbf{w}_k + offset_k < 0 \tag{3.7}$$

For more accuracy the effect of residual ISI (σ^2_{Ik}) may be broken to a peak distortion part and a Gaussian part [6]; however, we skip that stage here for brevity.

The peak voltage constraint at the transmitter output at time nT+(i/N)T can be described as:

$$\sum_{m=0}^{N-1} x_m^T(n)\mathbf{w}_m^i < \sum_{m=0}^{N-1} Max\left(x_m^T(n)\right)\left|\mathbf{w}_m^i\right| < \sum_{m=0}^{N-1}\sqrt{\frac{3\left(2^{b_m}-1\right)}{2^{b_m}+1}}\left|\mathbf{w}_m^i\right| = V_{Tx\_Max}^i \tag{3.8}$$

where w^i_m is the i^{th} poly-phase of w_m. Finally equalizer coefficients can be obtained through the following Second Order Conic optimization problem:

$$\begin{aligned} &Minimize\ V_{Peak} \\ &Subject\ to: \\ &BER_k < P_e \qquad for\ k = 1,2,...,N \\ &V_{Tx\_Max}^i < V_{Peak} \qquad for\ i = 1,2,...,N \end{aligned} \tag{3.9}$$

3.2.2 **BER Constrained ZFE-DFE Solution**

Even though the convex formulation of (3.9) provides the optimal equalizer taps, it cannot be easily employed in systems which require adaptation. Since most adaptive solutions are based on MMSE or ZF optimization, in this section we derive the ZFE-DFE solution for the system.

The important point to note is that in a system with transmit equalization, the optimum received signal level is a variable and cannot be treated as a constant. Let's assume that the equalizer taps are set so that the gain of the main tap for the k^{th} sub-channel is given by g_k. The error in the k^{th} sub-channel is:

$$e_k(n) = z_k(n) - g_k x_k(n-\Delta) = \left(\sum_{m=0}^{N-1}\mathbf{x}_m^T(n)\mathbf{Q}_k\,\mathbf{w}_m\right) - g_k\mathbf{x}_k^T(n)\mathbf{1}_\Delta \tag{3.10}$$

Defining the cost as the sum of the Mean Square Error (MSE) at all slicer inputs:

$$\sigma^2 = \sum_{k=0}^{N-1} \sigma_k^2 = \sum_{m=0}^{N-1} \left[\mathbf{w}_m^T \left(\sum_{k=0}^{N-1} \mathbf{Q}_k^T \mathbf{Q}_k \right) \mathbf{w}_m - 2 g_m 1_\Delta^T \mathbf{Q}_m \mathbf{w}_m + g_m^2 \right] \tag{3.11}$$

The term in the brackets in the above equation represents the total interference caused by the m[th] transmitter at all the receivers. Minimizing the above cost function leads to minimizing the total interference caused by the individual transmitters independent of the others:

$$\mathbf{w}_m = g_m \left(\sum_{k=0}^{N-1} \mathbf{Q}_k^T \mathbf{Q}_k \right)^{-1} \mathbf{Q}_m^T 1_\Delta \tag{3.12}$$

Using the sum of MSE as the cost function is an approximation, and therefore inferior to convex modeling, because it ignores the fact that different channels have different constellations and can tolerate different levels of interference.[18] However, this approximation leads to a result that has an interesting practical implication: the equalizer taps for a given sub-channel can, up to a scale factor, be obtained independent of other sub-channels. Therefore, in a real system they can be adapted when all other sub-channels are turned off and the corresponding gm is set to an arbitrary number (which leads to an open eye).

Now we would like to find the power allocation coefficients g_m in a joint manner for all the sub-channels. It is again possible to follow the procedure in the previous section and form an SOC problem in terms of g_m coefficients; however, a simpler solution is possible if we approximate the effect of residual interference as peak distortion. In other words, we make the assumption that since target BER is very small, the residual interference taps are very likely to add constructively and in the worst case reduce the eye-opening at the receiver by the sum of the absolute of all the taps. Peak distortion caused by interference at the k[th] slicer input can be expressed as:

---

[18] In practice it is possible to include different constant weights for the different ISI and ICI terms in the MSE cost function to achieve better performance using the same ZFE-DFE solution presented here. The constant weights can be found, for example, by solving the convex problem of (9) and finding the relative power of the ISI and ICI terms in the optimal setting. Since channel characteristics do not change significantly by time, the weight coefficients can be calculated only once at beginning and programmed to the adaptive system.

$$V_{ISI_k} = \sum_{m=0}^{N-1} g_m \beta_{km} \tag{3.13}$$

where:

$$\beta_{km} = \sqrt{\frac{3\left(2^{b_m}-1\right)}{2^{b_m}+1}} \left\| \mathbf{Q}_k \left(\sum_{k=0}^{N-1} \mathbf{Q}_k^T \mathbf{Q}_k \right)^{-1} \mathbf{Q}_m^T 1_\Delta \right\|_1 \tag{3.14}$$

and is constant. The BER constraint of (3.6) therefore reduces to:

$$Q^{-1}\left(\frac{P_e}{2\left(1-2^{-b_k}\right)}\right)\sigma_{Thermal} + offset_k < \sqrt{\frac{3}{2^{2b_k}-1}} g_k 1_\Delta^T \mathbf{Q}_k \left(\sum_{m=0}^{N-1} \mathbf{Q}_m^T \mathbf{Q}_m \right)^{-1} \mathbf{Q}_k^T 1_\Delta - \sum_{m=0}^{N-1} g_m \beta_{km} \tag{3.15}$$

The peak voltage at time $nT+(i/N)T$ at the transmitter output can also be described as:

$$V_{Tx}^i = \sum_{m=0}^{N-1} g_m \alpha_m^i \tag{3.16}$$

where α^i_m are constant and can be obtained by inserting poly-phases of the taps in (3.12) into (3.8) (for g_m=1). Using these constraints, the SOC problem of (3.9) is a linear programming problem:

$$\begin{aligned} &Minimize\ V_{Peak} \\ &Subject\ to: \\ &\mathbf{Bg} > \mathbf{b} \\ &\mathbf{Ag} < V_{Peak}\mathbf{1} \end{aligned} \tag{3.17}$$

In the above problem B is an NxN matrix with entries $B_{mk} = -\beta_{km}$ if k≠m and $B_{mk} = \sqrt{\frac{3}{2^{2b_k}-1}}\,\tilde{\beta}_{km}$ if k=m, A is an NxN matrix with entries $A_{km} = \alpha_{km}$, and b is a column vector whose entries b_k are equal to the left hand side of (3.15). By writing the Lagrangian for problem (3.17), it is easy to verify that the minimum V_{peak} (if it exists) is achieved when **Bg = b** or equivalently **g = B^{-1}b**. This means that with peak distortion approximation for interference, optimum power allocation is achieved when

all the receivers meet the BER constraint with equality. Therefore, in a real system once the equalizer taps for all the sub-channels are adapted, the power allocation coefficients gm can be obtained by measuring "eye closures" at the slicer inputs to form the B matrix followed by inverting the matrix.

3.2.3 **Jitter**

The uncertainty in the phase of the clock is one of the major performance limiting factors in high-speed links. The jitter in the transmitter clock can extend or shorten the transmit pulse duration at the output of the Digital to Analog Converter (DAC). Such edge modulation, when passed through the channel, would manifest itself as correlated noise at the sampling instant. On the other hand, the uncertainty in the receiver sampling clock would lead to sampling at sub-optimal points and therefore degrades the SNR. A thorough analysis of jitter for BB links was originally presented in [35]. In this section we extend that formulation to the case of MIMO systems and provide a more compact formulation.

3.2.3.1 **TX jitter**

As shown in [35] the effect of transmitter jitter can be modeled as Figure 3.6(a). In this figure r_0, r_1 … r_{N-1} represent the noise samples at the slicer inputs and h is the impulse response of the channel. Intuitively, when a pulse at the output of the DAC (y) is extended, the extended portion of the pulse occupies the place of the following pulse. For small jitter variance, the net effect may be modeled as an impulse noise at the transition point with magnitude of $\phi(n)(y(n) - y(n-1))$. From noise modeling perspective, this is equivalent to a filter s = [1 -1] placed at the DAC output followed by a multiplicative noise $\phi(n)$. This time-varying multiplicative noise creates a linear time-varying (LTV) channel between the transmitters and the receivers. Following our notation in the earlier parts of this section, noise at the k^{th} slicer input can be described as:

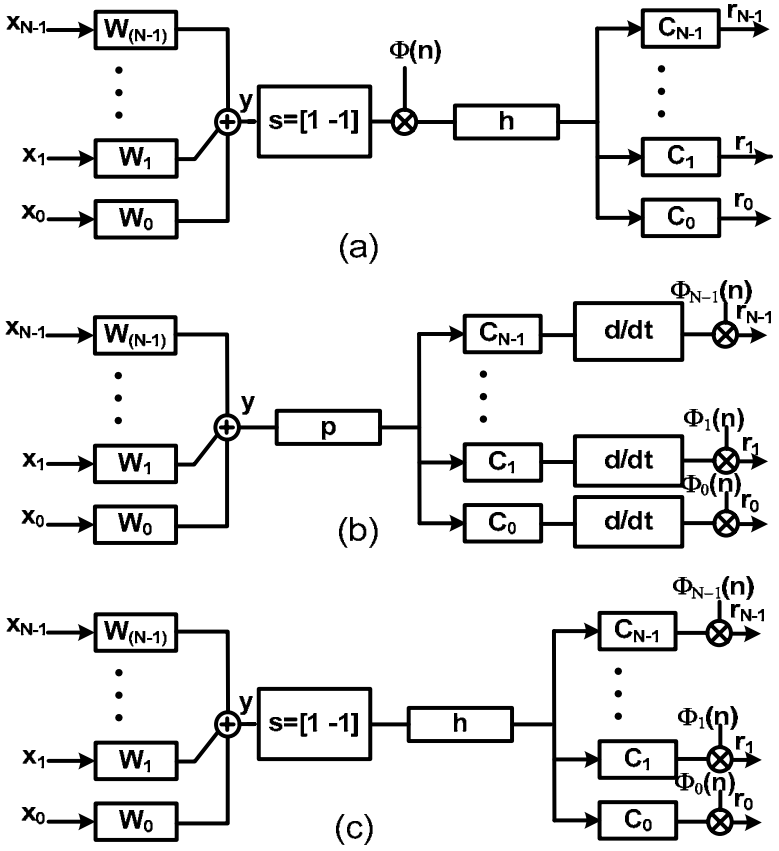


Figure 3.6: (a) Transmitter jitter model. (b) Receiver jitter model (c) Alternative receiver jitter model.

$$r_k(n) = \sum_{m=0}^{N-1} \mathbf{h}_k * \left[\phi(n) \times (\mathbf{s} * \mathbf{w}_m * \mathbf{x}_m(n))\right] = \sum_{m=0}^{N-1} \mathbf{x}_m^T (\mathbf{H}_k \bullet \boldsymbol{\psi}) \mathbf{S} \mathbf{w}_m \tag{3.18}$$

where $\bullet$ is element by element matrix multiplication, S is the convolution matrix corresponding to s, h_k and H_k are similar to p_k and P_k in (3.1) and (3.2) but correspond to impulse response instead of the pulse response of the channel, and $\boldsymbol{\psi}$ is the jitter matrix whose elements can be described by $\boldsymbol{\psi}_{ij} = \phi(n - i + j)$. It can be shown that the jitter noise variance is:

$$\sigma^2_{TX\_jitter_k} = \sum_{m=0}^{N-1} \mathbf{w}_m^T \mathbf{S}^T (\mathbf{H}_k^T \mathbf{H}_k \bullet \mathbf{R}_\Phi) \mathbf{S} \mathbf{w}_m \tag{3.19}$$

where $\mathbf{R}_{\Phi}$ is the autocorrelation matrix of the jitter.

3.2.3.2 **RX jitter**

The difference between the signal at the optimum sampling point and the signal the actual sampled point may be considered as receiver jitter noise. The deviation from the optimum point is equal to the slope of the received waveform times the deviation to the first order. Approximating the slope with finite difference (d/dt) at a fine sampling resolution, the model of Figure 3.6(b) can be obtained. Alternatively, the differentiator operator can be moved into the channel to convert the pulse response, p, to impulse response, h, convolved with s = [1 -1], Figure 3.6(c)[19]. Again:

$$\begin{aligned} r_k(n) &= \phi_k(n) \times \sum_{m=0}^{N-1} \mathbf{h}_k * \mathbf{s} * \mathbf{w}_m * \mathbf{x}_m(n) \\ &= \phi_k(n) \times \sum_{m=0}^{N-1} \mathbf{x}_m^T \mathbf{H}_k \mathbf{S} \mathbf{w}_m \end{aligned} \tag{3.20}$$

and the jitter variance is:

$$\sigma^2_{RX\_jitter_k} = \sigma^2_{\phi_k} \left(\sum_{m=0}^{N-1} \mathbf{w}_m^T \mathbf{S}^T \mathbf{H}_k^T \mathbf{H}_k \mathbf{S} \mathbf{w}_m \right) \tag{3.21}$$

Clearly both (3.19) and (3.21) are convex in equalizer taps and can therefore be incorporated in the BER constraint (3.7) of the SOC formulation, or in the cost function (3.11) of the ZF optimization.

---

[19] The models in Fig 7(b),(c) assume that the clock used to generate the carrier signals for the mixers is fully correlated with the clock used for starting and ending the integration period. In other words, it is implicitly assumed that the mixer cycle period always remains aligned with the integration cycle despite jitter, which justifies the use of the matched filter model for the receiver. Alternatively, in order to only model the effect of sampler jitter, it can be shown that the model of Figure 7(c) can be modified to have s = [1] and $c_k(t) = g_k(t)h'(t)$ where $g_k(t)$ is the mixing waveform for the k^{th} sub-channel and h'(t) is the derivative of the integrator impulse response.

3.3 Performance Analysis and Simulation Results

The AMT system can improve the performance of high-speed links in two ways: through better transmit power allocation, and by parallelizing the receiver front-end, leading to better receiver sensitivity. Using the convex analysis framework developed in the previous section, we will investigate these two possibilities using realistic channel models from different applications.

3.3.1 Power Allocation Gain

Figure 3.7(a) shows a multi-drop configuration where multiple memory modules are connected to two CPUs. C_i in this figure represents the capacitive loading of the devices and we assume that the memory modules are not terminated. In this configuration when the two CPUs communicate, the memory modules act as reflection generators. Therefore, notches in the frequency response of the channel are part of the characteristics of this system. However, the length and the characteristic impedance of the stubs connecting the modules to the main trace can be tuned such that all the stubs resonate at identical frequencies. Furthermore, since the stubs are pieces of waveguides, and therefore have periodic frequency responses, notches repeat at equally-spaced frequencies. This characteristic is ideal for AMT because the two constraints placed on the system to make it LTI result in equal bandwidth for all the sub-channels. Figure 3.7(b) shows the corresponding frequency responses for two values of C_i. Even though capacitances don't have periodic frequency responses, as long as the characteristics of the system are dominated by stubs, the notch frequencies remain (approximately) equally spaced.

Analysis of a 3-channel AMT system with 2-PAM sub-channels over the channel of Figure 3.7 (C_i = 1pF) indicates that the maximum achievable data rate with AMT is 5.25-Gb/s. A 2-PAM BB system with comparable equalization complexity and same transmit peak voltage, targeting same BER of 10^{-15}, only achieves 3.0-Gb/s. The AMT system can achieve 75% higher data rate by placing two 2-PAM sub-channels (or equivalently a 4-QAM sub-channel) after the first notch frequency.

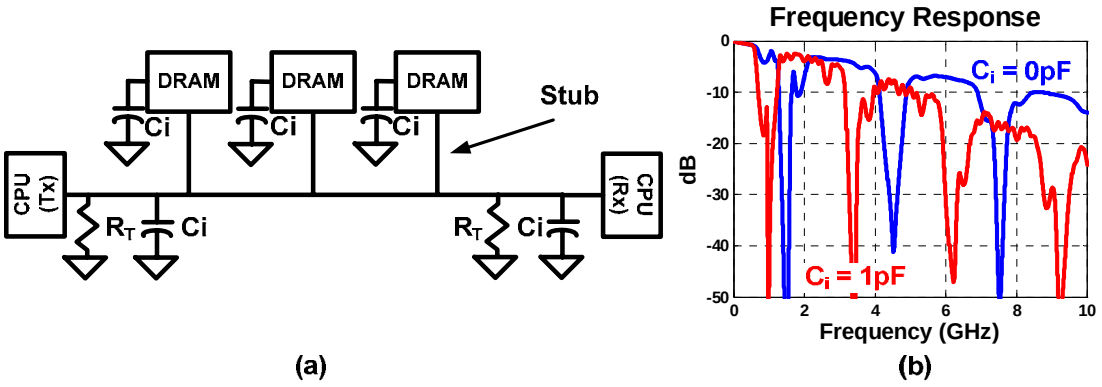


Figure 3.7: (a) A generic multi-drop configuration, where three memory modules are connected to two CPUs. Only the two CPUs have termination resistors (b) Corresponding channel frequency responses for 0pF and 1pF of parasitic capacitance per device.

The above example demonstrates the other difference between high-speed link systems and DSL or wireless systems. In link systems, the communication channel can be engineered and optimized together with the transmission algorithm within the physical constraints of the system to achieve optimum performance. In this example, the length and the impedance of the stubs were set such that they all resonate at same frequencies. A further step is to add additional stubs to the system to dominate the characteristics of the channel and tune it to the AMT system characteristics [36].

3.3.2 **Parallelization Gain**

Even over channels where power allocation gain is small, AMT can improve link performance by fundamentally parallelizing the receiver front-end, leading to better receiver sensitivity. Receiver sensitivity, which was modeled by the term "*offset*" in the formulations in previous section, represents the smallest voltage that is resolvable by the receiver. The value of "*offset*" typically varies in the range of 3mV to 20mV in different implementations of baseband links, while the root mean square thermal noise contribution is generally below 1mV. Therefore, improving receiver sensitivity can significantly improve link performance.

Figure 3.8(a) shows a 2PAM BB receiver operating at rate R utilizing a linear peaking amplifier and a DFE. The peaking amplifier partly compensates for the

magnitude distortion of the channel and improves receiver sensitivity. Generally a gain-bandwidth tradeoff exists for the design of the amplifier. The DFE has to subtract a weighted sum of the received symbols from the incoming signal. In particular, to close the feedback loop for the first tap, the entire operation of detecting the current symbol, multiplying it with the appropriate weight, and subtracting it from the incoming symbol should be performed in less than a symbol period, which is 200ps in a 5-GS/s link. As was mentioned in 0, in order to alleviate this constraint most BB links resort to loop-unrolling techniques [18][20], where two decisions are made for the next incoming bit assuming the current bit is +1 and -1. The correct decision is then chosen once the value of the current bit is known (Figure 3.8(b)). Therefore, a loop-unrolled DFE utilizes two samplers sampling at rate R with thresholds placed at $\pm\tilde{\alpha}$. The sensitivity of such samplers is adversely affected by the noise on the reference voltages.

Figure 3.8(c) shows the receiver of a 2-channel (2PAM, 2PAM) AMT link with a throughput of R. In this architecture, a peaking amplifier with a bandwidth of R/4 is placed only in the high-pass branch where it is needed. The AMT amplifier has half the bandwidth of the BB amplifier, and consequently, potentially higher gain. In addition, the DFE loop time increases by a factor of two compared to the BB DFE due to the inherent parallelization in AMT, and the two samplers operate at R/2 with thresholds placed at zero. Consequently, the AMT samplers operate at half the rate of BB samplers and do not require non-zero reference voltages. As a result, overall the AMT receiver can potentially achieve better sensitivity compared to a loop-unrolled BB link. Even though the actual gain is very implementation dependent, using the optimization framework we can evaluate the potential improvement in performance if the improvements in sensitivity are realized.

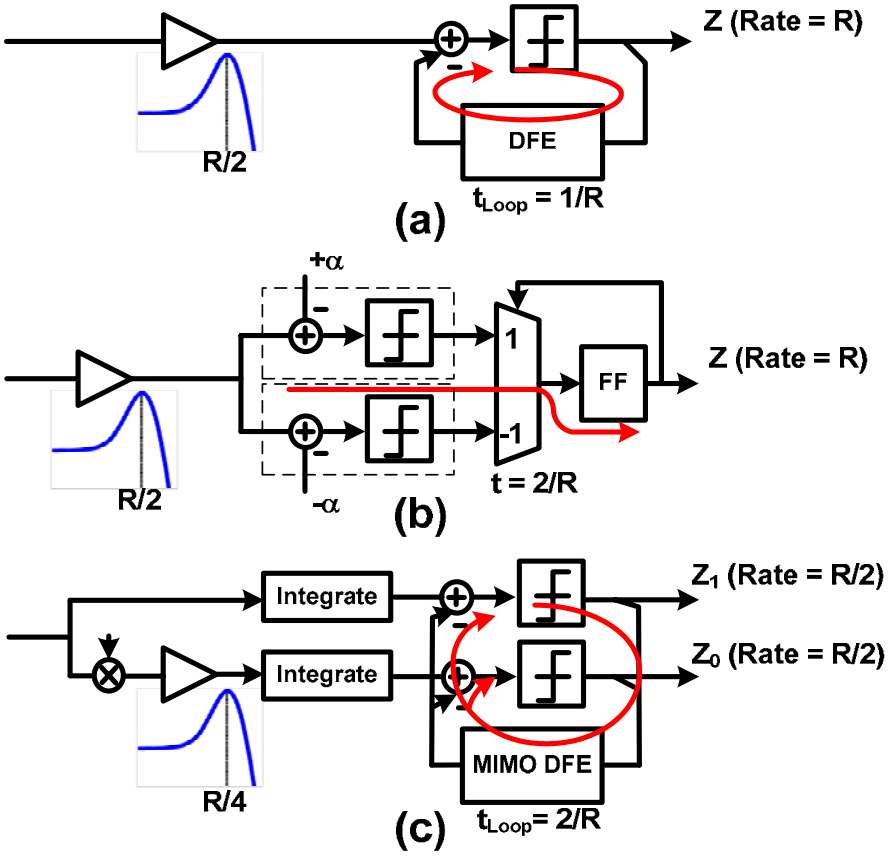


Figure 3.8: (a) A BB 2PAM receiver with a peaking linear amplifier and a DFE operating at rate R (b) A loop-unrolled BB 2PAM receiver. α represents the first DFE tap value. (c) A 2-channel (2PAM, 2PAM) AMT receiver operating at rate R.

Figure 3.9(a) shows the frequency response of 6 typical backplane channels. These channels correspond to 3", 10" and 20" NECLO PCB traces routed on either the top layer or the bottom layer of a PCB. All these channels all have relatively smooth frequency characteristics. Figure 3.9(b) compares the performance of a 2PAM BB system with a 2-channel (2PAM, 2PAM) AMT system in terms of the minimum required peak transmit voltage if they both operate over the channels of Figure 3.9(a), indexed from 1 to 6. Assuming that the samplers in the AMT system have better sensitivity, data in Figure 3.9(b) indicates that for same target BER of 10^{-15}, at the same peak transmit voltage that the BB system can support 6-Gb/s, the AMT system can support 9-Gb/s data transfer rate. Therefore, with just an overhead of a mixer and two integrators the AMT system can potentially achieve 50% improvement in data

rate by parallelizing the receiver structure. The mixer and the integrators do not constitute a large portion of the system power, and their power consumption overhead is (to some extent) offset by the half-rate samplers (as opposed to the full-rate samplers in the BB system). However, if the parallelization gain does not translate to better sensitivity due to other implementation issues, Figure 3.9(c) shows that the AMT system requires larger transmit voltage, which is due to its higher Peak to Average Power Ratio (PAPR).

