Power Supply Switching for a mm-Wave Asymmetric Multilevel Outphasing Power Amplifier System

by

Jonathon David Spaulding

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering In Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

Author .....
Department of Electrical Engineering and Computer Science
August 31, 2010

Certified by .....
Joel Dawson
Associate Professor
Thesis Supervisor

Certified by .....
Taylor Barton
Instructor
Thesis Supervisor

Accepted by .....
Dr. Christopher J. Terman
Chairman, Department Committee on Graduate Theses

Power Supply Switching for a mm-Wave Asymmetric Multilevel Outphasing Power Amplifier System

by

Jonathon David Spaulding

Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2010, in partial fulfillment of the
requirements for the degree of
Master of Engineering In Computer Science and Engineering

Abstract

This thesis demonstrates power switches to be used in our new Asymmetric Multilevel Outphasing (AMO) transmitter architecture at mm-wave frequencies. The AMO topology breaks the linearity vs. efficiency design objective in radio frequency power amplifiers (PAs) which has until now appeared to be fundamental. These power switches allow for the modulation of the PA supply rail between four discrete levels at a maximum sampling rate of 2 GHz. This modulation results in a higher average system efficiency by reducing the outphasing angle between the phase paths. This work was designed in a 130-nm SiGe BiCMOS process.

Thesis Supervisor: Joel Dawson
Title: Associate Professor

Thesis Supervisor: Taylor Barton
Title: Instructor

Acknowledgments

I would like to thank: My advisor, Professor Joel Dawson, for his advice, mentorship, and support through the project. Ms. Taylor Barton for her inexhaustable patience and encouragement. This project would not be without her. Mr. Zhen Li and Mr. Sungwon Chung for their fantastic problem solving abilities. Mr. Wei Tai for his power amplifier work on the project. The Frank and Eva Buck Foundation for their financial support. Finally, my family and friends, for their support through the arduous process.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

The wireless transmission of data is an increasing problem in our world today. There is an increasing need to transmit data faster wirelessly, while maintaining overall system efficiency. In his classic paper [1], Shannon proposed the equation which has become a cornerstone of radio frequency (RF) circuit design. This equation, shown below in equation 1.1, indicates that the channel capacity C (in bits/second) is proportional to the bandwidth B of the system, and the log of the signal-to-noise ratio (SNR). Thus, to increase the channel capacity of an RF link, one might either attempt to increase the bandwidth of the system, or increase the SNR. It is often easier to design a system for a small relative bandwidth, so it makes sense to increase the carrier frequency so that total bandwidth stays large while relative bandwidth decreases. One approach emerging in technology today is to use mm-wave frequencies for an RF link. mm-Wave frequencies are defined as frequencies who have a wavelength of 1 to 10 mm, which corresponds to frequencies in the 30-300 GHz range. This approach is beneficial because it allows the system to maintain a large bandwidth while keeping a small normalized bandwidth. That is, the system might have a large bandwidth, but may still be narrow band in nature given the high frequency of the carrier frequency.

$$C = B * log_2(SNR + 1) \tag{1.1}$$

The fabrication of smaller process nodes in recent years has allowed for the advent of mm-wave technologies. mm-Wave technology is currently being implemented to solve a number of problems, including visual detection for aircraft, automotive proximity detection, and faster wireless transmitters for communications. This thesis deals with a unique architecture which aims to help solve the third problem while maintaining good system efficiency.

Communications devices are ubiquitous in today's society. Cellular phones are quite common, and hand held devices which transmit large quantities of data are becoming so. With the advent of hand held devices, demand has risen for radio-frequency transmitters that are both power efficient and highly linear. Power efficient transmitters allow for long battery lifetimes, while high linearity maximizes the achievable bit rate. In short, there is demand to make devices last longer on a single battery charge while transmitting data faster. The difficulty lies in the fact that linearity and efficiency are competing design objectives. Linear amplifiers, such as the Class AB amplifier, are inefficient, while switching amplifiers are efficient, but nonlinear. While I used cellular phones as an example here, they are not the only technology which would benefit from a better RF transmitter. One could imagine wireless home theater systems or short range ad-hoc computer networks, for example, benefitting from RF advances.

One other problem lies in the fact that there are multiple communications standards. Because different standards have different specifications, it becomes difficult to design a single transmitter which will be efficient for multiple standards. Many IC manufacturers have chips which are optimized for one particular standard, but it is important to have a single architecture which will handle any standard well. Returning to the cell phone example, when cell phones use multiple standards, they often have a separate chip for each. Creating a single chip which can handle any current standard efficiently, and be adapted to new standards is thus very important. The goal of our project is to design and implement a new RF transmitter architecture which will be both efficient, as well as linear enough to support any existing communications standard. In this introduction chapter, I will review

some basic communication theory, give an overview of the current transmitter topologies in use, and provide the foundations to understand the operation of the system presented in this work.

1.1.1 Applications of mm-Wave Technology

Working in the mm-wave band of frequencies (30-300 GHz) offers some distinct advantages over working with lower frequencies. This section serves to highlight some of the applications and technology mm-wave work is being focused on. It has already been mentioned that higher bandwidth systems are available at mm-wave frequencies, simply because the center frequency is so large. Besides this, mm-wave systems require a much smaller antenna size, resulting in a more lightweight and portable system. Thus, mm-wave technology is useful for applications which require the wireless transmission of large amounts of data. Besides this, mm-wave signals can provide a very high resolution image for visual scanning due to small wavelength. This kind of spatial resolution can outperform an optical visual system under certain conditions (such as fog). This is useful in multiple applications, such as automotive sensing. Sensors on automobiles allow for the car to aid the driver by providing full 360° awareness. This kind of technology is also used in safety inspections, allowing for the visual search of concealed weapons. One last use for mm-wave technology lies in medical imaging. Some specific applications include oxygen content measurement in blood (oxygen attenuated signals at 60 GHz), or tumor detection. These applications, and more, are discussed more in-depth in [2]. Clearly, mm-wave technology is finding more and more uses in cutting edge applications.

1.1.2 Communications Standards

To approach the theoretical channel capacity for a system, modulation schemes are used to encode the baseband data. For example, imagine a system where one of four symbols is transmitted at a time, as opposed to a system where one of sixty-four symbols is transmitted at a time. Clearly, the system which has more possible symbols for transmission has a higher data content per symbol transmitted, so is more desirable for our purposes. The

idea of using more symbols is beneficial; however the difficulty comes in with the SNR of the system. Extrapolating this example out further, at some point it will be impossible for the receiver to distinguish between symbols sent because they will be so closely clustered together that