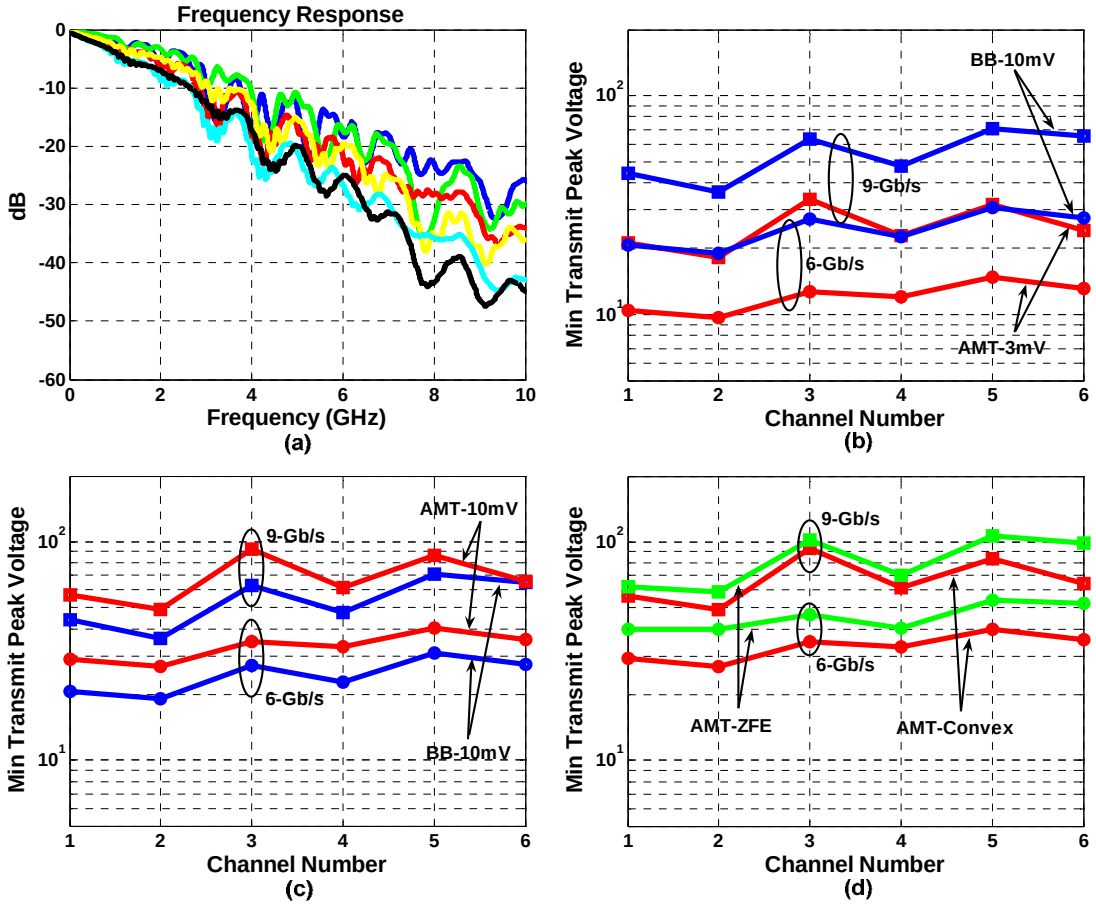


Figure 3.9: (a) Frequency responses of 6 NELCO backplane channels: 3", 10" and 20", routed on top or bottom layers of a PCB (b) Minimum required transmit peak voltage to achieve BER of 10^{-15} at 6-Gb/s and 9-Gb/s for 2PAM BB and 2-channel AMT systems, assuming AMT system has more sensitive samplers. Similar equalization complexity for both systems (c) Same as (b), but assuming samplers in both AMT and BB have same sensitivity of 10mV (d) Comparison of BER-constrained ZFE and Convex solutions for the AMT system under same conditions as (c).

Figure 3.9(d) compares the performance of the suboptimal BER-constrained ZF solution proposed in Section III with the optimal solution obtained from the convex framework proposed in the same section in terms of the minimum required transmit peak voltage under the conditions of Figure 3.9(c). The results indicate that the ZFE solution is reasonably close to the optimal. The difference between the two solutions increases as the number of sub-channels is increased, and in cases were substantial residual interference exists in the optimized system.

3.4 Summary

The AMT system presented in this chapter is an extension of a BB link to a MIMO system. As a result, the techniques utilized to mitigate the effect of interference are largely based on what is conventionally used in high-speed links. As a result, an AMT system can be implemented at link data rates without substantial increase in the total system power and complexity. In addition, the characteristics of an AMT system with only a small number of wideband sub-channels is tuned to the characteristics of link systems with only a small number impedance discontinuities in their signal path. Consequently, the AMT system can provides most of the benefits that a MT system can provide in such environments. Channel engineering and tuning the impedance discontinuities (instead of eliminating them) can additionally boost the performance of the system. Furthermore, the AMT system effectively parallelizes the receiver front-end, enabling more efficient circuit design.

Chapter 4

AMT Variations

The AMT system in its most general form requires feed-forward equalization at the transmitter, and mixing, integrations, and decision feedback equalization at the receiver to recover the individual sub-channels and eliminate ISI and ICI. However, if make certain assumptions about the channel characteristics, it is possible to find alternative implementations of the algorithm which ultimately result in better performance and less complexity.

As we described in 0, for the class of controlled-ISI channels categorized as partial-response channels, simple encoding of the data at the transmitter can result in considerable complexity reduction. Partial-response signaling can be particularly interesting for link systems, because channel characteristics can be controlled and changed to closely match a desired response [36]. In this chapter, we will describe how partial-response signaling and AMT can be merged together to extend the data rate of conventional link systems.

As we described in Chapter 3, one possible application for AMT is a multi-drop memory interface, where multiple devices are connected to a single trace. An important characteristic of memory interface links is that the DRAM side should be simple and compact, due to the limited number of metal layers on a DRAM process, the lower transistor speeds of a DRAM processes, and high target manufacturing yields. As a result, most of complexity in DRAM links is placed at the memory controlled side. DFE, for example, is typically not used at a DRAM receiver. In this

chapter, we also propose a variation of AMT receiver which is sample-based, rather than integrating, and is suitable when no DFE exists in the system.

4.1 Duobinary AMT with Channel Engineering

BB partial response methods mainly exploit the part of channel bandwidth before the first notch in the frequency. However, for example, in the case of a duobinary channel with a frequency response of $1+e^{-j2\pi fT}$, channel frequency response bounces back to non-zero values after the first notch frequency. In fact notches in the frequency response happen at equally spaced distances at (2k+1)/2T and additional channel bandwidth exists for signal transmission between every two notches, which is not utilized by duobinary transmission.

Now consider AMT transmission over the same channel, where each of the sub-channels is operating at rate T. The partial-response channel delays the combined output of the transmitter by one full sub-channel period T, and it to itself. However, since in an AMT system, all the sub-channel carrier frequencies are integer multiples of multiples of the sub-channel symbol rate, this operation does not affect the orthogonality between the sub-channels. Consequently, if duobinary precoding is performed on the sub-channels at the transmitter, they can be independently detected at the receiver after mixing and integration without any ISI or ICI. As a result at doubinary AMT can be an ideal candidate for systems with one dominant stub.

One remaining issue is that in link systems, it is often the case that the response of the channel is affected by multiple stubs caused by via stubs and package impedance discontinuities. For example, Figure 4.1 shows a chip-to-chip link where a dominant plating stub exists on one of the packages. A plating stub is a side-effect of package manufacturing process, and its removal involves a costly post manufacturing process.

Figure 4.1: A chip to chip system with a dominant stub on the package of chip B.

Figure 4.2 shows the frequency response of the link channel together with it loss tangent. Careful inspection of the frequency response indicates that the channel response starts to bounce back to reach the loss tangent after the first notch frequency (caused by the plating stub); however, the response is dampened due to the existence of other discontinuities in the signal path. If it wasn't for these other discontinuities, the response of the channel would bounce back to have less than 20dB attenuation below 20GHz. Therefore, part of the channel's useful transmission capacity is wasted beyond the first notch as a result of the interaction between the dominant and non-dominant stubs.

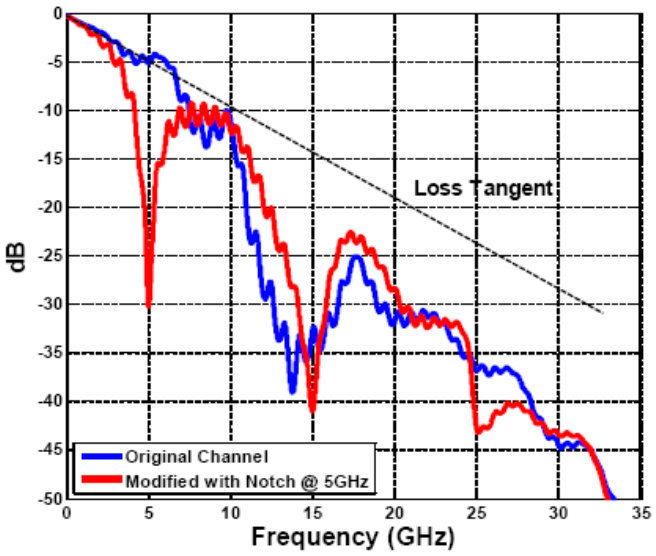


Figure 4.2: Frequency response of the original chip-to-chip channel, and the modified response after the length of the plating stub was increased.

We note that even though it is costly to reduce the length of the plating stub, increasing its length can be achieve very easily by extending the trace on the package.

Extending the length of the stub moves the dominant notch frequency to lower frequencies as shown in Figure 4.2. Even though this modification reduces the bandwidth of the channel before the first notch frequency, it allows the channel response to bounce back to the loss tangent, before being dampened by other discontinuities. Since now the frequency response of the channel is mostly determined by a single stub over useful frequencies, and stubs have periodic frequency responses, a second notch is created at 15GHz. As a result, a whole passband channel is opened up between 5GHz and 15GHz for pass-band signal transmission using an AMT transmitter. The AMT transmitter consists of a 10-Gb/s duobinary baseband channel, and two quadrature 10-Gb/s duobinary passband channels to achieve an aggregate data rate of 20-Gb/s. Since the modified channel does not completely bounce back to 0dB, as an ideal duobinary channel would do, the AMT system requires certain amount of transmit equalization or alternatively would need a linear equalizer with 10dB of gain at 10-GHz at the receiver front-end.

4.2 Sampler-Based AMT Receiver

The AMT architecture that was proposed in Chapter 3 includes mixers, integrators and a MIMO DFE. In a class of applications (for example memory interface) the receiver may need to be very simple. Therefore, a MIMO DFE is avoided and all the equalization in performed in the transmitter. In addition, all sub-channels are kept to 2-PAM. For these applications, a variation of the AMT receiver is possible that does not require mixing and integration. Instead, the proposed architecture employs at-rate (at system throughput) samplers per sub-channel to jointly detect all the sub-channels at the same time. We will describe the architecture for a 2-channel and a 4-channel sampler-based AMT receiver in this section.

4.2.1 **2-Channel Sampler-Based AMT Receiver**

Figure 4.3 shows a digital implementation of 2-channel AMT receiver with an ADC at the receiver front-end, and mixing and integration performed in the digital domain. It is assumed that the receiver does not require DFE.

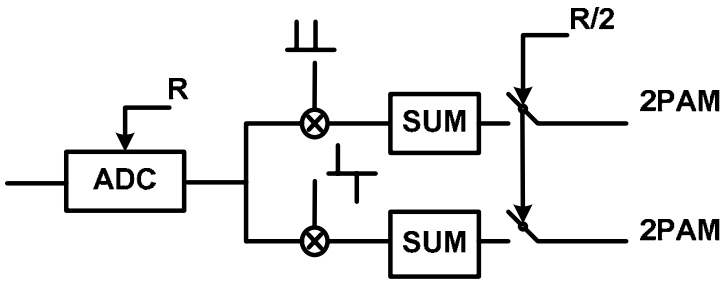


Figure 4.3: A digital implementation of a 2-channel AMT receiver.

For perfect detection of the two sub-channels, the received signal sequence at the input to the receiver of Figure 4.3 should be the summation of the two sequences shown in Figure 4.4. In this figure $x_{1,j}$ represents the transmitted sequence for sub-channel 1, and $x_{2,j}$ represents the transmitted sequence for sub-channel 2 (both possibly scaled by the channel). Let's assume that the signal level for the sub-channels is scaled by the channel to "a" and "b" for the 1st and 2nd sub-channels respectively and that "*a*" and "*b*" are large enough to be detected by a 1-bit sampler running at the throughput rate.

Figure 4.4 also shows the possible received signal values at different time instances. As evident from the figure, the received signal at any time-instant may have 4 values, similar to a 4-PAM constellation. However, without any assumptions on the relative magnitudes of "*a*" and "*b*", the points in the AMT constellation may not be equally distant. In fact, "*a*-*b*" and "-*a*+*b*" may turn out to have same values, reducing the number of points in the constellation to 3. As a result, it is not possible to detect the received signal with three samplers with equally spaced thresholds similar to a 4-PAM signal. Nonetheless, it is possible to utilize the structure of the received signal to perform something similar.

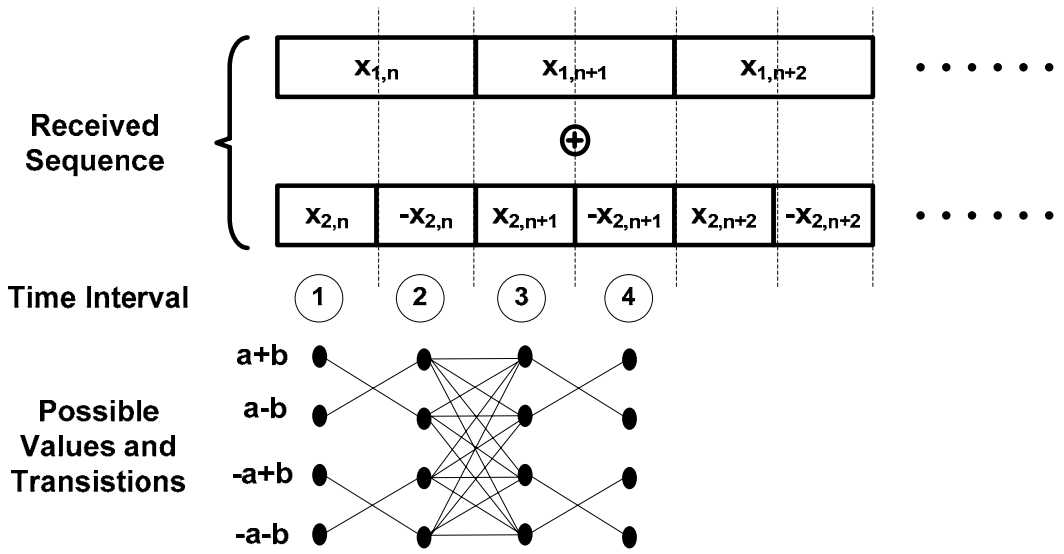


Figure 4.4: Decomposition of the received sequence at the receiver input with no DFE at the receiver. It is assumed a>b for this figure.

The first observation is that we can always detect whether we have received the "+*a*+*b*" point by placing a sampler with threshold at *+max(a,b)*. This is because the minimum distance between the point "+*a*+*b*" and all other points in the constellation is at least *2min(a,b)*. For the same reason, we can always detect whether we have received the point "-*a*-*b*" by placing a sampler with threshold at *-max(a,b)*. Therefore, the only two points we cannot detect easily are "-*a*+*b*" and "+*a*-*b*".

Figure 4.4 also shows the possible half-symbol transitions (1→ 2, 3 →4, etc.) of the AMT constellation. As opposed to a 4-PAM signal where all transitions are possible from one time instant to the next, the structure of the MT received signal sequence allows only a selected number of transitions at half-symbol transition points. For example, if the received signal at time "*1*" is "+*a*-*b*", the received signal at time "*2*" is guaranteed to be "+*a*+*b*". Therefore, by looking at the outputs of the two samplers described in the previous paragraph over two half-symbol periods we are able to completely decode the received signal. Figure 4.5 shows the architecture of the 2-channel AMT receiver based on this idea.

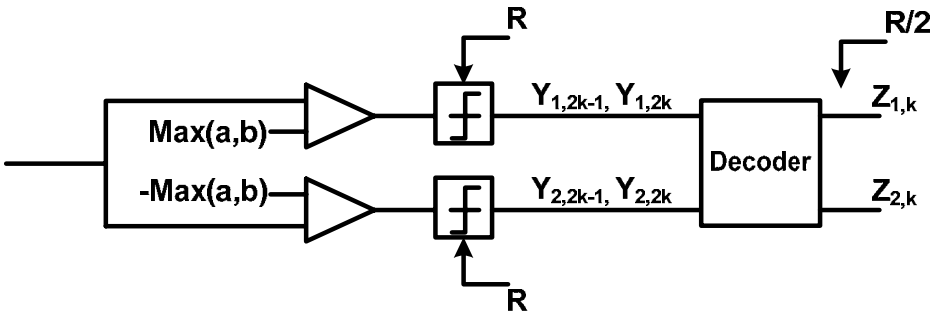


Figure 4.5: Sampler-based 2-channel AMT receiver.

The decoder block has the following truth table:

Table 4.1: 2-channel AMT decoder functionality. "X" represents a "don't care" value.

$Y_{1,2k-1}$	$Y_{2,2k-1}$	$Y_{1,2k}$	$Y_{2,2k}$	$Z_{1,k}$	$Z_{2,k}$
1	X	X	X	1	1
X	-1	X	X	-1	-1
X	X	1	X	1	-1
X	X	X	-1	-1	1

4.2.2 **4-Channel Sampler-Based AMT Receiver**

Figure 4.6 shows a digital implementation of a 4-channel AMT receiver.

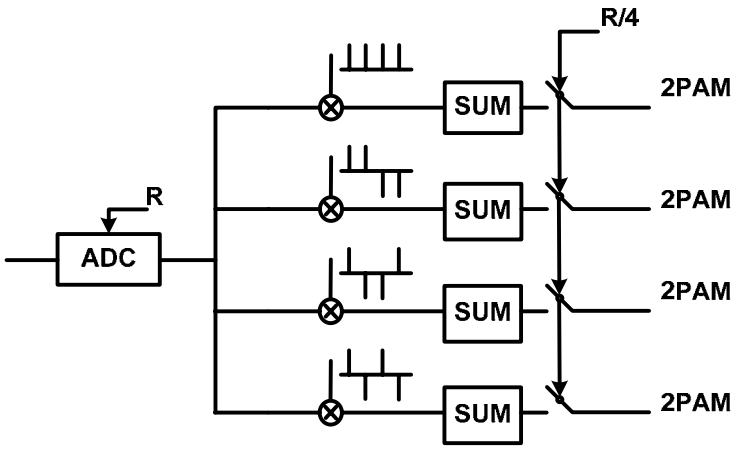


Figure 4.6: A digital implementation of a 4-channel AMT receiver.

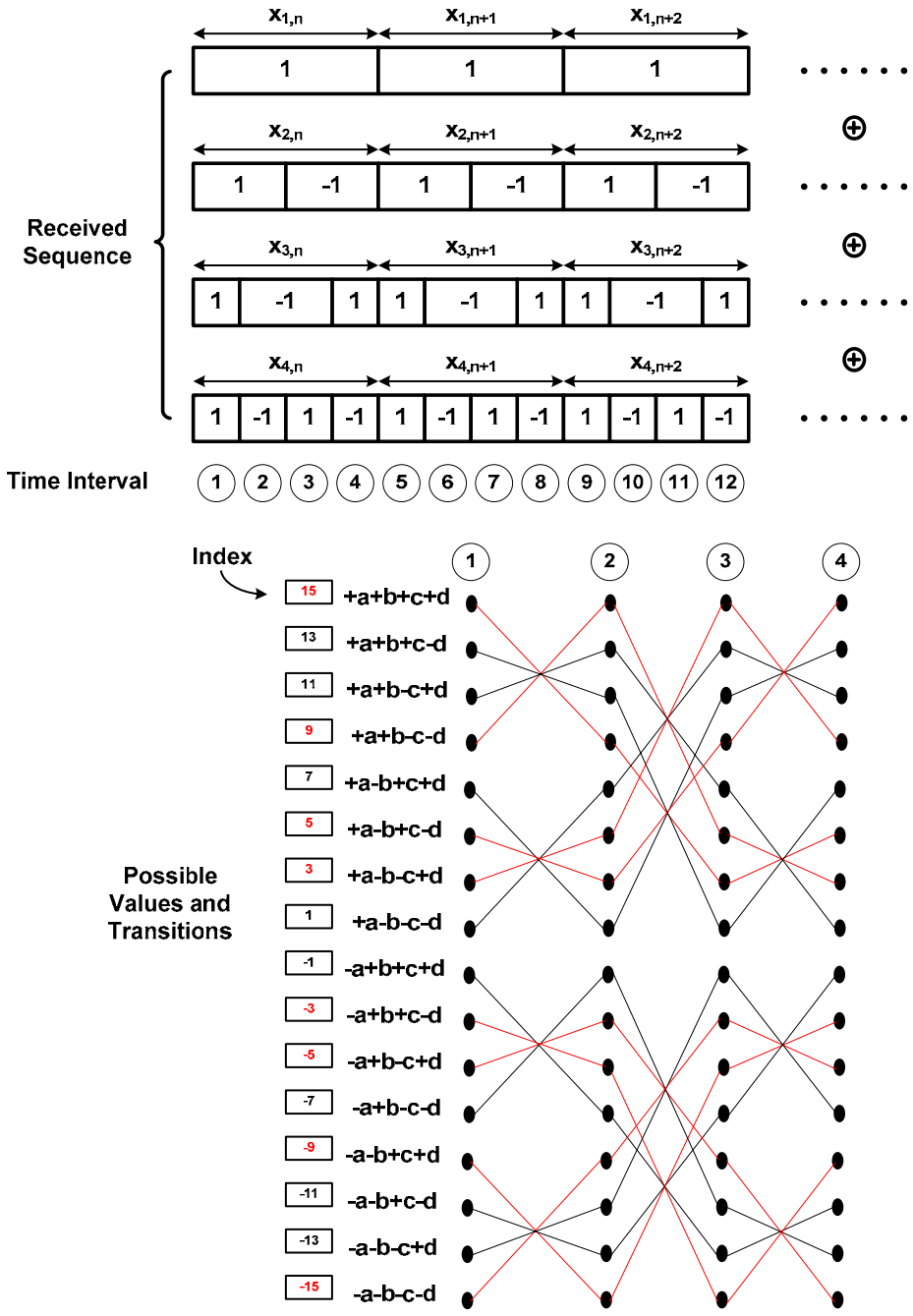


Figure 4.7: Received sequence pattern for perfect recovery decomposed to different sub-channels. It is assumed that the 1st sub-channel has a scale of "a", and the 2nd, 3rd and 4th sub-channels have scales of "b", "c", and "d" respectively. Also all possible transitions within one sub-channel symbol period are shown. Input sequences which make a transition to "+a+b+c+d" or "-a-b-c-d" are shown in red.

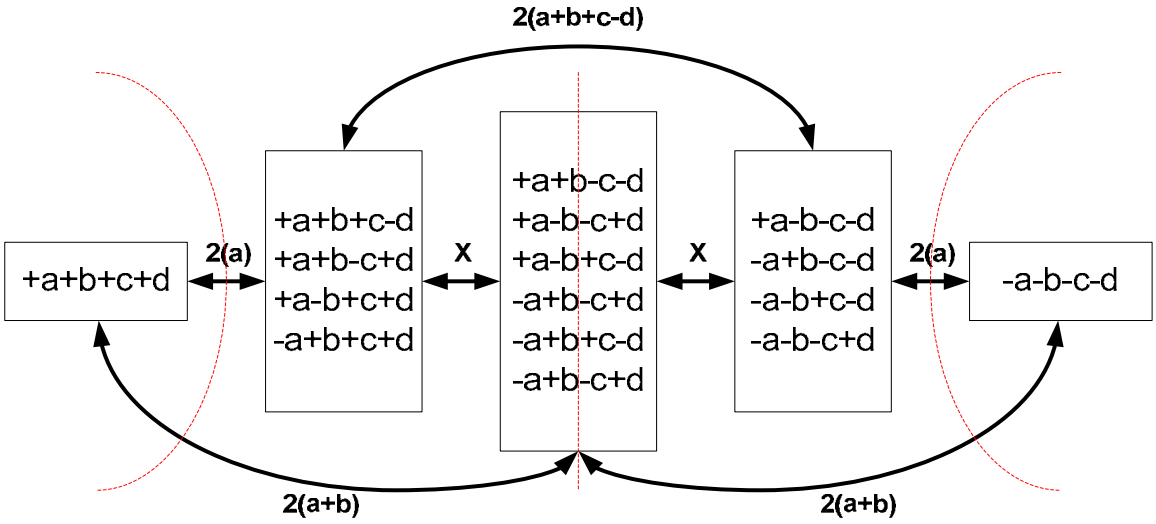


Figure 4.8: 16 received constellation points grouped into 5 sets. Minimum distance between the points also shown assuming a≤b≤c≤d.

Figure 4.7 shows the received sequence which leads to perfect detection decomposed to different sub-channel sequences. It is again assumed that no DFE exists in the receiver, and the received signal magnitude for the 1st, 2nd, 3rd, and 4th sub-channel is represented by "*a*", "*b*", "*c*", and "*d*" respectively. Without loss of generality we will assume $a \leq b \leq c \leq d$ in the following discussion. All possible transitions within one sub-channel symbol period are also shown in the same figure. Similar to the previous section, we will treat all possible received signal values as points in a 16-point constellation. Figure 4.8 shows the points grouped into 5 different bins with minimum distance between the bins also marked. "X" represents an unknown quantity; a quantity whose value depends on the exact relationship between "*a*", "*b*", "*c*", "*d*" and is not necessarily larger than *2min(a,b,c,d)*.

As figure shows, the minimum distance between the point "*+a+b+c+d*" and all other points is *2min(a,b,c,d)*. Therefore, a sampler with threshold at *+a+b+c+d - min(a,b,c,d)* can correctly detect whether the state "*+a+b+c+d*" has been visited. Same holds for the point "*-a-b-c-d*". A sampler placed at *-a-b-c-d + min(a,b,c,d)* can detect a transition to the state "*-a-b-c-d*". Among the 16 states shown in Figure 4.7, indexed from -15 to 15, the states -15, -9, -5, -3, 3, 5, 9, and 15 visit either "*-a-b-c-d*" or "*+a+b+c+d*" in exactly one time-slot (between 1 to 4) and the other 8 states do not.

We will call the former group of states the "*even*" group and call the latter group the "*odd*" group. The *even* group is shown in red in Figure 4.7. This group consist of the 4-bit transmit sequences with two or no "-1"s among them ("+*a*+*b*+*c*+*d*", "+*a*+*b*-*c*-*d*", "+*a*-*b*+*c*-*d*", etc). The *odd* group, on the other hand, consists of those points that have one or three "-1"s ("+*a*-*b*-*c*-*d*", "+*a*+*b*+*c*-*d*", "+*a*+*b*-*c*+*d*", etc). In the *odd* group, assuming $d \leq b+c$, half the points are larger than "*2a*" and half the points are less than "*-2a*". Therefore, if we know that the received point belongs to the *odd* group, a sampler placed at zero can correctly determine whether the point belongs to the positive half or the negative half. If we track the 8 points in the second group during the 4 time slots, we observe the following pattern:

Table 4.2: Sign of detected signal with a sampler placed at zero during the four quarter-symbol time slots for constellation points that belong to the odd group.

Index	**Slot 1**	**Slot 2**	**Slot 3**	**Slot 4**
13	+	+	-	+
11	+	+	+	-
7	+	-	+	+
1	-	+	+	+
-1	+	-	-	-
-7	-	+	-	-
-11	-	-	-	+
-13	-	-	+	-

Therefore, by looking at any three columns of the table, we can correctly decode the received constellation, if we know if belongs to the *odd* group. The assumption $b+c \geq d$ that is necessary for this decoding is not very restrictive because in an optimized system all "*a*", "*b*", "*c*", and "*d*" are supposed to be of the same order. Figure 4.9 shows the 4-channel AMT receiver architecture based on this concept.

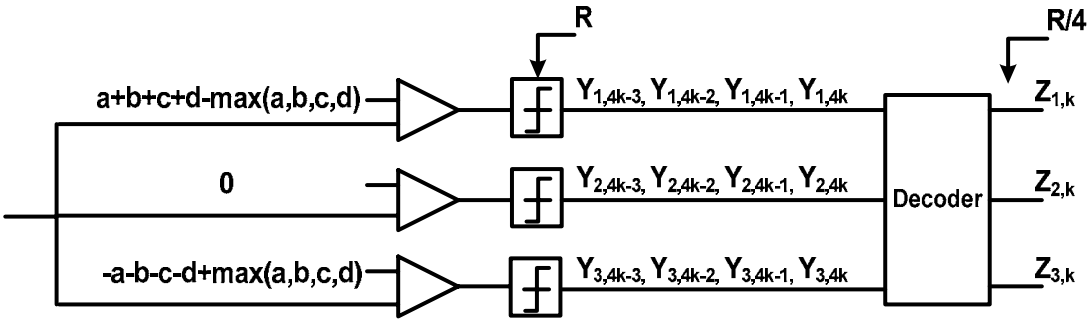


Figure 4.9: Sampler-based 4-channel AMT receiver.

The decoder truth table is as follows:

- If $Y_{1,4k-j} = 1$ or $Y_{3,4k-j} = -1$ for any of the time-slots (j = 1,2,3,4), then the received symbol belongs to the *even* group and can be decoded as follows:
    - $Y_{1,4k-j} = 1$
        - J = 1: (1,1,1,1)
        - J = 2: (1,1,-1,-1)
        - J = 3: (1,-1,-1,1)
        - J = 4: (1,-1,1,-1)
    - $Y_{4,4k-j} = -1$
        - J = 1: (-1,-1,-1,-1)
        - J = 2: (-1,-1,1,1)
        - J = 3: (-1,1,1,-1)
        - J = 4: (-1,1,-1,1)
- If $Y_{1,4k-j} = -1$ and $Y_{4,4k-j} = +1$ for all time-slots (j = 1,2,3,4), then the received symbol belongs to the *odd* group.
    - Decode based on Table 4.1.

4.3 Summary

We showed that for certain applications, alternative implementations of the AMT system are possible. In particular, the combination of channel engineering with AMT to optimize system performance can be a powerful technique. We also showed how simple decoding techniques can be applied to make the architecture of the AMT receiver very similar to a baseband receiver. The technique can be extended to implement 3-channel or duo-binary sampler-based AMT systems. The down-side, however, is that since sampler-based architectures make hard-decisions on the received signal, they have inferior performance to the corresponding integrator-based receivers from random noise SNR perspective. The loss in SNR for 2-channel and 4-channel systems is 3-dB and 6-dB, respectively.

Chapter 5

Experimental System

In Chapter 3 we showed that an AMT transmitter is a generalized version of a parallelized BB transmitter. This observation can therefore be applied to implement a transmitter that supports both AMT and BB signaling with little overhead. Such a transmitter can achieve the best performance over a very wide range of channel characteristics using either AMT or BB transmission depending on which one is optimal.

This chapter describes the design of a 24-Gb/s software programmable transmitter, implemented in a 90-nm CMOS technology, which supports 4-channel and 2-channel AMT as well as a variety of BB transmission modes including 2-PAM and 4-PAM. With 16 effective FFE taps, and each tap coefficient having a full 10-bit dynamic range, the transmitter has sufficient equalization capabilities to enable the study of both AMT and BB transmission algorithms over a wide range of environments and applications. In other words, the transmitter is an extremely flexible test instrument for high-speed link applications.

The high-speed transmitter employs a digital linear equalizer and a 12-GS/s 8-bit Digital-to-Analog Converter (DAC). We start this chapter by describing the architecture of the transmitter and continue by presenting some measurement results in AMT and multi-PAM transmission modes. The architecture of the equalizer also enables compensation of analog distortions caused by the mismatch in the paths of the on-chip time-interleaved DAC through cyclic time-variant equalization. A brief discussion of these compensation results will also be presented. A thorough analysis of

our wideband characterization and compensation technique will be presented in the following chapter.

5.1 Transmitter Architecture

The 24-Gb/s transmitter includes an on-chip pattern generator (PG), a linear equalizer, and a DAC. A digital equalizer with 16 10-bit full-range taps, makes the transmitter a platform with sufficient flexibility to enable evaluation of different transmission algorithms in different environments. Figure 5.1 shows the top level block diagram of the transmitter.

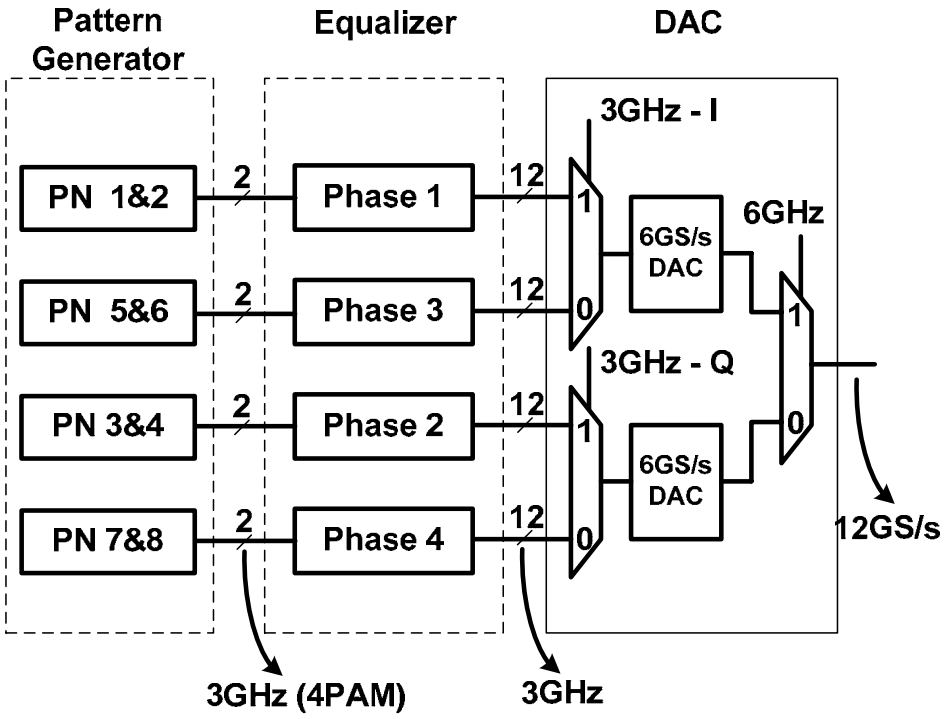


Figure 5.1: Transmitter block diagram. An on-chip clock generator circuit (not shown in the figure) generates the 6-GHz, 3-GHz, and 1.5-GHz clocks necessary for the system.

The PG creates 24-Gb/s of pseudo-random (PN) binary data for the system in the form of four 3-GS/s. Each of the four PG blocks is 2-way parallelized to operated with

a 1.5-GHz clock and can generate 4-PAM (or 2-PAM) sequence. The 16-tap linear equalizer is parallelized four ways with each parallel branch allowed to take independent tap values. This way the transmitter can be configured to support 4-channel and 2-channel AMT, as well as multi-PAM BB signal transmission ranging from 2-PAM to 256-PAM. Each parallel branch of the equalizer is 2-way parallelized to operate with a 1.5-GHz clock and receives a 4-PAM (or 2-PAM) sequence as input. Each equalizer tap coefficient has 10 bit resolution and the convolution operation is performed at full 10-bit precision, with the final result truncated to 8 bits. To ensure a reasonable Differential Non-Linearity (DNL), a thermometer encoder, included at the output stage of the equalizer, converts the 3 most significant bits of the output to unary code. An on-chip 8-bit 12-GS/s 2-way output multiplexed current-mode DAC converts the digital outputs to an analog signal and directly drives the 50-Ω chip output.

Figure 5.2 shows the DAC's architecture. High-speed digital streams from the preceding digital equalizer are multiplexed at the DAC interface. Each multiplexer inside the 4-2 MUX array converts four single-ended streams of data at 3-Gb/s into two differential lines, each operating at 6-Gb/s. Multiplexing is performed using 4 quadrature phases of a 3-GHz clock signal. The resulting 6-Gb/s streams are time multiplexed inside the DAC cell using a differential 6-GHz clock. Maximum allowable current mismatch and the number of unary-encoded bits are determined with system-level simulations for the target bit-error rate (BER) of 10^{-15} [37]. The design of the DAC is described in [39].

Figure 5.3 shows the clock network of the transmitter. A 12-GHz clock signal is applied to the chip as input reference frequency. This frequency is divided on-chip to create the 6-GHz and 3-GHz clocks required for the DAC, and the 1.5-GHz clock required for the digital equalizer and the PG. A phase interpolator and a phase detector are placed at the interface between the DAC and the equalizer to ensure the two circuits don't have setup and hold violations at their common interface, even though they have independent clock distribution networks. The phase detector samples a 1.5-GHz clock that branches off from the equalizer network with the 3-GHz in-phase clock and provides information on the phase alignment of the two clocks. The phase

interpolator is programmed offline based on this information. The 1.5-GHz clock distribution inside the equalizer is in the form of an 8-mesh with the clock being routed from the sides toward the center, in the direction of the data flow (Figure 5.4(a)).

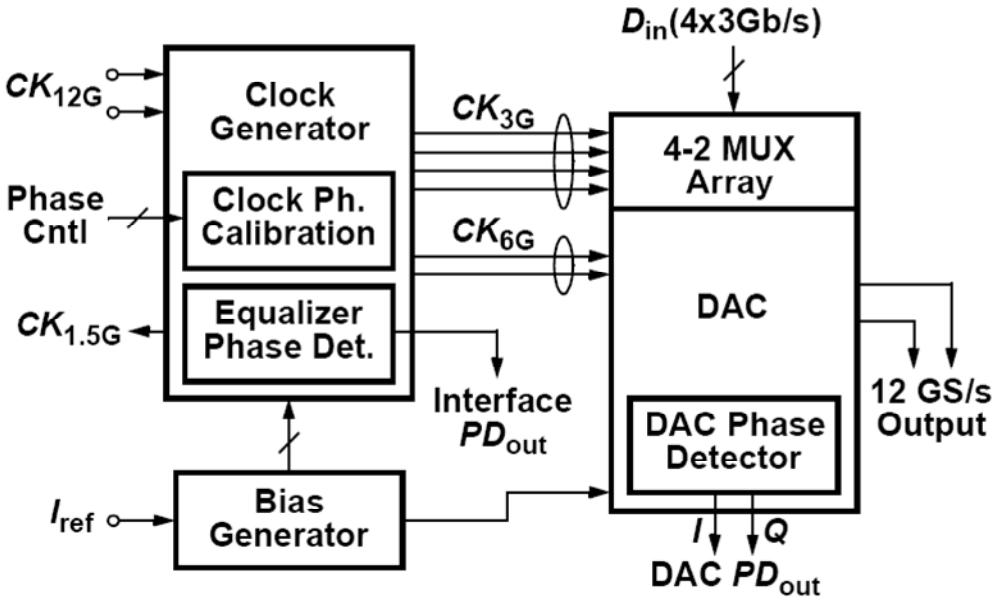


Figure 5.2: DAC architecture.

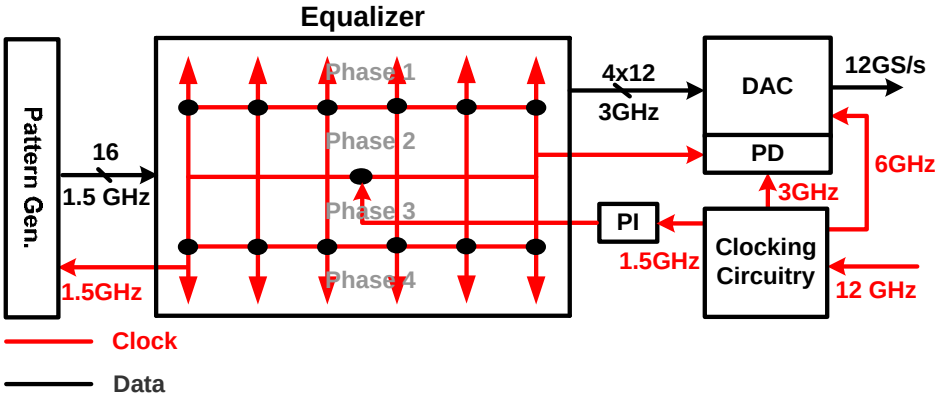


Figure 5.3: Transmitter clock network. Black dots in the equalizer indicate the placement of the clock drivers. A phase interpolator (PI) shifts the phase of the input clock to the transmitter based on information from a phase detector (PD). The phase detector samples a 1.5-GHz clock that branches off from a leaf of the equalizer's clock grid with the 3-GHz input clock to the DAC.

The PG block, implemented with standard library cells following a full ASIC flow, operates with a 375-MHz clock, with the exception of a small block of serializers at the output stage which operate with a 1.5-GHz clock. The 1.5-GHz clock for the PG also branches off from the equalizer network, and the 1.5-GHz clock distribution latency in the PG is included in the critical path of the data interface between the equalizer and the PG.

The equivalent functionality of the equalizer is a 16-tap FIR filter with 10-bit tap coefficients and 2-bit (4-PAM) inputs operating at 12-GS/s. Figure 5.4(b) shows the datapath of one phase of the 4-phase equalizer.

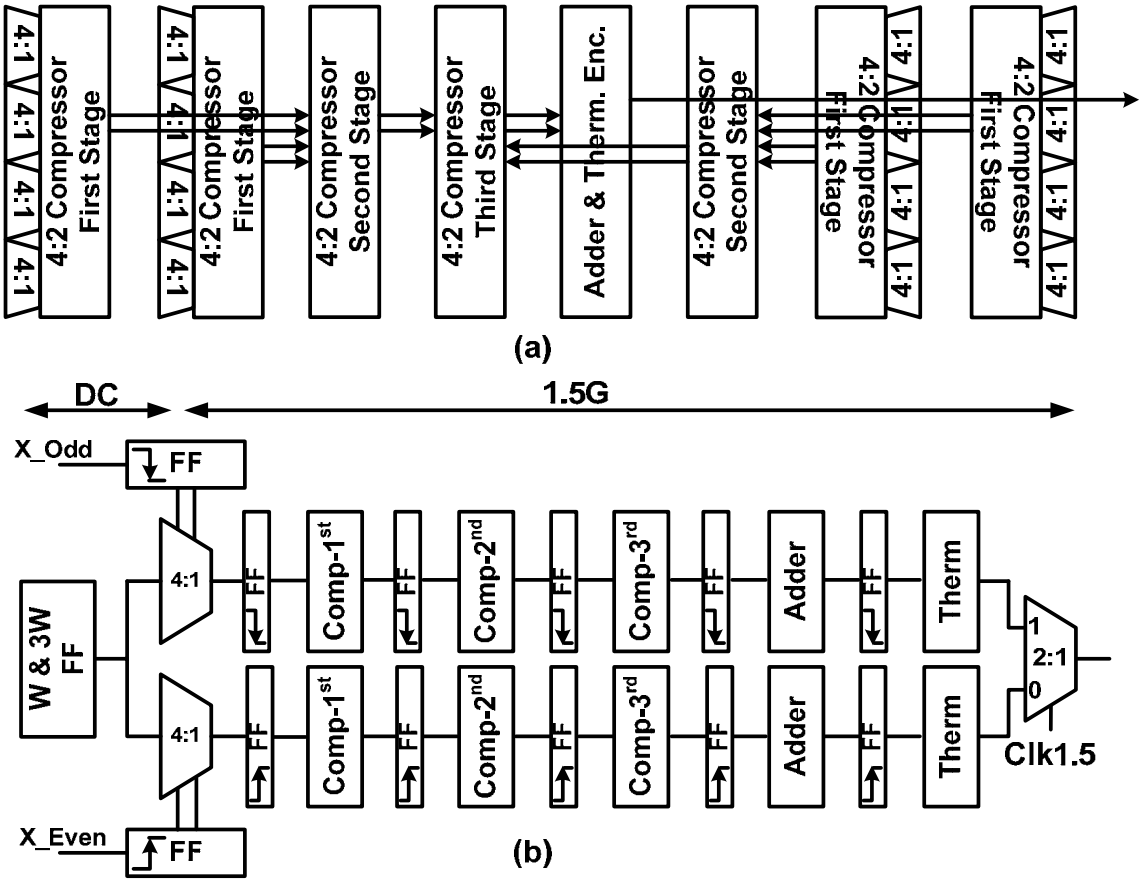


Figure 5.4: (a) Directions of data flow in the digital equalizer. (b) Datapath of one phase of the equalizer consisting of three stages of 4:2 compression, a pseudo Kogge-Stone adder and a thermometer encoder. Each block is 2-way parallelized and a 2:1 multiplexer (serializers) is included in the thermometer encoder.

In one phase, 16 10-bit by 2-bit multiplications are performed and the results are summed in one 3-GHz cycle. For multiplications, The 10-bit values of W and 3W (where W is an equalizer tap coefficient) are stored in flip-flops and a 4:1 multiplexer selects the correct multiplication output (±W or ±3W) based on the 2-bit data input (Figure 5.4(b)). The flip-flops storing the tap coefficients do not dissipate any active power and only impose a small area overhead. Additions are performed with three stages of 4:2 compressor units and a final pseudo Kogge-Stone adder. In this design, the compressor architecture proposed in [40] was used, except that parts of the logic were duplicated to create true and complementary outputs in parallel to reduce the total number of logic stages in the critical path (Figure 5.5). Optimizing the compressors to meet a 3.0-GHz cycle time would lead to larger area and higher power consumption than a 2-way parallelized architecture to meet a 1.5-GHz cycle. Consequently, the design was two-way parallelized to operate with a 1.5-GHz clock and each compression was placed in a separate pipeline stage. Overall, the equalizer has 5 stages of pipelining, with flip-flops placed before every compressor, before the final adder and before the thermometer encoder. A final 2:1 multiplexer that converts the 1.5-GHz odd and even outputs to a 3.0-GHz output is placed after the thermometer encoder and precedes the long output wires. The thermometer encoder was sized to settle in a 3.0-GHz cycle to reduce the transient power dissipated on the long output wires.

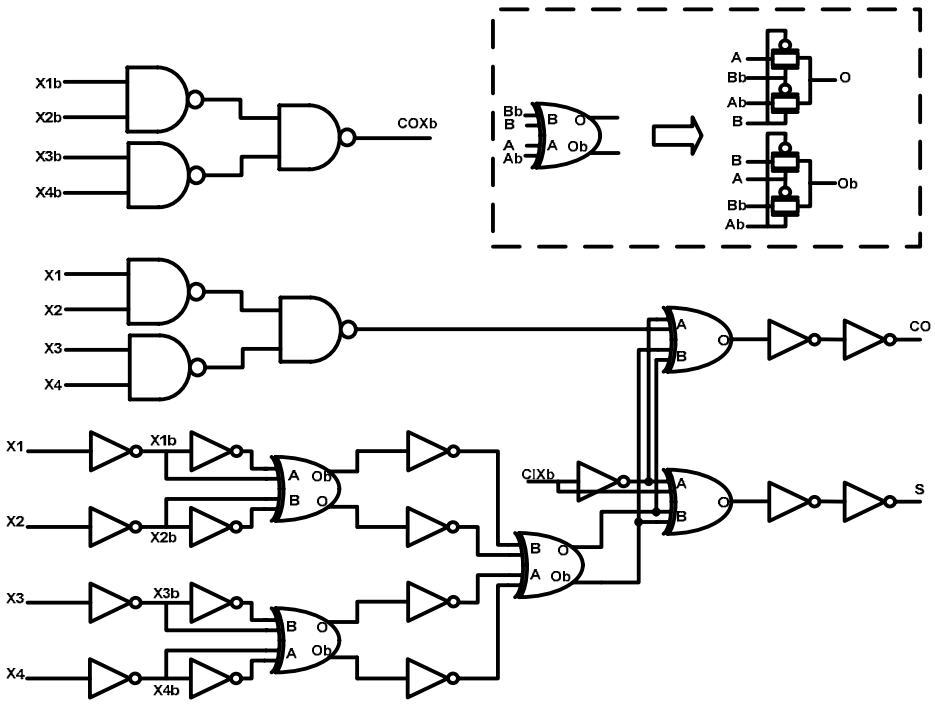


Figure 5.5: 4:2 Compressor Schematic. The critical path is from X1 to CO.

The equalizer is implemented in static logic and transistor sizing of the building components is fully custom. However, these building components are further characterized as standard cells to enable the automation of the design using a combination of several commercial ASIC design tools, including a synthesis and a place and route (P&R) tool, in addition to an in-house hierarchical MATLAB placement tool. The MATLAB tool interacts with the P&R tool to place high-speed components at exact desired locations. Figure 5.6 shows the complete floorplan of the equalizer. Routing is fully handled by the P&R tool and verification is performed with commercial Static Timing Analysis tools. The overall flow has the precision of a custom design while utilizing the vast automation and verification capabilities of the commercial ASIC design tools. The complete flow is described in more detail in Appendix C.

Figure 5.7(a) shows the chip micrograph. The main measured characteristics are included in Figure 5.7(b). At nominal operating rate, the entire transmitter has an energy-efficiency of 21mW/Gbps.

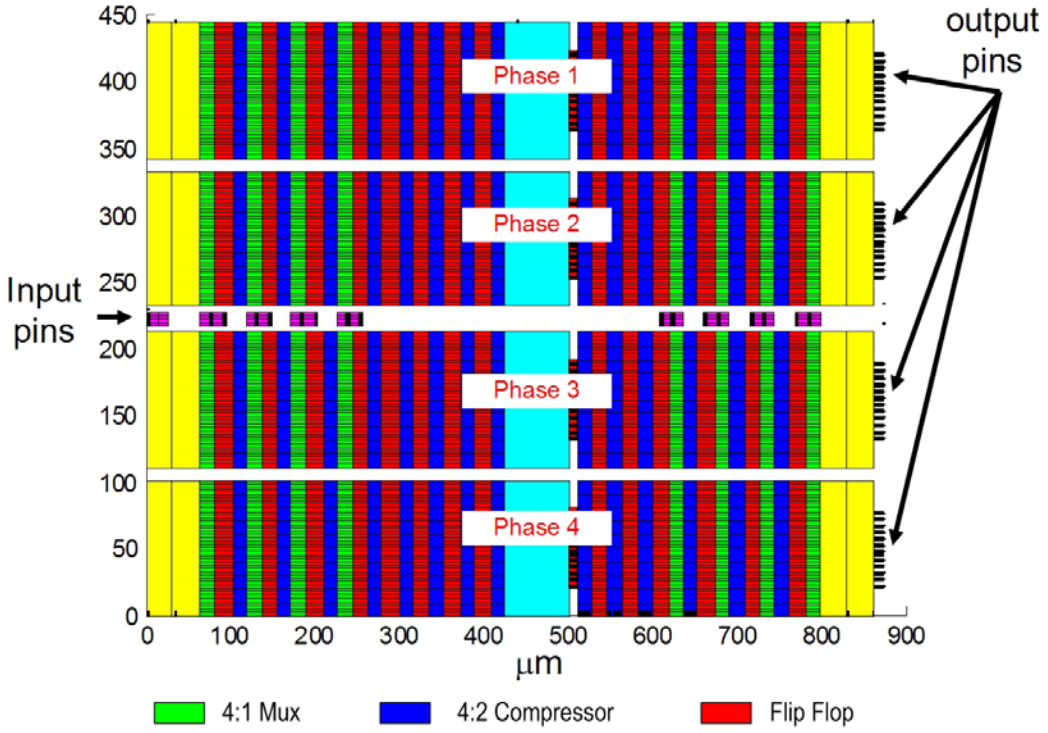
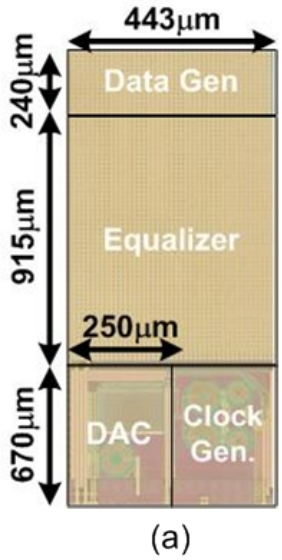


Figure 5.6: Equalizer floorplan in the MATLAB placement tool. Yellow rectangles on the sides are areas where low-speed flip-flops holding equalizer tap coefficients are placed by the P&R tool. The cyan rectangles in the middle are areas where the adder and the thermometer encoder are placed by the P&R tool. The pink rectangles in the middle row are latches, implementing the shift registers for the input data sequence. X and Y axes show actual sizes in microns.



Process	90nm CMOS
Maximum Rate	28Gb/s
Digital Power	350mW
Analog Power	160mW
Area	0.8mm^2
Output Swing	1.6V$_{pp}$

(b)

Figure 5.7: (a) Chip micrograph (b) Performance summary

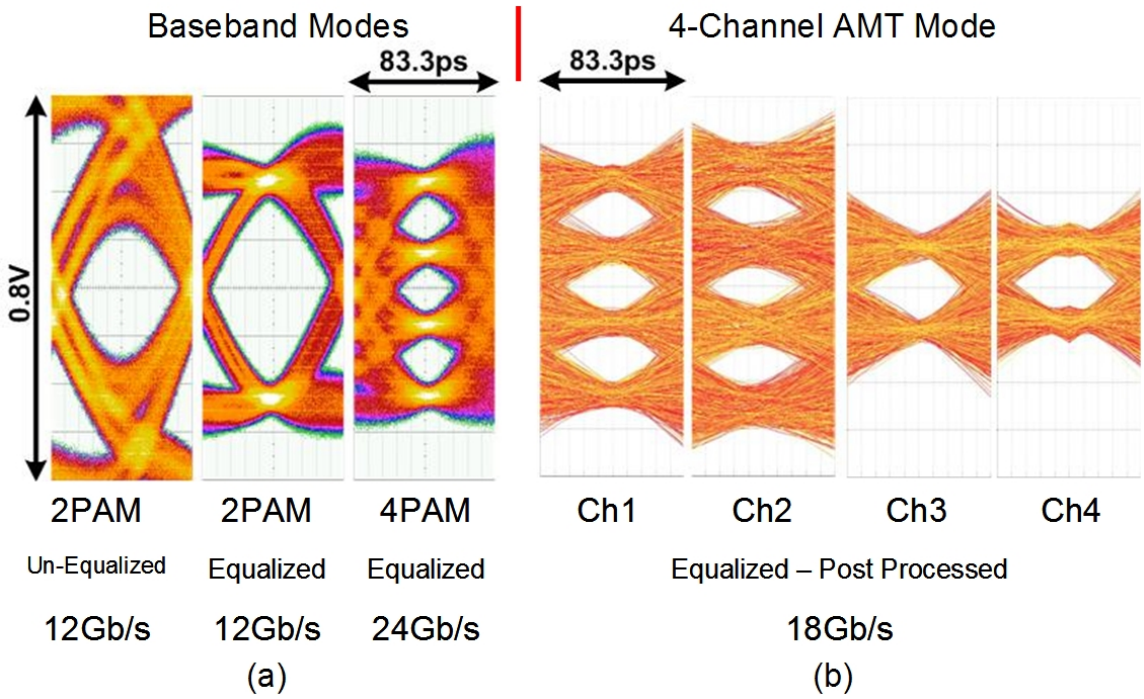


Figure 5.8: (a) Eye diagrams measured on a scope when the transmitter is operating in BB mode: un-equalized 2-PAM at 12-Gb/s (left), equalized 2-PAM at 12-Gb/s (middle) and equalized 4-PAM at 24-Gb/s. (b) Eye diagrams measured at the transmitter output with a scope and post-processed in MATLAB (mixing and integration per channel – no DFE) when the transmitter is operating in 4-channel 18-Gb/s AMT mode.

5.2 Measured Transmitter Performance

Figure 5.8 shows the measured eye diagrams in different operating modes. The three figures to the left show the eye diagrams captured on an equivalent-time scope when the transmitter is operating in BB mode. In AMT mode, the eye diagrams can only be drawn after mixing and integration at the receiver. Since an AMT receiver was not available, the analog signal at the output of the transmitter was sampled with a scope and mixing and integration was performed in MATLAB to generate the eyes. Figure 5.8(b) shows the four eye diagrams corresponding to the four sub-channels when the transmitter operates in 4-channel AMT mode. Two of the sub-channels are operating in 4-PAM mode and the other two in 2-PAM mode for an aggregate data rate of 18-Gb/s. The two 2-PAM sub-channels can also be programmed to operate in

4-PAM mode. For all the eye diagrams shown in Figure 5.8, the communication channel consists of the chip plastic package, 2 inches of differential traces on the PCB, 3 feet of cable, and the front-end of the scope. The overall pulse response of the channel exhibits 14-dB of attenuation at 6-GHz. Measured 3-dB bandwidth of the DAC with respect to an ideal zero-order-hold pulse is at 7.1-GHz.

As we described in Chapter 3, AMT can particularly improve the performance of high-speed links where stubs dominate the channel characteristic. Figure 5.9(a), shows the same multi-drop topology studied in Chapter 3, fabricated on an FR4 PCB trace. The main trace between the two CPUs is 16" and the three stubs are approximately 1". Figure 5.9(b) shows the measured frequency response of the channel when the prototype transmitter is connected to one end of the trace, acting as the transmitting CPU, and an oscilloscope is connected to the other end of the trace, acting as the receiving CPU. The first notch frequency is slightly above 1-GHz. A property of AMT, and the implemented transmitter in particular, is that the entire frequency response of the system can be scaled by changing a single clock frequency. In this case in order to place the 3dB bandwidth of the BB sub-channel around 1.1GHz, the input clock frequency to the transmitter is reduced from 12-GHz to 9.2-GHz. Figure 5.9(c) shows the measured eye-diagrams when only two of the sub-channels are used in 2-PAM mode for an aggregate data-rate of 4.6-Gb/s. Again the receiver (Figure 5.9(d)) is implemented in MATLAB to post process measured data collected by the oscilloscope. If the transmitter is configured in 2-PAM BB mode over the same channel, no open eyes can be observed on the scope beyond 2.6-Gbps.

5.3 Other Transmitter Operating Modes

The degrees of freedom available in the transmitter enable supporting more complex multi-PAM (for example 8PAM) BB modulations. In addition the freedom in choosing the tap coefficients for different branches of the parallelized equalizer enables digital compensation of the analog distortions through cyclic time-variant equalization. These modes are briefly described in this section.

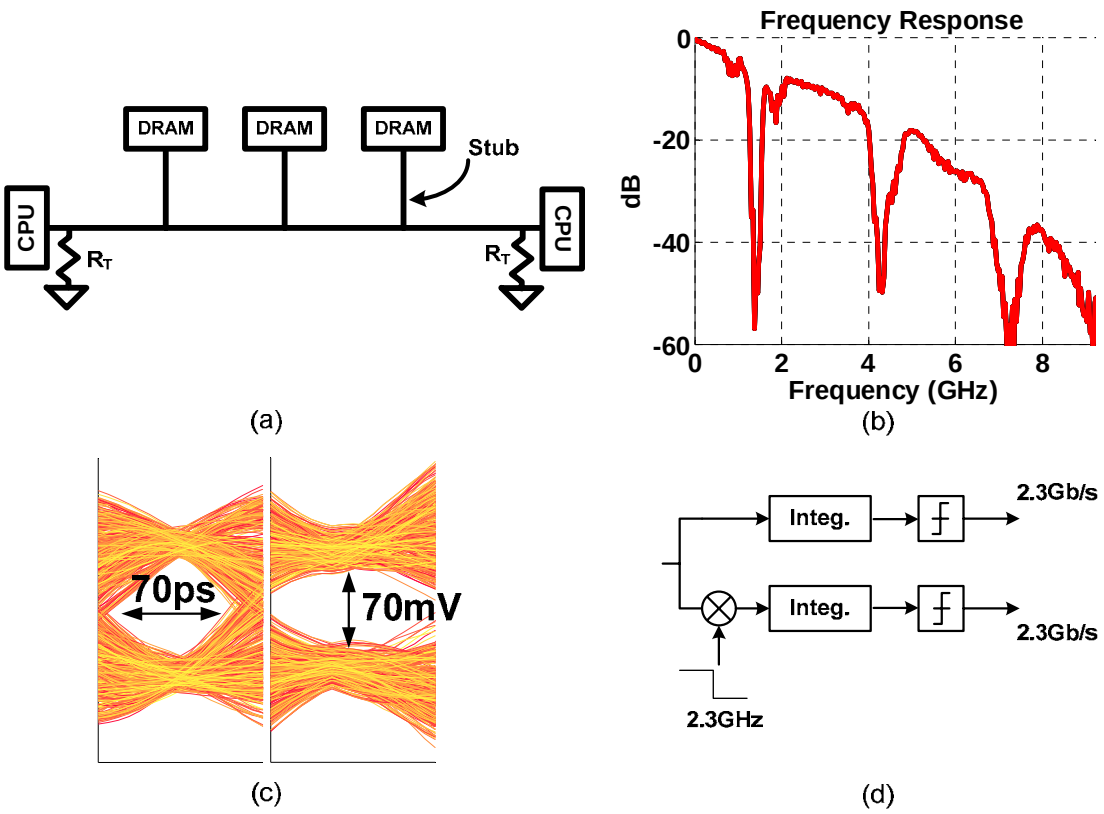


Figure 5.9: (a) a Multi-drop configuration (b) Measured frequency response of a three-drop 16" FR4 trace (c) Eye diagrams based on measured data when only two 2PAM, 2.3-GS/s sub-channels are used. Total throughput is 4.6-Gb/s (d) Block diagram of the receiver simulated in MATLAB to generate the eyes.

5.3.1 **Multi-PAM Operation**

By programming the correct tap coefficients to the equalizer, the transmitter can support signal transmission in 2^M-PAM ($1 \leq M \leq 8$) mode. Operation in 8-PAM, for example, is enabled based on the observation that an 8-PAM sequence can be decomposed to the summation of a 4-PAM sequence and a 2-PAM sequence as shown in Figure 5.10(a).

Figure 5.10(b) shows the transmitter configured to operate in 8-PAM mode, as a two-way parallelized system where each of the parallel branches consists of two branches which together generate an equalized 8-PAM output. In this mode the transmitter can only support up to 18-Gb/s of uncorrelated data since two of the

branches are operating in 2-PAM mode instead of 4-PAM. However, in order to show the operating limits of the analog circuits in the transmitter, some of the equalizer taps were used to delay the input sequence for a few cycles and add it back to itself. This effectively doubles the throughput by creating a virtual source with a weakly correlated sequence. As long as the delay is larger than the delay spread of the pulse response of the channel, such correlation has negligible effect on the eye diagrams. Figure 5.10(c) shows that discernible, although perhaps too small to be practical, eyes observed on the scope at 36-Gb/s in 8PAM mode. Similar ideas can be applied to program the transmitter in higher multi-PAM BB modes as well as 2-channel AMT with 6-GS/s (12-Gb/s) sub-channels.

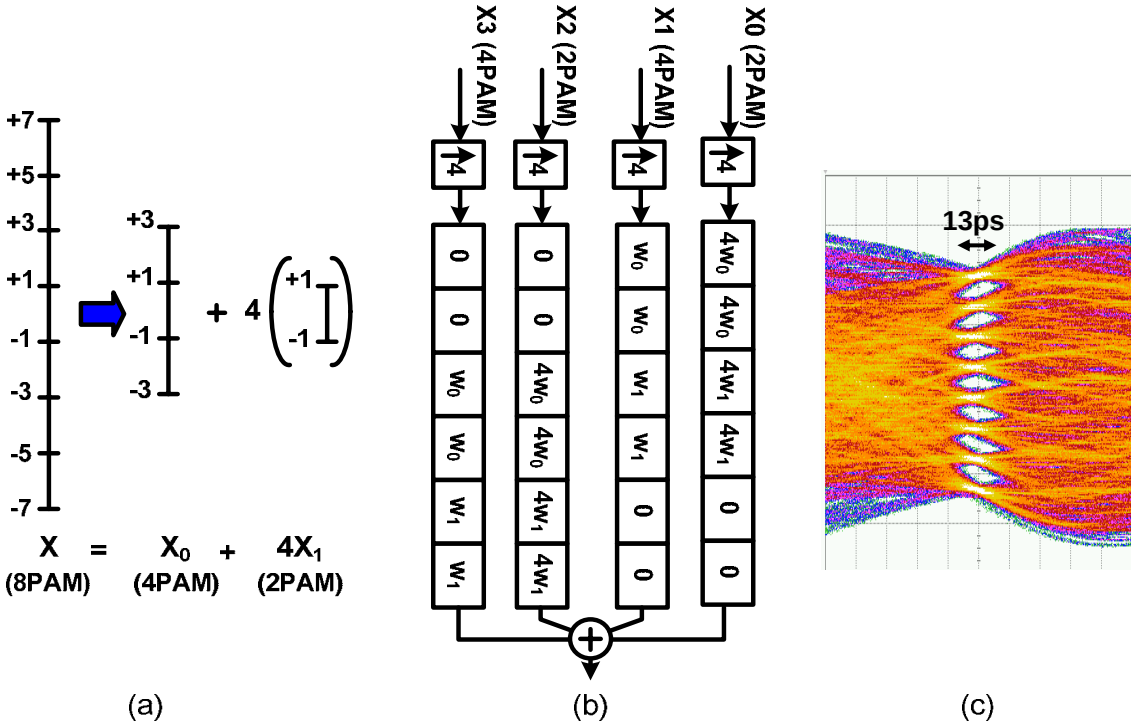


Figure 5.10: (a) 8-PAM symbol decomposition to a 4-PAM and a 2-PAM symbol. (b) Transmitter configured in 8-PAM mode (only 6 taps per phase shown) (c) Eye diagram on a scope in 8-PAM, 36-Gb/s mode.

5.3.2 **Linear Cyclic Time-Variant Equalization**

The transmitter also supports cyclic time-variant equalization. This mode is useful for applications utilizing wideband interleaved data converters, where the response from the input to the output may be different from time to time due to the mismatch between the interleaved paths. For example, in the transmitter described in this paper, four different paths can be identified from the input of the DAC to its output. Figure 5.11(a) shows the estimated pulse responses for these four paths. It can be seen that one path is visibly different from the other three. This difference in the responses makes the system cyclically time-variant rather than time-invariant. Therefore, as long as the system is treated like an LTI system, the time-variant nature manifests itself as analog distortion.

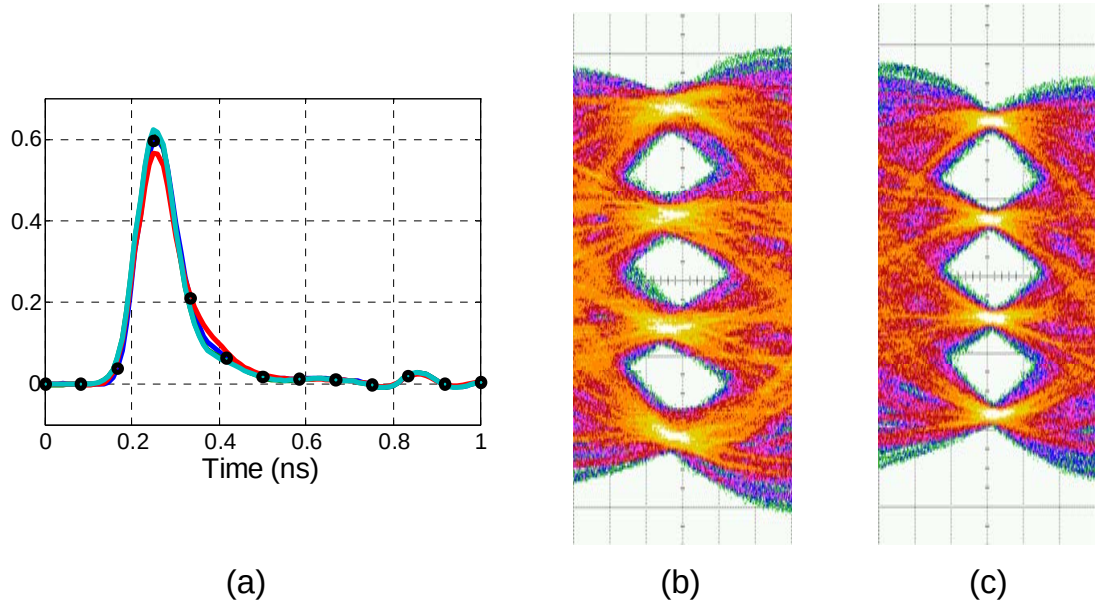


Figure 5.11: (a) Measured pulse responses of the 4 phases of the DAC. (b) LTI equalized BB 4PAM 28Gb/s eyes on a scope (b) Cyclically time-variant equalized BB 4PAM 28Gb/s eyes on a scope.

Figure 5.11(b) shows the measured eye diagrams when the transmitter is operating in BB 4PAM 28Gbps mode[20], and LTI equalization is performed. Measured Signal to Interference and Distortion Ratio (SIDR) at the middle of the eye is 26dB. However, if the four phases of the 4-way parallelized equalizer in the transmitter are programmed independently to perform cyclic time-variant equalization [41], measured SIDR improves to 31dB (Figure 5.11(c)), which indicates at least 5dB of the distortion is related to the mismatch.

5.4 Summary

We showed that a software programmable transmitter can be implemented by parallelizing a conventional FIR filter and letting each of the parallel branches to be programmed independently. The architecture of the transmitter enables supporting bandwidth scalable AMT and 2^M-PAM BB transmission. In addition, the degrees of freedom available in an AMT system and in the transmitter in particular, enable cyclic time-variant equalization to compensate for analog distortions of the system. The additional degrees of freedom come at a small cost, since the additional equalizer tap values which are needed only add a small additional area to the chip, but don't increase the active power. In the following section we provide more details on how the characterization for the DAC was performed.

[20] Input clock frequency is increased to 14-GHz.

Chapter 6

Characterization and Compensation of Wideband Circuits

The design of the 12-GS/s DAC included in the 24-Gb/s transmitter, posed a few challenges early on in the design process. Given the important role of the DAC in the transmitter, we needed a design methodology to guide the design process to make sure the DAC meets the system level specifications, while avoiding wasting power and area by over-designing the DAC.

The performance of a high-speed circuit like the DAC is inevitably affected by various kinds of distortion. These non-idealities need to be correctly characterized, and if the design doesn't meet the target performance, they further need to be compensated for in analog or digital domain. Digital compensation techniques can be particularly efficient [42][43] since they exploit the raw speed and high density of digital gates in modern CMOS processes. However, utilizing such techniques sometimes requires an accurate model of the linear and non-linear distortion terms, as well as an estimate of the potential performance improvement if such compensation techniques are applied.

Single-tone and two-tone tests are typically used to measure the linearity of circuits like DACs [44]-[47]. If the circuits are designed to operate with narrowband signals, tone frequencies are chosen around the frequency band of the signal for accuracy. However, as the bandwidth of the input signal increases, this characterization method provides varying results dependent on the choice of the input frequencies because many sources of non-linearity are frequency dependent (e.g. voltage-dependent capacitors).

Multi-tone (multi-sine) tests have also been devised to characterize the system over wide frequency ranges [48][49]. In a multi-tone test, the input consists of multiple sinusoidal signals, wherein tone frequencies are chosen to allow separation of odd-order and even-order non-linearities. The drawbacks of this technique include the random choice of sine amplitude and phases, as well as the unwanted overlap of different non-linear terms. In addition, the distortion metrics defined by tone tests are not directly related to the system-level performance metrics such as the BER.

In this chapter we propose a new approach to characterizing wideband circuits, which provides an accurate wideband linearity measure, as well as a platform to characterize an entire communication system at the speed of a high-level simulator yet with the precision of a transistor-level simulator. This technique also provides detailed insight into the circuit dynamics and introduces useful metrics such as impulse response and large-signal circuit bandwidth. The extension of this method to the estimation of higher-order Volterra filters of a weakly non-linear circuit as well as its cyclo-stationary bias provides insight into the sources of non-ideality, and provides an accurate system model that can be used for compensation of analog non-idealities. Further extension of the characterization technique to the broader category of Cyclic Time Variant (CTV) circuits makes it applicable to a very large family of circuits, including mixers and time-interleaved data converters (DAC and analog-to-digital-converter (ADC)).

6.1 Wideband Circuit Characterization

Our approach is quite simple. A circuit consists of a *Circuit Model* and a *Distortion term* (Figure 6.1). The circuit model represents deterministic aspects of the circuit behavior that do not affect system performance, or the terms that can be compensated for in the digital domain. Distortion substitutes anything not captured by the model, and may be considered as (signal dependent) noise.

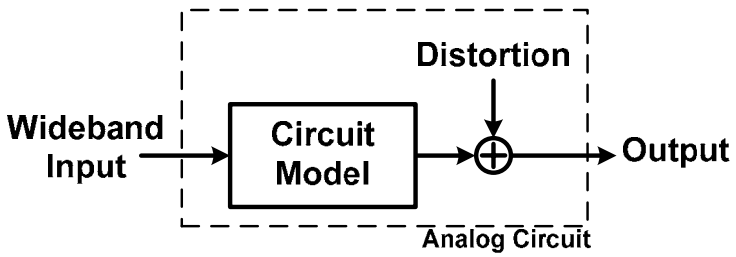


Figure 6.1: System model for characterization purposes

Finding the circuit model is a system identification problem [50]. The goal of system identification is to model a system with a set of basic components (like filters) and estimate their values such that the model correctly predicts the deterministic nature of the system output, given the inputs. As long as the outputs predicted by the model are consistent with the physical system, whether the components in the model correspond to physical phenomena is of little consequence. Nonetheless, good models generally provide insight to the way the physical system operates. For example, the simplest form of a weakly non-linear circuit's model is a linear time invariant (LTI) representation, captured by the circuit's impulse response, which is an FIR filter. The FIR filter coefficients can be estimated using Least Squares Estimation (LSE) techniques, for example.

The estimation phase of the system identification problem is referred to as "characterization" in this paper. The variance of the estimation error, which represents the energy of unwanted distortion, can consequently be used as the distortion metric to define parameters such as wideband Signal-to-Distortion Ratio (SDR). The estimated impulse response, on the other hand, can be used to define the bandwidth of the circuit, or be used in conjunction with a linear equalizer to compensate for the linear dispersion of the circuit. The circuit model can also include a dc offset term if the system has a dc bias which does not affect system performance. The dc bias will consequently be estimated along with the linear response, and the resulting SDR would not include the effect of the dc bias. In the following, for clarity, we describe the characterization process for a DAC, although the approach is applicable to a large family of wideband circuits.

6.1.1 **LSE-Based Characterization**

The characterization process has to be performed in discrete domain at a rate at least two times larger than the expected bandwidth of the circuit to satisfy the Nyquist Theorem. Figure 6.2 shows the block diagram of a DAC modeled as a discrete linear filter **h,** characterized at *r* times the input rate. For an ideal DAC, the filter **h** is a zero-order-hold pulse. For a realistic DAC, the best filter, $\hat{\mathbf{h}}$, is the one that provides the best fit between the outputs predicted by the linear model $y(t)=h(t) * x(t)$, where * is the convolution operator, and the circuit inputs. $\hat{\mathbf{h}}$ is called the linear response of the circuit and is used to find the large signal bandwidth of the circuit. The difference between the DAC output and the output predicted by $\hat{\mathbf{h}}$ is the estimation error. A Least Squares solution to **h** finds the best fit by minimizing the energy in the estimation error.

In order to find $\hat{\mathbf{h}}$ through LSE, we need the actual circuit outputs and the corresponding inputs. For example a piece-wise linear input signal can be generated using MATLAB to be used in a transient simulation of the circuit in SPICE. Temporal and probabilistic distribution of the input signal should mimic a typical operating condition. The output of the transient simulation is read back into MATLAB for estimation.

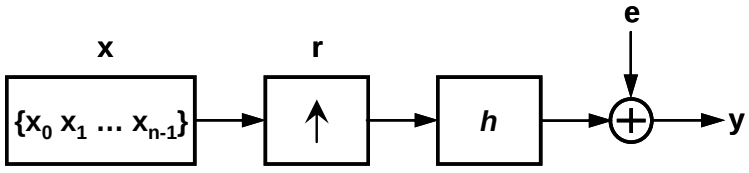


Figure 6.2: Wideband characterization using a random sequence

Using the model of Figure 6.2, we define the system as:

$$\mathbf{y} = \mathbf{Xh} + \mathbf{e} \tag{6.1}$$

where,

$$\mathbf{X} = \begin{pmatrix} x_0 & 0 & \cdots & 0 & \cdots & x_{n-1} & 0 & \cdots & 0 \\ 0 & x_0 & \cdots & 0 & & 0 & x_{n-1} & & 0 \\ & & \ddots & & & & & \ddots & \\ 0 & 0 & \cdots & x_0 & \cdots & 0 & 0 & \cdots & x_{n-1} \end{pmatrix}^T \tag{6.2}$$

In this equation, **X** is the convolution matrix of the input sequence with *nr* rows and *l* columns (*n* is the number of input symbols, *r* is the up-sampling rate, and *l* is the length of filter h). The length of **h** should be sufficiently large to model all dispersion created by the circuit. Next we define *Error* and minimize it to find $\hat{\mathbf{h}}$:

$$Error = \mathbf{e}^T\mathbf{e} \tag{6.3}$$

This will yield to:

$$\hat{\mathbf{h}} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{y} \tag{6.4}$$

$$Error = \mathbf{y}^\mathbf{T}\mathbf{y} - (\mathbf{y}^\mathbf{T}\mathbf{X})(\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}(\mathbf{X}^\mathbf{T}\mathbf{y}) \tag{6.5}$$

Linearity of the system is measured using its SDR which can be defined as:

$$SDR = \frac{(\mathbf{y}-\mathbf{e})^\mathbf{T}(\mathbf{y}-\mathbf{e})}{Error} = \frac{\hat{\mathbf{h}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{X}\hat{\mathbf{h}}}{Error} \tag{6.6}$$

Further computational simplicity can be achieved by representing $\hat{\mathbf{h}}$ with its poly-phase form [34]. This reduces the size of the matrices by a factor of *r* (a factor of r^2 for $\mathbf{X}^\mathbf{T}\mathbf{X}$) and significantly speeds up the matrix inversion required to obtain the solution.

Even though the DAC in our transmitter has 8 bits of resolution, to demonstrate the ideas, we first apply the techniques to a simple 2-bit current-mode DAC simulated in SPICE (Figure 6.3(a)) with controlled non-idealities. The DAC consists of three identical 1-bit output multiplexed driver cells as shown in Figure 6.3(b). Each cell consists of two half rate current-mode drivers with a 2:1 multiplexer at the output. A few asymmetries have been introduced to the structure of the DAC to create predictable analog distortion. The asymmetries include a mismatch in the duty cycle of

the 6-GHz clock (CK_{6G}) by 6% and a mismatch between the two current sources inside the cells (I_1 and I_2) by 5%. Both of these parameters are skewed such that the side that drives the output longer also has larger tail current. The fixed output capacitances (C_1 and C_2), modeling the ESD and output wiring capacitances, are mismatched by 10% to show the impact of output clock feed-through.

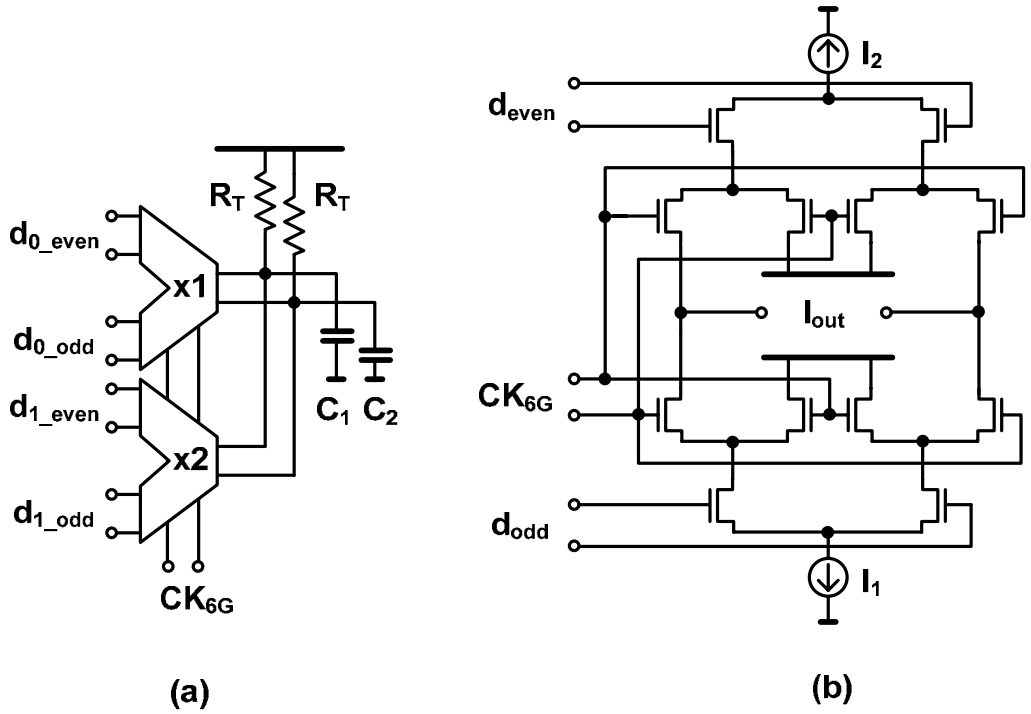


Figure 6.3: (a) Block diagram of a 2-bit DAC (b) 1-bit output-multiplexed driver cell

The circuit is characterized based on the results obtained from SPICE transient simulation. The system-level circuit model and the performance metrics of the circuit are constructed in MATLAB. The input to the system-level circuit model is a PAM signal with 4 levels: -3,-1,1, and 3 and the output is a real valued wideband signal. Figure 6.4 shows the estimated linear response of the DAC using a simple LTI model. The estimated SDR is 22-dB in this case.

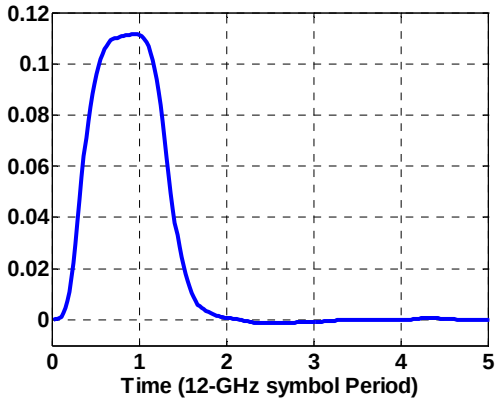


Figure 6.4: 2-bit DAC single bit response.

6.1.2 **CTV Characterization**

The estimated SDR for the 2-bit DAC is low due to the mismatches between the interleaved paths. The odd output samples of the DAC are wider and larger than the even outputs. As a result, the odd and even output samples have different linear responses. In other words, the circuit is CTV with a period of 2 instead of LTI. Consequently, a CTV model with a period of 2 captures the linear behavior of the DAC more accurately by letting the estimated distortion term only represent the non-linear behavior.

Many wideband circuits are CTV rather than time-invariant. Examples include mixers and time-interleaved data converters with mismatch between the interleaved paths. Figure 6.5(a) shows a 1st order time-variant model of the circuit being characterized, where the wideband circuit is modeled by a linear time-variant linear response $h1^{\tau}(n)$, and a distortion term $e(n)$ which includes any other sources of distortion in the system. The assumption of cyclic time variation means that the linear response $h1^{\tau}(n)$ is periodic in "τ". In other words, input samples to the system that are separated in time by the variation cycle excite the same linear response.

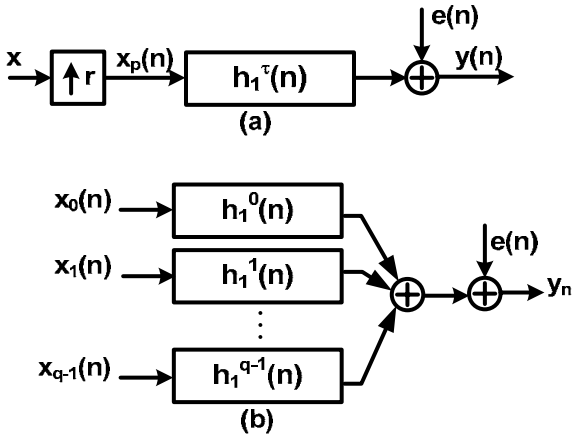


Figure 6.5: (a) CTV system model for linear term estimation ($h_1^{\tau}(n) = h_1^{\tau+q}(n)$), (b) equivalent model. $x_s(n)$ ($0 \le s < q-1$) is a decimated version of the original sequence x(n), but every sub-sequence is decimated at a different offset.

Then the input-output relationship can be described as:

$$y = x * h + b + e \tag{6.7}$$

where * denotes time-variant convolution. If the variation cycle is q, we can decompose the input sequence to a set of q input sub-sequences $x_s(n)$ ($0 \le s < q-1$), each of which is a decimated version of the original sequence, but every sub-sequence decimated at a different offset.

$$\begin{aligned} &x_s(kq+s) = x(kq+s), \;\; 0 \le s < q, k = 0,1,... \\ &x_s(kq+s) = 0, \;\; otherwise \end{aligned} \tag{6.8}$$

Due to the cyclical nature of the circuit, every sub-sequence observes a time invariant transfer function, and therefore, by superposition, the system can be described as shown in Figure 6.5(b) and written as:

$$y(n) = \sum_{s=0}^{q-1} x_s(n) * h_1^{\,s}(n) + e(n) \tag{6.9}$$

where * now denotes linear time-invariant convolution. In matrix form we have:

$$\mathbf{y} = \sum_{s=0}^{q-1} \mathbf{X}_s \mathbf{h}_1^{\,s} + \mathbf{e} \tag{6.10}$$

Or equivalently as:

$$\mathbf{y} = \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1 & \cdots & \mathbf{X}_{q-1} \end{bmatrix} \begin{bmatrix} \mathbf{h}_1^0 \\ \mathbf{h}_1^1 \\ \vdots \\ \mathbf{h}_1^s \end{bmatrix} + \mathbf{e}$$

$$\mathbf{y} = \mathbf{X}_{\mathbf{TV}}\, \mathbf{h}_{\mathbf{TV}} + \mathbf{e} \tag{6.11}$$

Finally, using the LSE technique:

$$\begin{aligned} \mathbf{h}_{\mathbf{TV}} &= (\mathbf{X}_{\mathbf{TV}}^{\mathbf{T}} \mathbf{X}_{\mathbf{TV}})^{-1} \mathbf{X}_{\mathbf{TV}}^{\mathbf{T}} \mathbf{y} \\ \mathbf{e} &= \mathbf{y} - \mathbf{X}_{\mathbf{TV}} (\mathbf{X}_{\mathbf{TV}}^{\mathbf{T}} \mathbf{X}_{\mathbf{TV}})^{-1} \mathbf{X}_{\mathbf{TV}}^{\mathbf{T}} \mathbf{y} \end{aligned} \tag{6.12}$$

And SDR is:

$$SDR = \frac{(\mathbf{y} - \mathbf{e})^T (\mathbf{y} - \mathbf{e})}{\mathbf{e}^{\mathbf{T}} \mathbf{e}} = \frac{\mathbf{h}_{\mathbf{TV}}^{\mathbf{T}} \mathbf{X}_{\mathbf{TV}}^{\mathbf{T}} \mathbf{X}_{\mathbf{TV}} \mathbf{h}_{\mathbf{TV}}}{\mathbf{e}^{\mathbf{T}} \mathbf{e}} \tag{6.13}$$

Figure 6.6 shows the estimated time-variant responses of the 2-bit DAC for the two clock phases. The difference in the linear responses clearly shows the effect of duty cycle mismatch (the difference between pulse-width of the linear responses) and current source mismatch in the driver cells (the difference between pulse amplitude of the linear responses). The estimated SDR in this case is 28-dB, which is 6-dB higher than LTI estimation and indicates that CTV estimation better models the system.

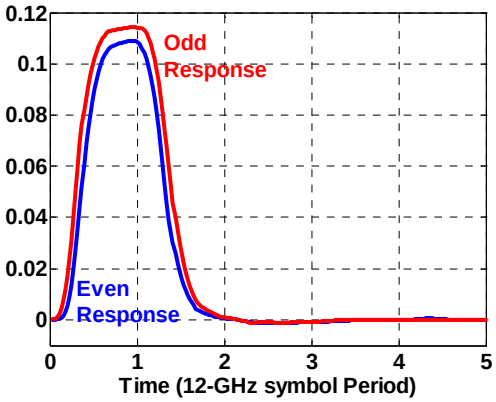


Figure 6.6: 2-bit DAC pulse responses corresponding to the 2 interleaved paths,

6.1.3 **Non-linearity Decomposition**

The characterization techniques described in the previous section can also be extended to modeling non-linearity in weakly non-linear circuits. In this paper we use the Volterra series representation of weakly non-linear a circuit for estimating its non-linearity [51][52]. Volterra series approximation for systems is similar to a Taylor series approximation for functions. Using Volterra series, the output can be written as:

$$y(n) = h_0(n) + \sum_{m_1} x(n-m_1)h_1(m_1) + \sum_{m_1,m_2} x(n-m_1)x(n-m_2)h_2(m_1,m_2) + \cdots \qquad (6.14)$$

where $h_0(n+g) = h_0(n)$, and g is a fixed number[21]. $h_0(n)$ represents data-independent cyclic bias at the output of the circuit. Inclusion of this term in the model instead of a scalar bias term acknowledges the fact that many noise sources in the system, including supply noise and clock feed-through, are cyclo-stationary in nature [53]. The second term is the convolution of the input signal and the linear response. The other terms are higher order non-linearity terms. All the convolutions in Equation 6.14 can represent CTV convolution and the corresponding filters can be replaced by CTV models to perform CTV non-linearity estimation. We will skip this generalization in the formulations that follow to reduce the number of variables and indices. The

[21] g = r assumes that the period of the cyclo-stationary noise is the same as the input rate. g= 2r, 3r, … assumes longer period.

number of non-linearity terms included in the model and the value of *g* are chosen based on the designer's expectation of the behavior of the circuit, or by trial and error. For example, in order to derive the model with non-linearity up to the m^{th} order, we describe the output of our system using a matrix form.

$$\mathbf{y} = \mathbf{I_0}\mathbf{h}_0 + \mathbf{X_1}\mathbf{h_1} + \mathbf{X_2}\mathbf{h_2} + \cdots + \mathbf{X_m}\mathbf{h_m} + \mathbf{e} \tag{6.15}$$

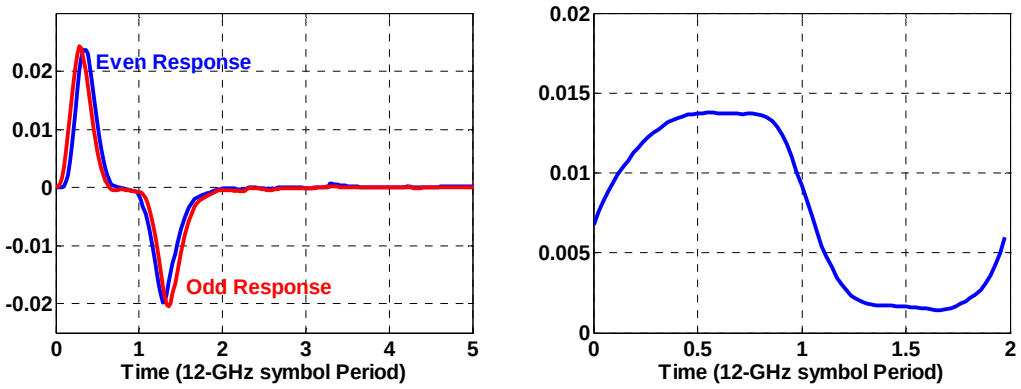
In this equation, $\mathbf{X}_1$ is the convolution matrix of the input sequence, similar to the matrix described in (2). Other matrices ($\mathbf{X}_2$ to $\mathbf{X}_m$) are higher order input convolution matrices in which x_0 to x_{n-1} are replaced by x_0^i to x_{n-1}^i, where *i* is the index of the matrix. $\mathbf{h_0}$ is the cyclic bias with *g* terms, $\mathbf{h_0} = [h_0, h_1, \ldots, h_{g-1}]^T$, and $\mathbf{I}_0 = [\mathbf{I}_{gxg}\ \mathbf{I}_{gxg}\ \ldots\ \mathbf{I}_{gxg}]^T$, where $\mathbf{I}_{gxg}$ is an identity matrix. Non-linear cross-product terms can also be included if expected to be large.

The above equation can be written in a more compact matrix form as follows:

$$\mathbf{y} = \begin{bmatrix}\mathbf{I_0} & \mathbf{X_1} & \cdots & \mathbf{X_m}\end{bmatrix}\begin{bmatrix}\mathbf{h_0} & \mathbf{h_1} & \cdots & \mathbf{h_m}\end{bmatrix}^T + \mathbf{e} \tag{6.16}$$

The equation is similar in form to the LSE presented in previous section. Again, we minimize error using LSE and derive $\mathbf{h_0}$, $\mathbf{h_1}$, $\mathbf{h_2}$, **…**, and $\mathbf{h_m}$. This approach will provide us with linear filters for the input signal and its higher-order non-linear terms. Knowledge of these terms provides insight to the behavior of the circuit, which can be utilized to reduce their effect by design. The estimated SDR also indicates the potential improvements if such non-linearity terms are compensated by digital signal processing.

We can enhance the model of the 2-bit DAC by adding cyclic bias and 3rd order non-linearity, which raises the SDR to 30 dB and 43 dB respectively. Figure 8(a) shows the estimated filters associated with 3rd order non-linearity of the DAC for the two phases of the circuit, and Figure 8(b) shows the estimated cyclic bias. The period of the cyclic bias is 2 symbol periods, which is a full cycle of time-variance in the circuit.



(a) (b)

Figure 6.7: 2-bit DAC (a) 3rd order non-linearity responses corresponding to the 2-phases (b) Cyclic bias (dc-offset and clock feed-through)

6.1.4 **Effect of Observation Length**

Short input random sequences result in optimistic SDR values because such a sequence does not accurately reflect the statistical characteristics of the input data. In other words, an LSE solution is biased towards the finite-length input sequence used for the estimation. In order to measure this dependency we can estimate **h** using one random sequence, then use a new sequence and directly compute SDR using the previously estimated **h**. A discrepancy between these two SDR values indicates that the original input sequence is too short and estimation must be based on a longer sequence. It can be shown that the difference between these two SDR values decreases almost linearly with the observation length [38]. In practice, the number of observation points should be increased until its dependence on the SDR value is no longer significant.

6.2 Real-Time Characterization of the 12-GS/s DAC

In this section we apply the time-variant characterization method to the 12-GS/s DAC in the transmitter. In order to obtain the estimated vector $\mathbf{h}_{TV}$ and the SDR in the lab, both the input and output sequences must be simultaneously accessed. The input

sequence used in this case is generated by the eight independent PN-sequence generators, running at 3-GHz and generating 4-PAM data for the 4 phases of the equalizer. PN generators can be programmed with different initial phases and the equalizer taps can be programmed to create appropriate levels at the DAC input during the characterization phase. For example, the four phases of the equalizer can be programmed as in Figure 6.8 to create a 12-GS/s sequence at the DAC input that has white correlation properties and hits all values in the range of (0,255).

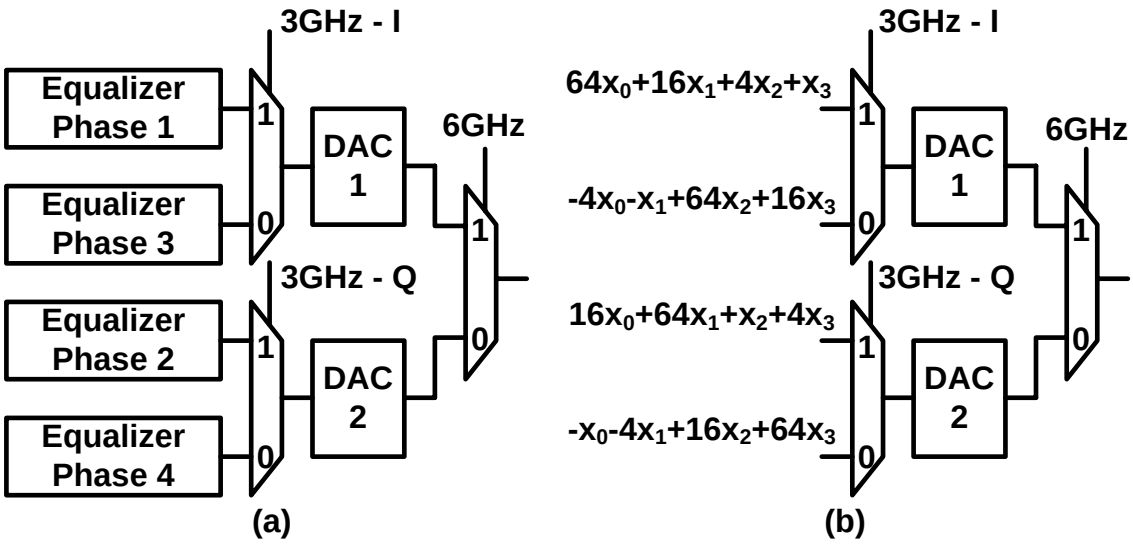


Figure 6.8: (a) Conceptual transmitter architecture (b) Equalizer configured to create a white input for the DAC that hits every DAC input level

The periodic nature of a PN sequence enables the use of an equivalent-time scope for obtaining the output samples. Equivalent time scopes are preferred to real-time scopes due to their superior linearity, resolution and bandwidth. Since the goal is to characterize the repetitive time varying response, the scope should be in averaging mode during measurements to eliminate the effect of random signal-independent non-idealities, including jitter and thermal noise. Once output samples are captured by the scope, a correlation receiver is implemented in MATLAB to find the phase of the periodic input sequence that corresponds to the captured output.

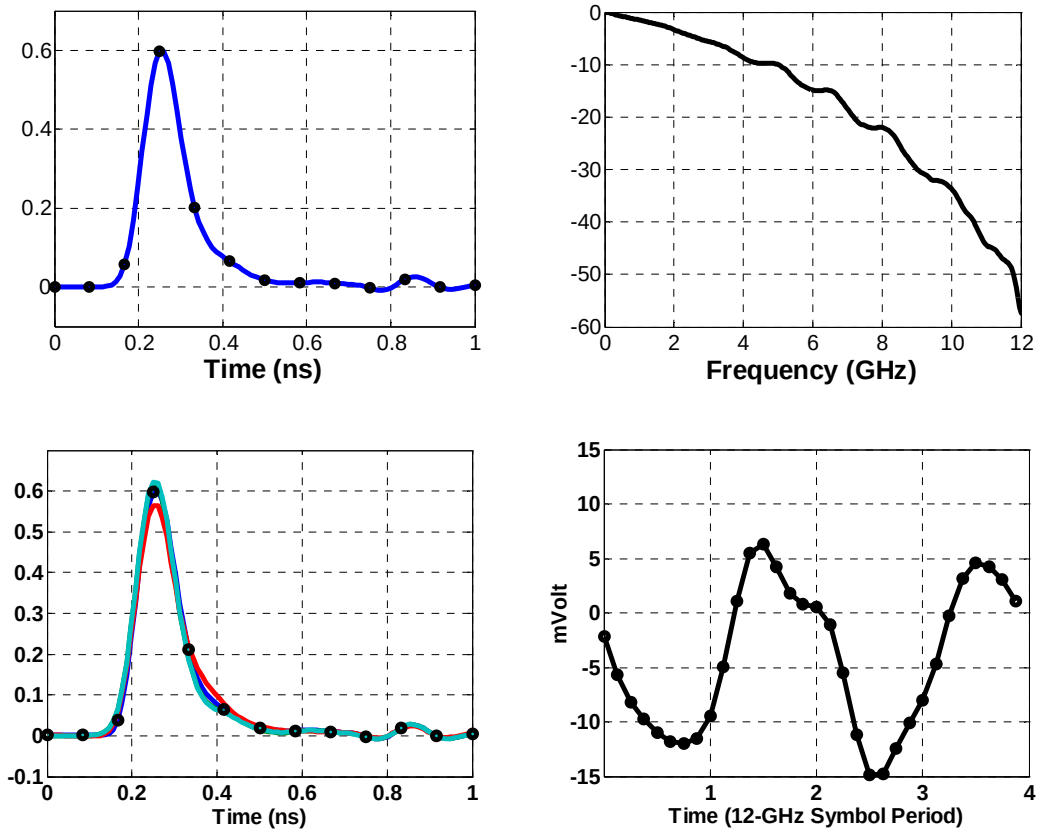


Figure 6.9: DAC characterization plots based on measure data: (a) LTI pulse response, (b) LTI frequency response, (c) CTV pulse responses corresponding to the 4-phases of the 4-way interleaved DAC, (d) Cyclo-stationary clock feed-through and noise plus DC offset.

Figure 6.9(a) shows the estimated time-invariant response of the DAC based on measured data and Figure 6.9(b) shows the corresponding frequency response. The estimated SDR is 26-dB in this case. However, since the DAC is 4-way parallelized, a CTV estimation with a period of 4 would better model its behavior. Figure 6.9(c) shows the four estimated time-variant responses of the DAC ($h_l^{0\text{-}3}$ in Equation 6.11). One can see that one of the responses is significantly different from the others. The estimated SDR in this case is 32-dB, which is 6-dB higher than time-invariant estimation and proves that the time-variant estimation better models the system. The frequency of the estimated cyclic bias, shown in Figure 11(d), is 3-GHz corresponding to g=4 in Equation 6.14. Single tone test of the same DAC at 750-MHz tone frequency

indicates 41-dB of SNDR [39], which clearly does not represent the wideband behavior of the circuit.

Estimation of higher order non-linear terms does not significantly improved the SDR beyond 32-dB. We speculate that the estimation is limited by coupling of data-dependent digital noise to the inputs of the DAC. Data-dependent noise cannot be eliminated from measurements by averaging. The strong coupling exists because the drivers at the input to DAC are connected to the digital supply by mistake.

6.3 Time Variant Digital Compensation

The results from the previous section indicate that 6 dB of improvement in SDR can be achieved if the interleaved DAC is treated as a CTV system instead of a time-invariant system. For example, these results predict that a time-variant linear equalizer can be employed to equalize the linear response of the DAC to achieve 32-dB of SDR at the middle of the eye-diagram, while a linear time-invariant equalizer can only achieve 26-dB of SDR. In most backplane link systems, the transmit equalizer is time-invariant. One way to transform a time-invariant equalizer to a cyclically time-variant equalizer with a period of 4, for example, is to parallelize the equalizer to four parallel branches, and let the four sets of equalizer taps in the four branches take independent values [54]. Such an equalizer already exists in our transmitter.

A block diagram of CTV compensated system is shown in Figure 6.10(a). The equalizer is a cyclically time-variant FIR filter with period of 4. Figure 6.10(b) shows an equivalent model of the system with the time-variant equalizer explicitly shown as four distinct equalizers operating on decimated sub-sequences (Equation 6.8) of the original input sequence[22]. Figure 6.10(c) shows the equalized model of the system where $c_\tau(n)$ represents the equalized time-variant response of the DAC:

$$c_s(n) = w_s(n) * h_\tau(n) \qquad s = 0,1,2,3 \tag{6.17}$$

[22] Please note that this system simplifies to a parallelized time-invariant equalizer if the FIR filters are the constrained to be the same.

In the above equation * represents time-variant convolution. The goal of the equalization is to satisfy the following conditions for all values of *s*.

$$\begin{aligned} &c_s(0) = 1, \\ &c_s(kp) = 0, \quad k \neq 0 \end{aligned} \tag{6.18}$$

where *p* is the up-sampling rate. In other words, we would like all the phases of the equalized response $c_\tau(n)$ to have a main tap value of 1, and zero interference. Therefore, a modified version of zero-forcing solution where regular convolution is replaced by time-varying convolution[23] can be used to find the optimal tap values for the four phases of the equalizer $w_s(n)$ (s = 0,1,2,3) independently.

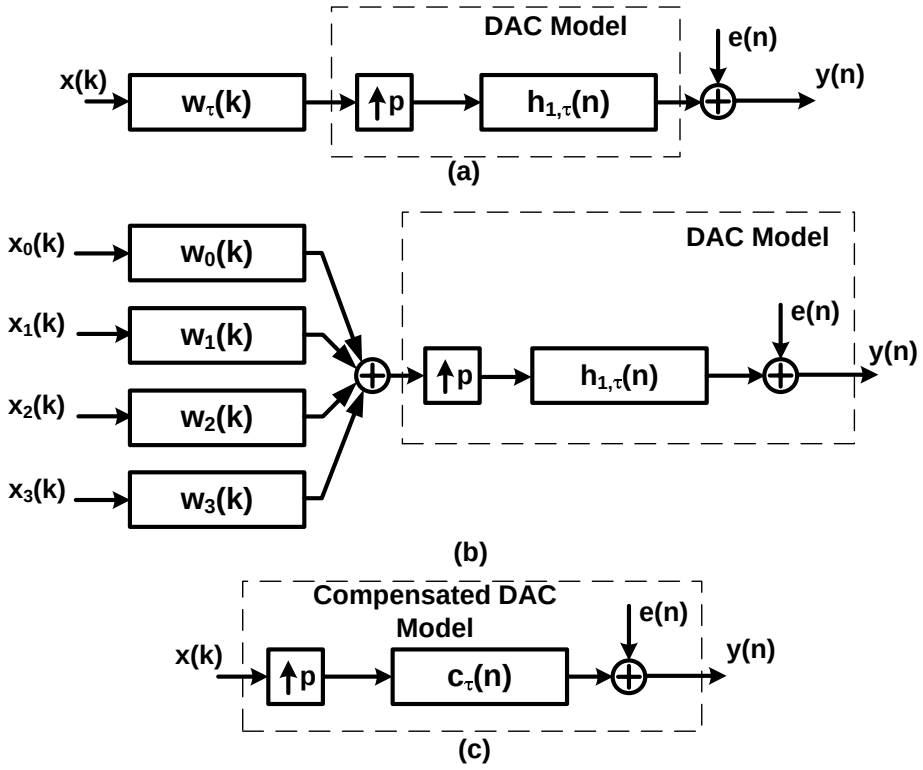
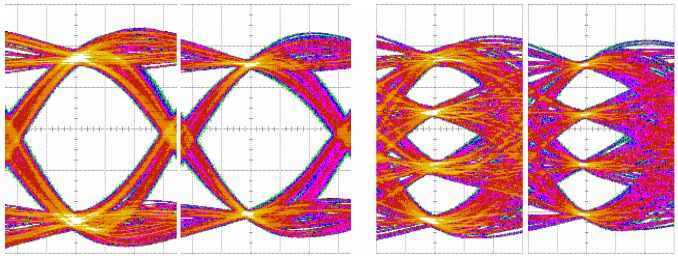


Figure 6.10: (a) DAC preceded by a time-variant equalizer (b) Equivalent model with the time-variant equalizer explicitly shown as four distinct equalizers operating on decimated sub-sequences of the original input sequence (c) Equivalent compensated system model.

[23] Time-varying convolution was defined in Section II.

Figure 6.11 depicts the eye diagrams of the 2-PAM and 4-PAM output signals with LTI and CTV equalization. CTV compensation produces a bigger eye opening by correctly suppressing interference from all phases of the interleaved DAC at the middle of the eye. Measured Signal to Interference and Distortion Ratio (SIDR) at the middle of the eye is 26-dB for the LTI equalized eye and 31-dB for the CTV equalized eye. The 1-dB difference between the measured SIDR after CTV compensation and the estimated SDR after CTV characterization is due to the residual interference. The CTV transmit equalizer can only eliminate Inter-Symbol Interference (ISI) up to 34-dB. This effect is not seen in time-invariant compensation because the 26-dB SDR is way below the estimated SIR of 34-dB.



(a) (b)

Figure 6.11: Time-invariant (left) versus time-variant (right) equalization for (a) 2-PAM, 12 Gb/s, and (b) 4-PAM, 24-Gb/s

6.4 Summary

In this chapter, we presented a new approach to characterizing wideband circuits based on LSE modeling of the circuit. This technique characterizes the system at full speed using a wideband random data, and provides an accurate wideband linearity measure as well as an accurate model for the system enabling the digital compensation of analog distortions. In addition, the technique can be used to analyze the effects and the nature of circuit non-idealities including non-linearity, periodic disturbances, and mismatches. The extension of the characterization method to CTV systems makes it applicable to a very wide range of circuits including mixers and time-interleaved data-converters. The techniques can also be extended to a cascade of weakly non-linear

circuits in the signal path. For example if a system includes a linear equalizer, a DAC, and a mixer in the transmitter, and a mixer, an ADC and a linear equalizer in the receiver, characterization technique can be applied to characterize the entire communication system combined with the communication channel. The model can then be included in a top level system simulator. In summary the techniques introduced in this chapter can be applied to guide the design according to the following methodology:

- ***Preliminary design***: Apply the Volterra series estimation techniques to gain insight to the sources of distortion in the circuit and quantify the potential performance improvements if the distortion terms are compensated for in the digital domain. Decide on analog-digital compensation partitioning.
- ***Circuit Specification***: Characterize the circuits, taking into account only the distortion terms that are not compensated for in the digital domain. Use the estimated circuit model and the estimated SDR in a system-level simulator to quantify the effect of analog distortion on system performance.
- ***Measurement***: Apply the same methods in the lab to characterize the fabricated circuits to make sure the implemented circuit meets the goals of the design. Estimate the Volterra series filters to gain insight to the non-idealities that are present in the fabricated circuit, but were not modeled correctly in pre-fabrication simulations.

The CTV equalization technique described in this chapter can also be extended to CTV Decision Feedback Equalization. The combination of CTV feed-forward and feedback equalization can be used to compensate for time-variant behavior of the entire communication.

Chapter 7

Conclusions

As the data rates of high-speed backplane links continue to increase, finding power efficient ways for signal transmission over dispersive channels become more important. Preliminary analysis of link channels indicates that data rate achieved by state of the art high-speed links is significantly lower than the available channel capacity achieved by optimal transmit power allocation. While discrete MT techniques achieve optimal transmit power-allocation with reasonable complexity in many other applications, the power/complexity overhead needed to create the A/D converter at the receiver front-end and the gigaops needed to perform the required digital signal processing make them inefficient for high-speed link applications. We need a lower complexity multi-tone approach for high-speed links and AMT is one such approach.

We show that an AMT system using linear transmit equalization and MIMO DFE circumvents these overheads, and can be built at comparable complexity to a BB solution. The AMT system achieves superior performance compared to BB systems by better utilization of transmit power over channels with equally spaced notch frequencies, and by parallelizing the receiver architecture, leading to better receiver sensitivity. The equalization parameters for this system can be found by solving a Second Order Conic optimization problem, or (sub-optimally) by finding the BER-constrained ZF solution.

AMT transmission can be combined with other conventional transmission techniques to create new transmission algorithms. In particular, we demonstrated how AMT and partial-response signaling can be combined to exploit the capacity of the

communication channel beyond a notch frequency. Such techniques combined with channel engineering techniques that modify the channel characteristics to suite the transmission algorithm can extend the data rate of conventional link systems. We further described a low complexity implementation of an AMT receiver which relied on redundant 1-bit samplers rather than mixers and integrators to decode the received data.

The AMT system in many ways is a generalization of a BB system. In particular, by parallelizing a conventional FIR filter and letting each of the parallel branches to be programmed independently, we built a transmitter that supports bandwidth scalable AMT and 2^M-PAM BB transmission. In addition, the degrees of freedom available in an AMT system enable cyclic time-variant equalization to compensate for analog distortions of the system. Such transmitter architecture is useful even for BB systems, since the additional degrees of freedom come at the cost of small additional area and no increase in active power.

To characterize the resulting system, we demonstrated an approach based on LSE modeling of the circuit. Our technique characterizes the system at full speed using wideband random data, and provides an accurate wideband linearity measure as well as an accurate model for the system enabling the digital compensation of analog distortions. In addition, the technique can be used to analyze the effects and the nature of circuit non-idealities including non-linearity, periodic disturbances, and mismatches. The extension of the characterization method to CTV systems makes it applicable to a very wide range of circuits including mixers and time-interleaved data-converters. We showed that the insight obtained from this characterization method can be applied to digital compensation of time-interleaved systems to achieve significant improvement in performance.

Overall, we believe that the AMT algorithm combined with channel-engineering can be a powerful method for extending the life of high-speed links, particularly in cost-sensitive applications where reducing package pin count has a large impact on the cost of the overall system. However, in order to further verify the feasibility of the algorithm, several further issues need to be addressed:

Our prototype transmitter, while ideal for investigating different transmission algorithms, is not necessarily the most power-efficient solution for practical AMT systems. In conventional BB systems, the cost of equalization is generally reduced by performing the necessary additions by current summation. In Appendix A we describe how a similar idea can be applied to AMT transmitters to reduce the power consumption of the transmitter. Implementation of a current-mode AMT transmitter would lead to a more accurate estimated of the power-consumption and complexity of an AMT transmitter.

This dissertation mainly focused on the system level analysis of the AMT system and creation of a very versatile transmitter platform. The next obvious step to verify the feasibility of this algorithm is to implement the receiver. A prototype receiver can serve to quantify the effect of jitter on the performance of the system, for example. Our preliminary analysis of jitter indicates that, at least for 2-PAM sub-channels, the effect of jitter is insignificant, but the analysis is dependent on certain assumptions about correlation and statistical properties of jitter that should be verified.

We also did not perform a rigorous study of clock and data recovery for AMT systems. Due to the interference from neighboring sub-channels, the data-edges in an AMT system are not as clear as they are for BB systems. As a result, the application of edge-based CDR to AMT systems could be difficult. Our preliminary study of a forwarded-clock CDR, in which clock is transmitted with the data over the same wire, however, indicates that such techniques may lead to excellent performance for both AMT and BB systems. The forwarded-clock CDR architecture is reviewed in Appendix B.

One way that AMT can improve link performance is by utilizing the portions of the channel bandwidth beyond a notch in the frequency domain, which are not accessible to BB systems. However, due to the same problems that limit the performance of 4-PAM systems, as we described in 0, multi-PAM sub-channels for AMT are also probably not practically possible. Therefore, the next step for improving performance is increasing bandwidth efficiency in terms of bits/Hz by implementing more accurate wideband receiver front-ends.

Furthermore, most of the published research so far has focused on links as single-input single-output systems. However, in many applications such as memory interfaces or chip-to-chip systems, data is transferred on multiple parallel lines forming a bus. Joint encoding and decoding of data on the data bus can potentially be a powerful technique to increase data rate. A parallel bus, with N lines each transmitting binary data can be viewed as a system with 2^N constellation points. One possible approach, for example is to add redundant lines to the system to form a larger constellation consisting of 2^{N+M} points and chose the best 2^N points in the constellation that would lead to the lowest power detection circuits, or minimum amount of noise (cross-talk, supply noise, etc) in the system. The challenge is to find the best codes where the hit in data rate due to redundancy is compensated for by the reduction in system power or system noise.

Appendix A

Current-Mode AMT Transmit Equalizer

In conventional BB systems, the cost of equalization is generally reduced by performing the necessary additions by current summation. Figure A.1 shows a block diagram of current-mode 4-tap symbol-spaced equalizer.

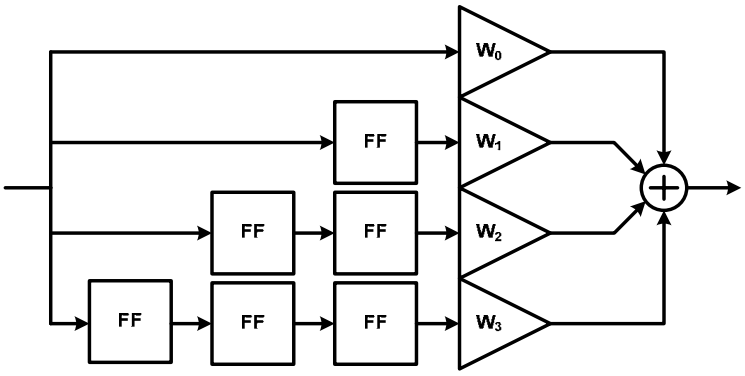


Figure A.1: Current-mode symbol-space equalizer. The triangles show current-mode drivers with strengths proportional to the tap coefficient.

An AMT equalizer consists of multiple fractionally-spaced (over-sampled) equalizers. The implementation of a fractional equalizer is different from a symbol-spaced equalizer because the former operates at a rate higher than its input data rate. Nonetheless, for practical systems where only a few equalizer taps are necessary, a similar idea can be applied to perform AMT equalization in current-domain to eliminate digital additions. Figure A.2 shows a functional diagram of a 4-tap 2x over-

sampled equalizer. For simplicity, we will focus on this particular example in this section.

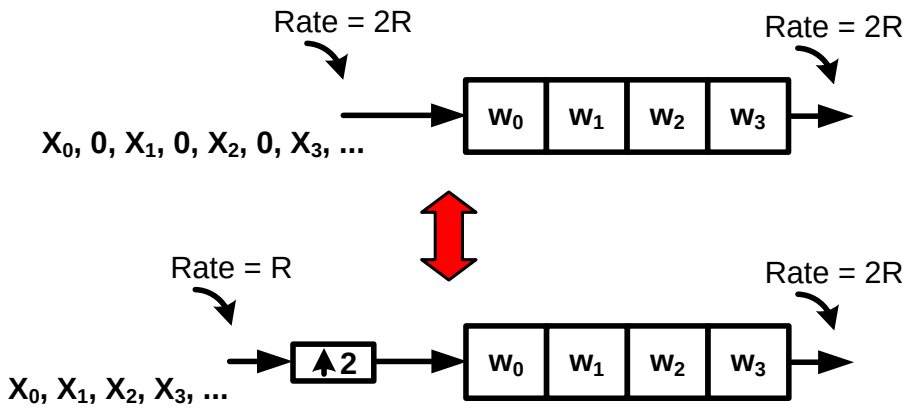


Figure A.2: Functional diagram of a 2x over-sampled equalizer.

A conventional way of implementing a fractional equalizer is poly-phase implementation [34]. Figure A.3 shows the poly-phase representation for a 4-tap 2x over-sampled equalizer.

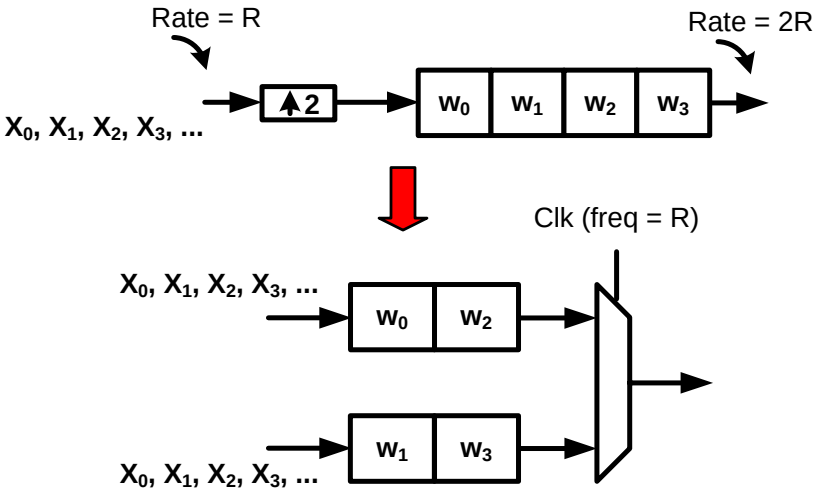


Figure A.3: Poly-phase representation of a 4-tap 2x over-sampled equalizer

In a current-mode implementation of the poly-phase representation of the fractional equalizer, tap values are proportional to the total output current (I_{Total}). Then the current corresponding to W_j (j = 1,2,3,4) is:

$$I_{Wj} = W_j / \max(W_0 + W_2, W_1 + W_3) * I_{Total}^{24} \quad (A.1)$$

For simplicity, let's first assume *W0 + W2 = W1 + W3* and that *W1>W0*. Figure A.4(a) shows a possible implementation for the 4-tap filter in Figure A.3.

The problem with this implementation is that the current sources are turned off while they are not driving the line. Figure A.4(b) shows an alternative architecture where the multiplexing is pushed to the input of the differential pairs. Figure A.4(c) shows the architectures in A.4(a) and A.4(b) in a block diagram form. If *W0 + W2 > W1 + W3*, or equivalently *(W0 + W2) – (W1 + W3) = 2Wfake*, the architecture can be changed to Figure A.4(d).

In general, for an arbitrary number of taps, a current-mode implementation can be obtained by dividing the line driver to segments and including 2:1 multiplexers in front of the segments as shown in Figure A.5. W_f represents a fake current cell in the case where the sum of the odd numbered taps is less than the even numbered taps. Other permutations of the current cells in Figure A.5(a) are also possible which leads to implementations different from Figure A.5(b).

The implementation above does not introduce any additional capacitance to the output node of the transmitter as long as the tap values are constant. However, in practical systems, the system should have the ability to program tap coefficients to arbitrary values without incurring much current overhead. Figure A.6 for example shows a dual data rate (DDR) implementation of a symbol-spaced equalizer where the main equalizer tap can take values in the range of [0,1] and the pre and post-cursor taps can take values between [0,3/8], sharing current with the main tap.

[24] For negative tap values, the polarity of the differential inputs to the current drivers is changed.

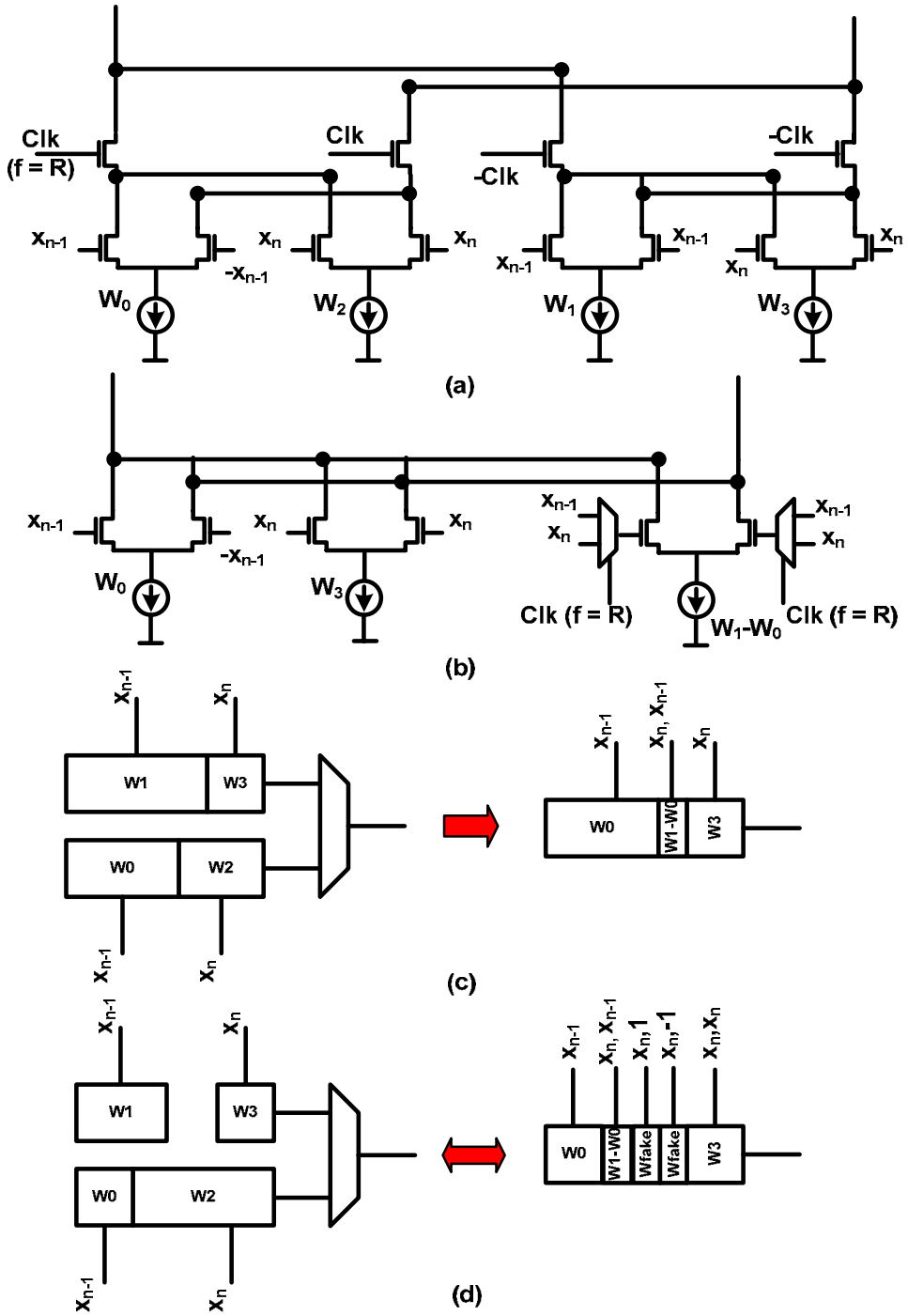


Figure A.4: (a) Current-mode implementation of the two-tap per phase filter assuming same total current per phase. (b) A better current-mode implementation (c) Block diagram representation of (a) and (b). (d) The case where W0+W2 > W1+W3 and W1 > W0.

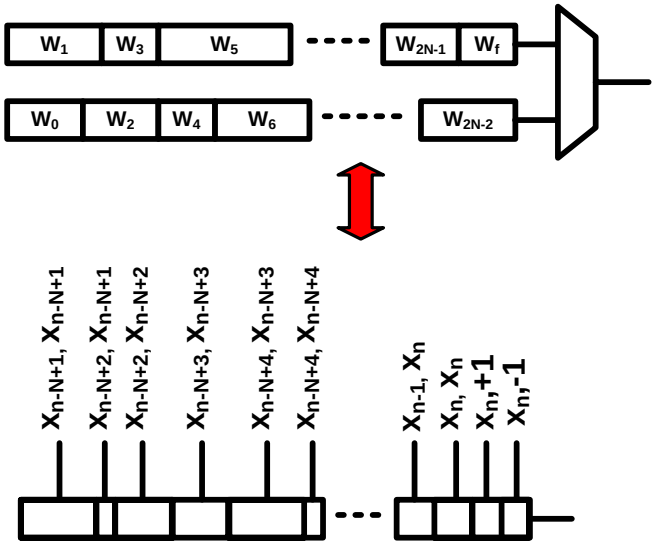


Figure A.5: Current-mode implementation for arbitrary number of taps

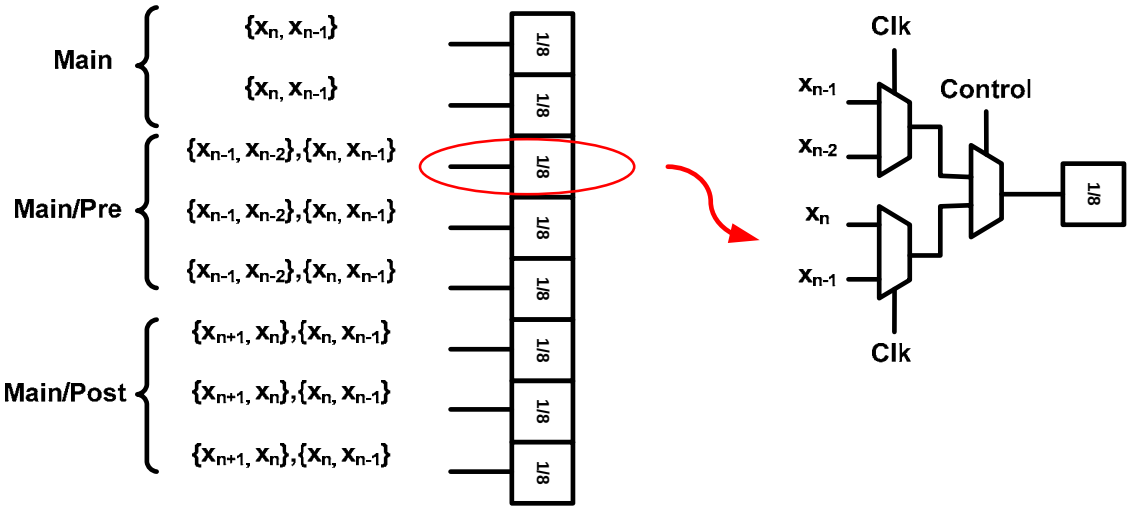


Figure A.6: A DDR current-mode symbol-spaced equalizer with tap sharing.

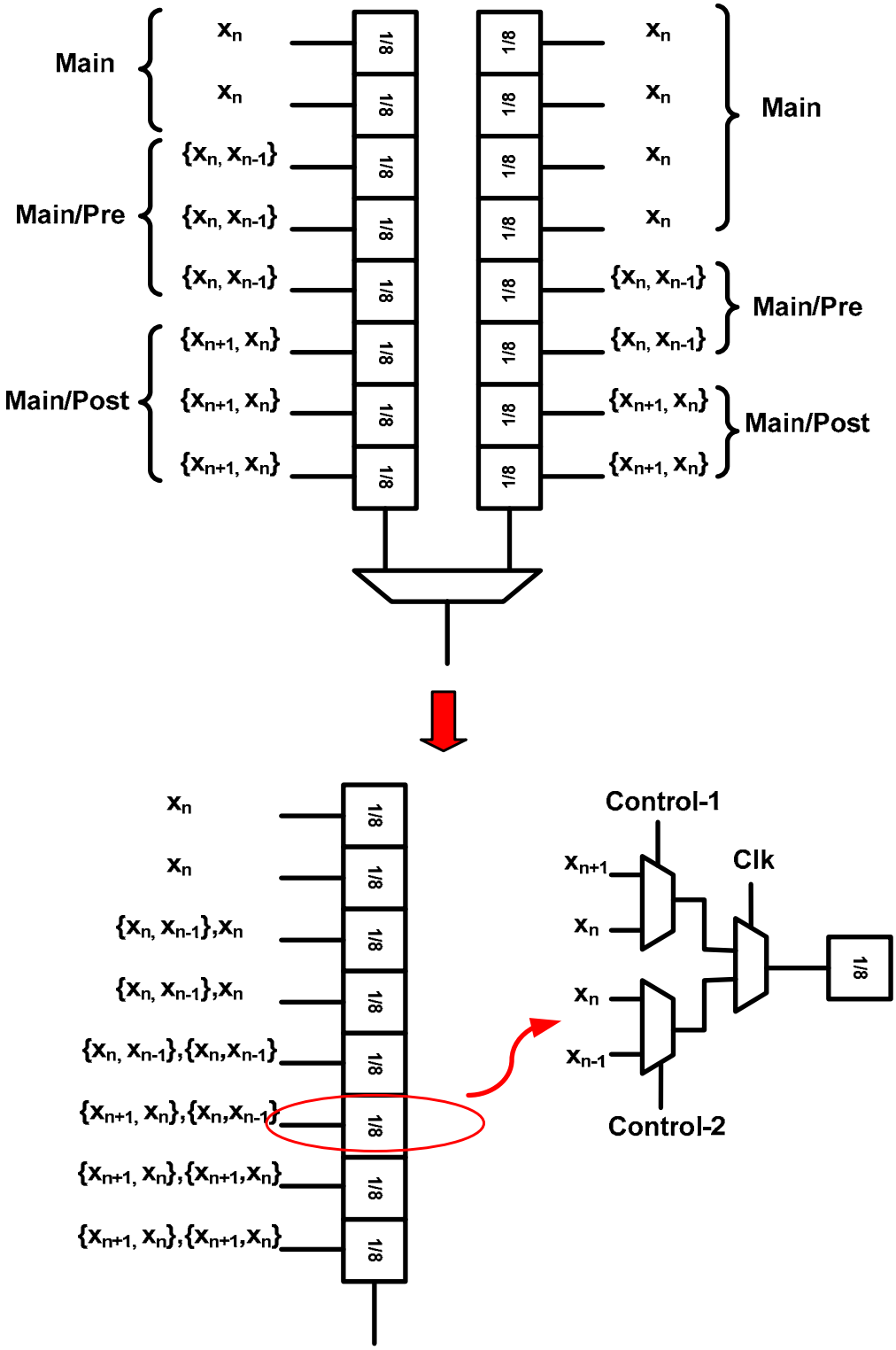


Figure A.7: Tap sharing example for a 2x over-sampled equalizer with 3 taps per phase.

Sharing in a current-mode implementation of an over-sampled equalizer is performed in the same way except that sharing is first performed per phase, and then the phases of the over-sampled equalizer are combined through 2:1 multiplexers at the input to the differential pairs. Figure A.7, for example, shows a 2x over-sampled equalizer with three taps per phase where the programmability range for the pre and post taps is different in the two phases of the equalizer. The "control" signal in both Figure A.6 and Figure A.7 are set before the equalizer operation starts, or slowly during adaptation.

It can also be shown that a cyclic time-variant symbol-spaced equalizer can be transformed in a way that is has the same structure as a fractional equalizer. As we described in Chapter 6, CTV equalization can be useful for compensating the time-variant nature of time-interleaved data converters (or systems). Therefore, an efficient implementation of a fractional equalizer would lead to an efficient implementation of a CTV equalizer.

Appendix B

Pilot-Based Clock and Data Recovery for AMT

Most state of the art high-speed links perform clock and data recovery at the receiver by detecting received signal transitions. In this method, since clock information is extracted from random data transitions, residual interference at zero-crossings can affect the performance. In addition, due to the interaction between the transmit equalizer's adaptation loop and the CDR loop, systems with both loops operating at the same time may lock to sub-optimal sampling points. The interaction exists because the transmit equalizer's adaptation loop changes the position of the zero-crossings at the receiver, and the zero-crossings affect the optimal sampling point, which in return affects the information used for equalizer adaptation [55]. Such edge-based methods are in particular not very useful for AMT since the zero crossings in AMT for a particular sub-channel carry strong interference caused by the other sub-channels.

An alternative to this approach is clock and data recovery based on a clock signal transmitted by the transmitter on top of data. Such recovery techniques have been known in other communication systems for a very long time [54], and have the advantage that the reference for clock recovery is data-independent. As a result, the interaction between CDR and equalizer adaptation loops does not exist in a pilot-based CDR. We will describe one such technique in the context of high-speed links in this appendix. The proposed CDR technique particularly fits the structure of an AMT receiver.

The line driver in link systems always has a zero-order-hold response. As a result the transmit spectrum generally has a null at (1/T), where T is the system symbol rate. As a result, transmit signal energy is significantly attenuated around the frequency corresponding to symbol rate (1/T). Therefore, if the transmitter adds a clock signal with frequency 1/T to the transmitted data, the receiver can extract clock frequency information by integrating the received signal over certain number of periods. Figure B.1 shows the transmitted signal spectrum and the corresponding receiver architecture. Once clock frequency is determined, a phase interpolator shifts the clock phase and sets the sampling point (τ) based on an eye maximization adaptive loop. In system with adaptive equalization, the adaptation rate τ should be at least as fast as the equalizer tap adaptation so that the system can track the optimal sampling point. Since the added clock to the spectrum repeats every symbol period, an integrating receiver, as it exists in an AMT receiver, entirely eliminates the clock from signal. Even without an integrating receiver, since clock is a repetitive pattern, it is always a constant offset at the sampling points at the receiver. Therefore, in BB system for example, it can be eliminated from signal before sampling with DC offset cancellation. Alternatively, the phase of the added clock can be adjusted at the transmitter such that its zero crossing falls at the sampling point.

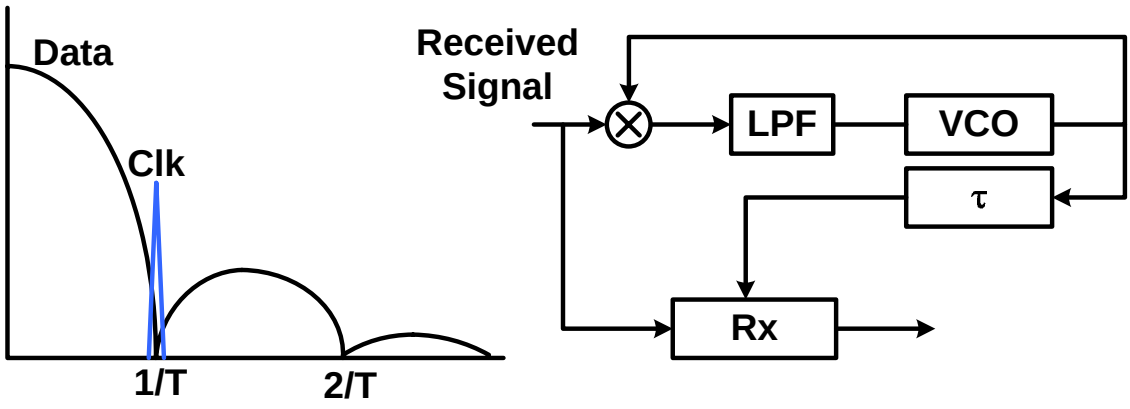


Figure B.1: Spectrum of the transmit signal with a clock tone added to frequency 1/T and the corresponding receiver.

A preliminary study of the performance of this type of CDR indicates that even for reasonably large CDR bandwidths, the performance of the CDR is affected by the thermal noise at the receiver input, rather than the transmitted data. Since the level of thermal noise within the bandwidth of the CDR is very small compared to link voltage levels, this means that even with large channel attenuation at 1/T, the pilot adds only a small voltage overhead to the transmitted signal.

Appendix C

Semi-Custom High-Speed Datapath Design Using Commericial ASIC Design Tools

Fully custom-designed integrated circuits can achieve significantly better performance compared to circuits designed following an automated ASIC design flow [57]. This higher performance is generally achieved through optimal logic design and sizing, better placement of the cells and optimized routing. For this reason many time-critical circuits are generally fully custom-designed. However, full custom-design of large complicated circuits is extremely time-consuming and difficult. In addition, since a fully custom flow is not closely linked to automated verification tools, the designer is often prone to making mistakes.

An intermediate approach which is often pursued is dividing a large system to smaller macros. The macros are individually designed and characterized into abstract timing and power tables similar to the way standard cells are characterized. Subsequently, an automated ASIC design and verification flow is performed at the top level using macros as the basic elements [58]. The individual macro are either automatically or custom-designed depending on their design constraints.

In this appendix we propose a semi-custom design methodology for the design of time-critical datapath macros for which most aspects of the design including cell design, sizing, placement, and routing needs to be controlled by the designer to meet the tight constraints. Our approach is based on the observation that even though ASIC

design tools may not match a human's perception of structure and regularity to perform optimal datapath design, they can be very efficient once they are told exactly what to do. Fortunately, modern tools provide sufficient knobs that can be exploited to control their behavior. However, to write efficient scripts for this purpose, the designer has to follow a strict design methodology. We start this appendix by explaining our proposed methodology to enable exploiting the capabilities of ASIC CAD tools at a high-level.

The proposed design flow was successfully applied to the design of the 24-Gb/s transmitter described in Chapter 5. We will use the equalizer as an example to describe each step of the flow in more detail. We will conclude by a summary of the advantages and disadvantages of the flow.

C.1 Semi-Custom High-Speed Datapath Design

Custom datapath design generally starts with finding the right logic architecture for the design and creating the basic building blocks. The sizing of the building blocks has to be performed together with floorplanning such that gate sizes account for wire loads as well. Even though tens of thousands of instances may exist in a datapath, the number of unique building blocks is generally limited due to the regular bit-sliced structure of datapaths. Multiple levels of pipelining with similar logic in between the stages further reduce the number of required gate sizes. Once the preliminary schematic design is finalized and the layout of the building blocks is available, the design has to be placed and routed. To achieve optimal performance, it is often necessary to place the building elements at exact locations to minimize the length of the wires. A datapath with a regular structure and optimal placement and routing inevitably will include many critical paths that need to be identified in the post-routed design after parasitic extraction. Timing verification of the critical paths is the last stage of the design.

Overall the above mentioned flow can become very long and tedious for datapaths with large number of logic instances. In particular, timing closure problems that lead

to multiple iterations of the design steps starting all the way from gate architecture or sizing can substantially add to the design time. Consequently, designers often try to leverage the power of the existing CAD tools to some extent to automate at least part of this flow.

Many of the design automation tools, however, only deal with abstractions of the digital gates, called library cells. Therefore, a key step towards leveraging the capabilities of such tools is to characterize the basic datapath building blocks as library cells [58]. This step can be done once the gate sizes are finalized, and if the appropriate characterization tools are properly set up, the characterization process takes only a few hours to a few days per cell. This is particularly efficient for a datapath since the number of unique building blocks can be very small [59]. With a library of cells at hand, the design entry can start at Hardware Description Language (HDL) level, and a synthesis tool can be used to create a preliminary timing report and a flat (Register Transfer Level) RTL netlist. The synthesizer would merely act as a translator mapping a hierarchical netlist to a flat RTL netlist without changing the custom designed gate sizes. The main advantage of an HDL design entry compared to a schematic level design entry is the possibility of a fast thorough functional verification. In addition, since commercial cell characterization tools perform a reliable logic verification of the cells versus their schematics, the functional verification of the datapath at HDL level is directly connected to the original schematic of the gates without any broken links and human intervention.

Creating a custom library of cells, however, doesn't solve the place and route problem if an exact placement and routing of the cells is desired. Automatic placement tools generally don't perform optimal placement of datapaths since they cannot extract regularity and the structure of the design as well as a human. However, all placement tools accept scripts as inputs and allow the user to determine the placement of every single cell. The problem is that to create such a script, the designer has to be able to name every individual cell in the RTL netlist, and calculate its exact placement coordinates. Fortunately, this step can be done easily and efficiently for a datapath if the HDL design is written hierarchically as we will describe in detail in Section C.2.

Performing the placement through an ASIC CAD tool has numerous advantages. First, the ASIC tool can handle the routing of the design which is the most timing consuming part of the design process. Once the optimal placement of a datapath is set by the designer with routing in mind, routing is greatly simplified from algorithmic point of view and the performance of an automatic routing tool can considerably improve. If necessary, some of the critical wires can be routed first by the tool or through a script before other less critical wires are automatically routed. In addition, the Place and Route (P&R) tool can perform a thorough static timing analysis of the design providing feedback on potential bottlenecks. Even if the designer is struggling for the last picoseconds of timing and ASIC P&R tools are not trusted for that precision, the tool can be used to identify all the paths that have small timing margins and create a transistor level netlist of those paths for accurate transistor level simulations. Lastly, the placement of parts of the design that are not critical, including the distribution of low-speed clocks can be left to the P&R tools within areas determined by the script.

Without a cell-based description of a design timing verification of the post-routed design has to be done with transistor level simulators. For large design such simulations can take an extremely long time and the designer is responsible for guaranteeing full coverage. However, by following the proposed flow, a static timing analysis tool can be used to perform a through and fast timing analysis on the post P&R netlist.

Overall, with the proposed flow, total changes in the architecture are possible with relatively simple changes in the scripts. Consequently, different architectures and floorplans can be tried efficiently in relatively short amount of time. In the case of the design of the equalizer, a total change of architecture from a 3-GHz datapath to a 2-way parallelized 1.5GHz datapath took about four man-weeks starting from modifying the basic cell architectures all the way through to post route timing analysis.

We will cover more specific details of this flow in the next section in the context of the design of the digital equalizer described in Chapter 5. The equalizer consists of more than 10,000 1.5-GHz instances, each of which is in a critical path due to the

symmetric nature of the design. In addition, the equalizer has more than 1000 slow flip-flops for storing the tap coefficients, each of which has a custom reset value. The tap coefficients stored in these flops are shifted in and out with four slow clocks, one per each channel.

C.2 Equalizer Design Flow

In terms of design methodology, the design of the equalizer poses a challenge. The operating rate of the 1.5-GHz datapath is so high that it requires the full attention of a custom design throughout the design process including circuit design, placement, routing and clock and power distribution. On the other hand, design and modification of the +10K-gate high-speed datapath in a fully manual flow can be very inefficient. The flow proposed in this appendix was specifically developed to address these issues. In our approach the entire design is treated as one entity, and while the high-speed part gets the attention of a custom design, the vast capabilities of ASIC design tools in placement, routing and versification are leveraged whenever possible. In addition the low-speed parts of the design (mainly the low-speed flip-flops and their clocks) are fully handled by the tools.

A flow chart of the design flow is shown in Figure 2. In our specific design the design entry was in Verilog, and design output was in post P&R DEF format. The flow makes use of HSPICE (Synopsis), Star RC (Cadence), Cell Rater (Magma), Abstract (Cadence), Verilog NC (Cadence), RTL Compiler (Cadence), MATLAB (Mathworks), SOC Encounter (Cadence), and PrimeTime (Synopsys). We will describe each stage of the design in more detail in the following sections.

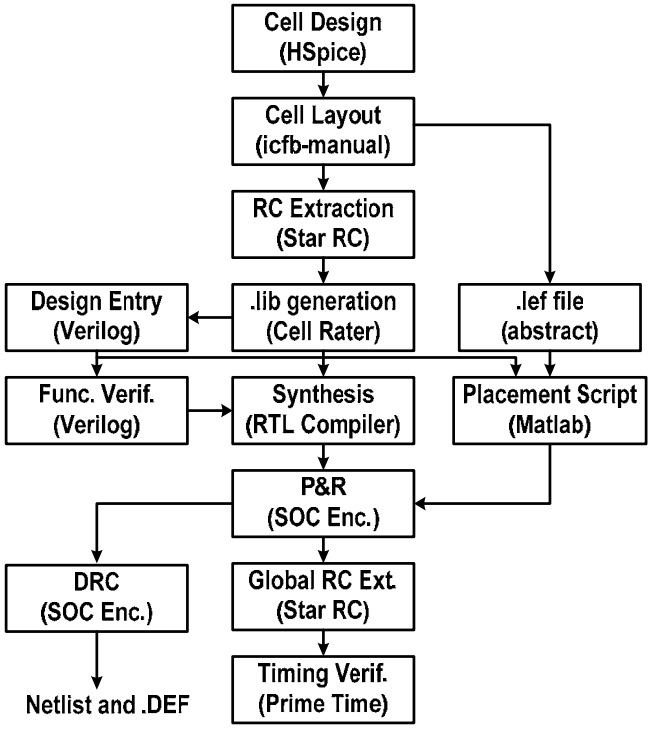


Figure C.1: Design flow using commercial ASIC design tools.

C.2.1 Cell design

The flow starts with the design of a high-speed cell library. The idea here is that at circuit level we custom design the high-speed part of the equalizer assuming we will have control over every aspect of the design including placement, routing, etc. Then we break the design into small cells, which are characterized as library cells by *Cell Rater* from *Magma*. The library is different from commercial cell libraries in that it only includes the exact cells and exact cell sizes that are used in the design. In this particular design, due to pipelining, only one size of 4:2 compressor, only one size of 4:1 multiplexer, 2 sizes of positive edge triggered flops and 2 sizes of negative edge triggered flops and a few other cells are required. As part of the characterization process, the characterization tool compares the functionality of the schematic to the Verilog functional description of the cell. This is the first stage of functional verification, starting from the schematic.

C.2.2 Cell layout

Library cells are laid out such that they can be placed next to each other by a tool like *tiles*. Therefore it is required that all cells have same height and their sides lie on a grid. Multiple row cells can also be designed with some restrictions. The compressor cell in this design is for example laid out as a four-row cell. The layout of the cells should be further abstracted to a LEF format to be usable by P&R tools. This stage was done using *Abstract* from *Cadence* in our design.

C.2.3 Hierarchical Verilog

The design entry is in Verilog. The high-speed part of the design is defined as a hierarchical netlist using the library of high-speed cells, fixing the type and size of the cells used in the design. The Verilog code for the high-speed part does not include any behavioral descriptions. The low-speed part of the design is written as a synthesizable Verilog code as is done in conventional ASIC design using a commercial standard cell library. Top level functional verification can be performed at this level for the whole design in Verilog.

Hierarchical Verilog netlist is directly linked to the regular and repetitive nature of a datapath. For example in this design a cascade of 10 compressors (used for adding 4 10-bit numbers) is repeated 7 times, and every time a bank of 4 flip-flops precedes every compressor. At a higher level, the whole equalizer consists of 4 identical phases, where each phase adds sixteen numbers. Hierarchical Verilog is essential to the flow, because it establishes a naming convention for the instances that appear in the post synthesis RTL netlist. For example, the compressor in the 2^{nd} bit-slice, in the 3^{rd} compressor column in the first stage of compression has a well defined name as a result of the hierarchical coding, which can be easily constructed for placement script generation. In addition this coding methodology reduces functional errors by using same modules repeatedly.

C.2.4 Synthesis

In order to generate an RTL netlist which only consists of unique modules (necessary for place and route) a synthesis CAD tool is used (*RTL Compiler* from *Cadence* in this design). The high-speed data-path is protected from synthesis by a "don't touch" or "preserve" attribute and directly appears in the output netlist. Therefore, the main tasks of the synthesizer on the high-speed part are uniquification and timing analysis. On the low-speed part, however, the synthesizer performs regular synthesis. Due to this capability, for example having parameterized values for the reset state of the low-speed flops is a relatively trivial task.

C.2.5 Hierarchical Placement

Hierarchical placement is the core of the flow and is motivated by the observation that placing very large number of high-speed cells can be very efficient through a script generator like MATLAB if the design has a good structure. MATLAB is used specifically because it has convenient arithmetic processing and reasonable graphics capabilities. The hierarchical placement can be best explained by an example:

- A function places a cell like a flip-flop or a compressor ("place_cell") at a given (x,y) coordinate
- A function places 4 flip-flops and a compressor ("place_regcomp") at a given (x,y) coordinate by calling the "place_cell" function.
- A function places 10 register-compressor units (place_regcomp_10) at a given coordinate (x,y) by calling the "place_regcomp" function
- …

Every placement function, at every level of hierarchy, is capable of placing its "leaf blocks" at correct coordinates relative to its (x,y) input coordinate. In addition, every function is capable of creating correct hierarchical instance names for all its leaf blocks relative to its input "root name". Therefore, as functions call each other, going down the hierarchy, correct coordinates and instance names are built for the cells lying

at the bottom of hierarchy. A placement in here is equivalent to writing a corresponding "placeInstance" command in a "tcl" file for later use by the p&R tool (*SOC Encounter* from *Cadence* in this design) and also showing a rectangle on MATLAB script for verification purposes. The GUI feature of the tool is very valuable for catching bugs and visualizing the floorplan.

The placement tool is provided with the height and width of the library cells as parameters and recursively calculates the height and width of different macro modules at different levels of the hierarchy. With this knowledge of heights and widths, at any level of hierarchy, the placement of different macro modules next to each other is performed by calling different placement functions and incrementing the x and y coordinates accordingly. The following is part of the placement code at the place_datapath level to further demonstrate the procedure:

```
.....
W_1ST_STAGE = W_MUX+W_REG+W_COMP;
% place u_comp_Lev1_No1
x = x0; y = y0; w = W_1ST_STAGE;
inst_name=strcat(root_name,'u_comp_Lev1_No1'/');
place_first_stage(x, y, don't_flip, inst_name);
% place u_comp_Lev1_No2
x = x+w; y = y0; w = W_1ST_STAGE;
inst_name = strcat(root_name, 'u_ comp_Lev1_No2/');
place_first_stage(xf, y, don't_flip, inst_name);
...
```

Since the tool uses relative coordinates for placement, changing the placement of a cell or even the floorplan is equivalent to changing the relative placement of functions in the hierarchical MATLAB code.

Figure C.2 shows different levels of hierarchical placement as shown by MATLAB. Figure C.3 shows fully placed equalizer by MATLAB including clock grid buffers. Figure C.4(a) shows the fully placed design in *SOC Encounter*.

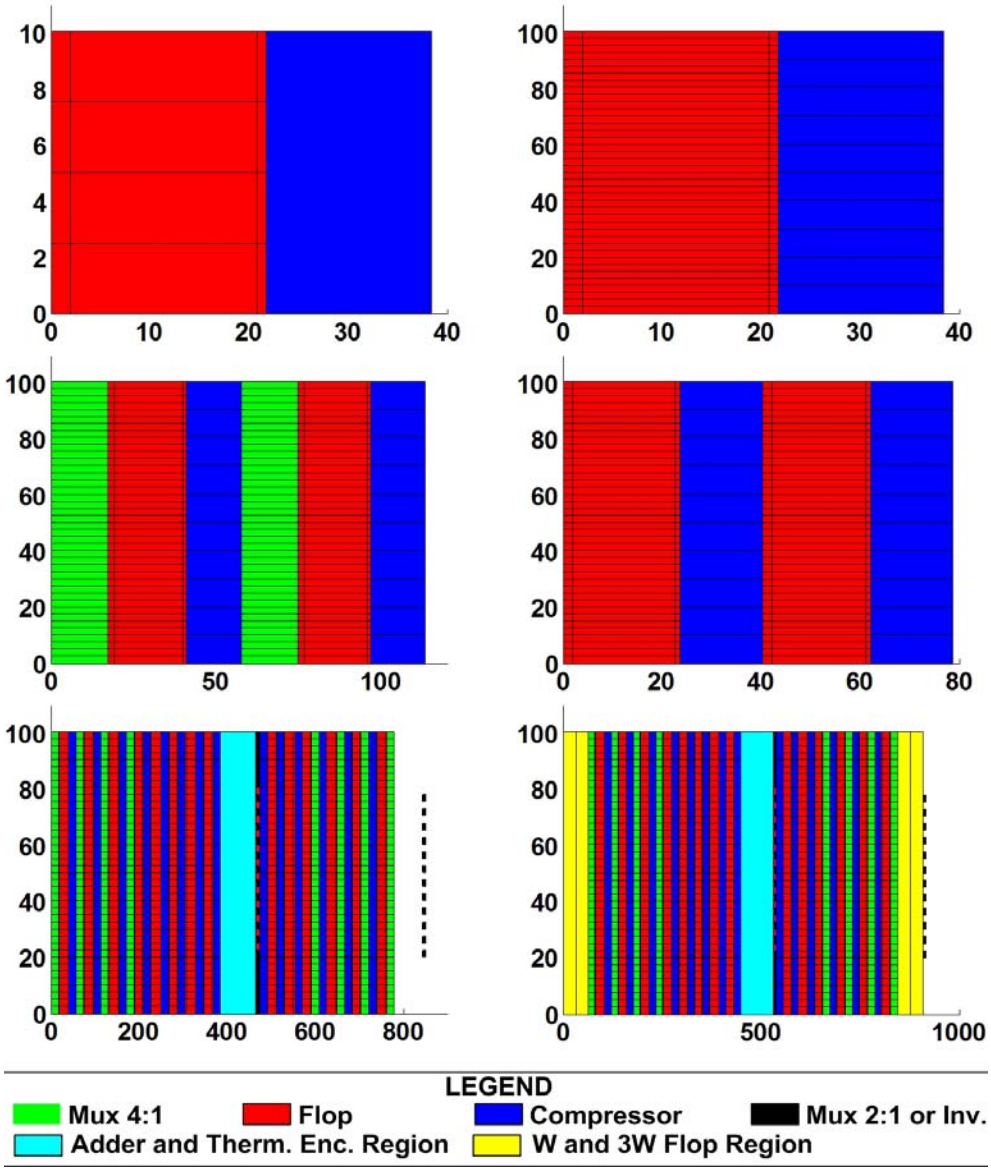


Figure C.2: Hierarchical placement with MATLAB. X and Y axes are in micron and are to scale. (a) compressor with 4 flops at input (regcomp) (b) cascade of 10 regcomp units (regcomp_10) (c) first stage of compression – two regcomp_10 units, one with positive edge flops and one with negative, each flop preceded by a 4:1 multiplexer (d) second or third stage of compression – two regcomp_10 units (e) full high-speed data-path for one phase of the equalizer (f) One phase of the equalizer.

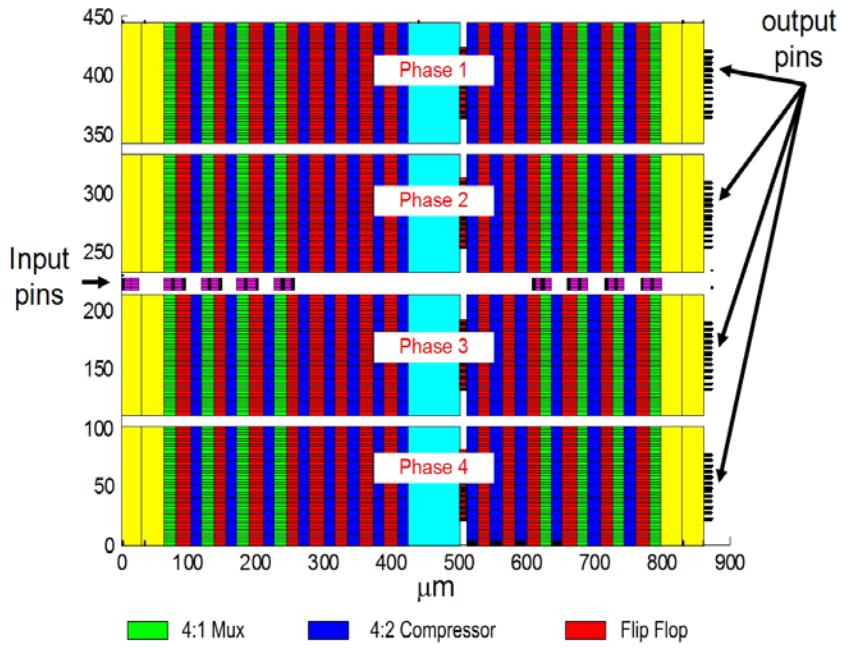


Figure C.3: Equalizer floorplan in the MATLAB placement tool. Yellow rectangles on the sides are areas where low-speed flip-flops holding equalizer tap coefficients are placed by the P&R tool. The cyan rectangles in the middle are areas where the adder and the thermometer encoder are placed by the P&R tool. The pink rectangles in the middle row are latches, implementing the shift registers for the input data sequence. X and Y axes show actual sizes in microns.

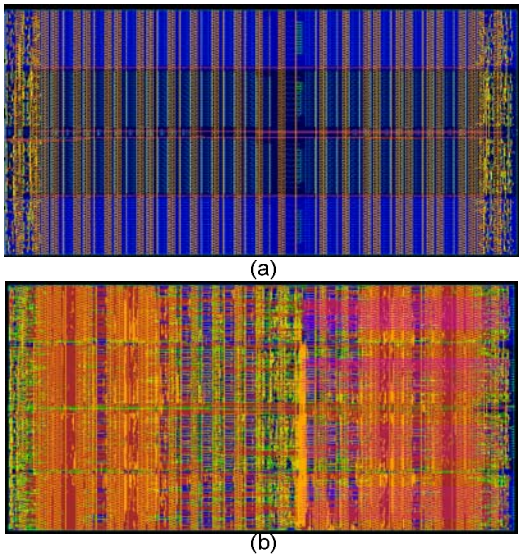


Figure C.4: (a) Placed equalizer in *SOC Encounter*. Yellow lines on the sides show low frequency clock grid. Red grid shows the 1.5GHz clock mesh. (b) Fully routed equalizer.

C.2.6 Routing

In this design all routing was done automatically by *SOC Encounter*. However, in order to make sure critical nets like the long wires connecting thermometer encoder outputs to equalizer output pins are routed efficiently, routing was done in several stages with critical nets routed first. Figure 5(b) shows the fully routed equalizer in *SOC Encounter*.

C.2.7 Clock Distribution

As mentioned earlier, this design has 5 clock domains, four of which are low-speed clocks with only hold-time constraints. These clocks are automatically distributed in the yellow regions (two left-most and two right-most columns) of Figure 4 by the tool. The high-speed 1.5-GHz clock domain has both hold and setup time constraints and has to reach all high-speed flops (red) in Figure 4. This clock is routed in the form of an 8-grid. A MATLAB script is used for placing clock buffers and routing of clock wires. Clock pins for the flip-flops are purposely brought to a top level metal layer where the last level of clock grid is routed so that pins make automatic contact with clock grid without the need for local clock routing. To achieve a higher level of confidence, the clock grid was also modeled and simulated in Spice.

C.2.8 Timing Verification

PrimeTime from *Synopsys* was used in this design for timing verification on the parasitic extracted post-route netlist generated by the P&R tool. Based on the timing analysis engine, our design met the 1.5-GHz cycle time at S/S/125^0/0.95V corner. Measurement results in the lab at room temperature indicate correct operation up to 1.9GHz (corresponding to a throughput of 29-Gb/s). Analog clocking circuits in the transmitter failed to operate beyond this point. The digital equalizer overall occupies an area of 0.4mm^2 and consumes 350mW of power at 24-Gb/s in nominal conditions.

C.3 Advantages and Disadvantages

The flow described above has several advantages and a few disadvantages compared to a manual flow.

Script based versus schematic based: The flow presented here is a script-based flow as opposed to a schematic-based flow, and consequently, carries all the advantages associated with a script-based flow. For example, it is much easier to change architecture because changes occur in a Verilog file describing the circuit. It is much easier to change the floorplan, clock or power distribution as well because significant changes are generally equivalent to simple modifications of the MATLAB code. On the other hand, the flow also carries the disadvantages of a script-based design including script generation for placement, power and clock distribution. It also has the overhead of library generation and stricter layout rules.

Static timing verification versus transistor level: This flow allows for static timing verification which is a lot faster and has better coverage compared to transistor level simulators like SPICE and NanoSim. However, overall timing verification is certainly slightly less accurate than transistor level simulations since inaccuracies from cell generation, parasitic estimation and timing analysis engines all add up, which may lead to about 10% inaccuracy in overall simulated results. Of course it is always possible to simulate parts of the design in Spice at the very last stage for additional confidence in the results. In addition, the flow overall leads to very reliable functional verification as well since functional verification starts from schematic at the cell characterization stage and extends all the way to top level Verilog.

The design flow presented in this appendix indicates that even though CAD tools may not be optimized for high-speed applications, they generally provide many knobs that can be employed to leverage their vast capabilities for a reliable design. Although the original design may take a long time because of dealing with many tools in a non-conventional way, once the issues are identified, redesigns and modifications can be significantly faster and more reliable than non-automated flows. This flow is applicable to any high-speed datapath with regular structure. The hierarchical

placement portion of the flow may even be employed for the design of analog blocks with regular layouts like Digital to Analog Converters. MATLAB code may be replaced by a Skill code or any other suitable programming language.

Bibliography

[1] K. Change *et al.*, “Clocking and circuit design for a parallel I/O on a first-generation CELL processor,” *IEEE International Solid-State Circuits Conference*, Feb. 2005, pp. 526-527.

[2] J. A. Kahle *et al.*, “Introduction to the Cell multiprocessor,” *IBM Journal of Research and Development*, July-Sep. 2005, pp. 589-604.

[3] M. Meghelli *et al.*, "A 10-Gb/s 5-Tap DFE/4-Tap FFE Transceiver in 90-nm CMOS Technology," *IEEE Journal of Solid State Circuits*, pp. 2885-2900, December 2006.

[4] R. Palmer, J. Poulton, et al, "A 14mW 6.25Gb/s Transceiver in 90nm CMOS for Serial Chip-to-Chip Communication," *IEEE International Solid-State Circuits Conference*, pp. 440-441, February 2006.

[5] B. Casper *et al*., “A 20Gb/s Forwarded Clock Transceiver in 90nm CMOS,” *IEEE International Solid-State Circuits Conference*, pp. 90-91, Feb. 2006.

[6] V. Stojanović, A. Amirkhany and M.A. Horowitz, “Optimal linear precoding with theoretical and practical data rates in high-speed serial-Link backplane communication,” *IEEE International Conference on Communications*, June 2004.

[7] B. Saltzberg, "Performance of an Efficient Parallel Data Transmission System," *IEEE Transactions on Communications*, vol. COM-15, no. 6, Dec. 1967, pp. 805-11.

[8] Weinstein and P. M. Ebert, "Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform." *IEEE Transactions on Communications*, vol. COM-19, no. 5, Oct. 1971, pp, 628-634.

[9] B. Hirosaki, "An Orthogonally Multiplexed QAM System using the Discrete Fourier Transform," *IEEE Transactions on Communications*, vol. COM-29, no. 7, Jul. 1981, pp. 982-89.

[10] G. Cherubini, E. Eleftheriou, S. Oker and J. M. Cioffi, "Filter Bank Modulation Techniques for Very High Speed Digital Subscriber Lines," *IEEE Communications Magazine*, Vol. 38, issue 5, pp. 98-104, May 2000.

[11] A. Amirkhany, A. Abbasfar, V. Stojanović and M.A. Horowitz, "Practical Limits of Multi-Tone Signaling over High-Speed Backplane Electrical Links," *IEEE International Conference on Communications*, June 2007.

[12] K. Poulton *et al.*, "A 20GS/s 8b ADC with a 1MB memory in 0.18µm CMOS," *IEEE International Solid-State Circuits Conference*, Feb. 2003, San Francisco.

[13] Y. Lin, H. Liu, and C. Lee; "A 1-GS/s FFT/IFFT Processor for UWB Applications," *IEEE Journal of Solid State Circuits*, pp. 1726-1735, Aug. 2005.

[14] M. Harwood et al., "A 12.5Gb/s SerDes in 65nm CMOS Using a Baud-Rate ADC with Digital Receiver Equalization and Clock Recovery," *IEEE International Solid-State Circuits Conference*, pp. 436-437, February 2006.

[15] J. Zerbe et al., "Equalization and clock recovery for a 2.5–10-Gb/s 2-PAM/4-PAM backplane transceiver cell," *IEEE Journal of Solid-State Circuits*, pp. 2121–2130, Dec. 2003.

[16] T. Toifl *et al.*, "A 22-gb/s PAM-4 receiver in 90-nm CMOS SOI technology," *IEEE Journal of Solid State Circuits*, pp. 954-965, April 2006.

[17] J. Stonick et al., "An adaptive PAM-4 5-Gb/s backplane transceiver in 0.25-mm CMOS," *IEEE Journal of Solid State Circuits*, pp. 436–443, Mar. 2003.

[18] A. Bessios, W. Stonecypher, A. Agarwal and J. Zerbe, "Transition-limiting codes for 4-PAM signaling in high speed serial links," *Global Communications Conference*, Dec. 2003.

[19] S. Kasturia and J. H. Winters, “Techniques for high-speed implementation of nonlinear cancellation,” *IEEE Journal on Selected Areas in Communications*, pp. 711–717, June 1991.

[20] V. Stojanovic *el al.*, “Autonomous dual-mode (PAM2/4) serial link transceiver with adaptive equalization and data recovery,” *IEEE Journal of Solid State Circuits*, pp. 1012-1026, April 2005.

[21] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004

[22] J. H. Sinky, M. Duelk, and A. Adamiecki, "High-speed Electrical Backplane Transmission Using Duobinary Signaling," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 53, no. 1, January 2005, pp. 152-160.

[23] J. Lee, M-S. Chen, and H-D. Wang, “A 20Gb/s Duobinary Transceiver in 90nm CMOS,” *IEEE Solid States Circuits Conference*, pp. 102-103, Feb 2008.

[24] J. M. Cioffi, *EE379A course reader*, Stanford University (available at http://www.stanford.edu/class/ee379a/)

[25] A. Wiesel, Y.C. Eldar and S. Shamai, “Linear precoding via conic optimization for fixed MIMO receivers,” *IEEE Transactions on Signal Processing*, pp. 161-176, Jan. 2006.

[26] J. Campello, “Practical bit loading for DMT,” *IEEE International Conference on Communications*, pp. 796-800, 1999.

[27] A. Amirkhany, A. Abbasfar, V. Stojanović and M.A. Horowitz, “Analog Multi-Tone Signaling for High-Speed Backplane Electrical Links,” *Global Communications Conference*, Nov. 2006.

[28] K. Maharatna, E. Grass, and U. Jagdhold, “A 64-point fourier transform chip for high-speed wireless LAN application using OFDM,” *IEEE Journal of Solid-State Circuits*, pp. 484–493, Mar. 2004.

[29] C. Chiu et al., “A 64-point Fourier transform chip for video motion compensation using phase correlation,” *IEEE Journal of Solid-State Circuits*, pp. 1751–1761, Nov. 1996.

[30] H. Sorensen, D. Jones, M. Heideman and C. Burrus, “Real-valued fast Fourier transform algorithms,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pp. 849 – 863, June 1987.

[31] P. Schvan *et al.*, “A 24GS/s 6b ADC in 90nm CMOS,” *IEEE Solid States Circuits Conference*, pp. 544-545, Feb 2008.

[32] A. Amirkhany, V. Stojanović and M.A. Horowitz, “Multi-tone Signaling for High-speed Backplane Electrical Links,” *Global Communications Conference*, Nov. 2004.

[33] S. Sidiropoulos and M. Horowitz, “A 700-Mb/s/pin CMOS signaling interface using current integrating receivers,” *IEEE Journal of Solid State Circuits*, pp. 681-690, May 1997.

[34] P. Vaidyanathan, *Multirate systems and filter banks*, Prentice Hall, 1993.

[35] V. Stojanović and M. Horowitz, “Modeling and analysis of high-speed links,” *Custom Integrated Circuits Conference*, Sep. 2003.

[36] W. Beyene, “Controlled ISI Design Techniques of Conventional Interconnect Systems for Data Rates Beyond 20Gb/s”, *Electrical Performance of Electronic Packaging*, Oct. 2006.

[37] J. Savoj *et al.*, “A New Technique for Characterization of Digital-to-Analog-Converters In High-Speed Systems,” *Design Automation and Test in Europe*, March 2007.

[38] S. Crozier, D. Falconer and S. Mahmoud, "Least sum of squared enom (LSSE) channel estimation", *IEE Proceedings F*, Volume 138, Aug. 1991, Pages: 371-378.

[39] J. Savoj *et al.*, “A 12-GS/s Phase-Calibrated CMOS Digital-to-Analog Converter,” *Symposium on VLSI Circuits*, June 2007.

[40] M. Nagamatsu *et al.*, “A 15-ns 32×32-b CMOS multiplier with an improved parallel structure,” *IEEE Journal of Solid State Circuits*, pp. 494-497, April 1990.

[41] A. Amirkhany, A. Abbasfar, J. Savoj and M. Horowitz, “Time-Variant Characterization and Compensation of Wideband Circuits,” *Custom Integrated Circuits Conference*, Sep. 2007.

[42] B. Murmann and B. E. Boser, ”A 12-bit 75-MS/s Pipelined ADC Using Open-Loop Residue Amplification,” *IEEE Journal of Solid State Circuits*, pp. 2040-2050, Dec. 2003.

[43] A. Panigada and I. Galton, "Digital Background Correction of Harmonic Distortion in Pipelined ADCs," *IEEE Transactions on Circuits and System I*, Sep. 2006.

[44] Walt Kester, *High-Speed Design Techniques*, Analog Devices, 1996.

[45] M. Vanden Bossche, J. Schoukens and J. Reenneboog, “Dynamic Testing and Diagnostics of A/D Converters,” *IEEE Journal of Solid State Circuits*, pp. 775-785, Aug. 1996.

[46] B. Razavi, *RF Microelectronics,* Prentice Hall, 1998.

[47] N. Giaquinto, A. Trotta, “Fast and Accurate ADC Testing Via an Enhanced Sine Wave Fitting Algorithm,” *IEEE Transactions on Instrumentation and Measurement*, pp. 1020-1025, Aug. 1997.

[48] J. Schoukens, R. Pintelon and T. Dobrowiecki, ”Linear Modeling in the Presence of Nonlinear Distortions,” *IEEE Transactions on Instrumentation and Measurement*, pp. 786-792, Aug. 2002.

[49] K. Vanhoenacker, T. Dobrowiecki and J. Schoukens, "Design of Multisine Excitations to Characterize the Nonlinear Distortions During FRF-Measurements," *IEEE Transactions on Instrumentation and Measurement*, pp. 1097-1102, Oct. 2001.

[50] J.S. Bendat, *Nonlinear System Analysis and Identification from Random Data*, John Wiley and Sons, 1990.

[51] A. Zhu, J. Dooley and T. J. Brazil, "Simplified Volterra Series Based Behavioral Modeling of RF Power Amplifiers Using Deviation Reduction," *IEEE MTT-S International Microwave Symposium*, June 2006.

[52] S. Chan, T. Stathaki, A. Constantinides, "Adaptive Weighted Least Squares Algorithm for Volterra Signal Modeling," *IEEE Transactions on Circuits and Systems I*, April 2000.

[53] E. Alon, V. Stojanovic, M. Horowitz, "Circuits and Techniques for High-Resolution Measurement of On-Chip Power Supply Noise," *IEEE Journal of Solid-State Circuits*, April 2005.

[54] V. Stojanović, G. Ginis, and M. A. Horowitz, "Transmit pre-emphasis for high-speed time-division-multiplexed serial-link transceiver," *IEEE International Conference on Communications*, April 2002.

[55] J. Ren *et al.*, "Precursor ISI Reduction in High-Speed I/O," *Symposium on VLSI Circuits*, June 2007.

[56] A. Viterbi, "Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques," *Proceedings of the IEEE*, pp. 1737–1753, 1963.

[57] D. Chinnery and K. Keutzer; "Closing the gap between ASIC and custom: An ASIC perspective", *Design Automation Conference*, June 2000.

[58] G. A. Northrop and P. Lu, "A Semi-custom Design Flow in High-performance Microprocessor Design," *Design Automation Conference*, June 2001

[59] W. J. Dally and A. Chang, "The role of custom designs in ASIC chips," *Design Automation Conference*, June 2000.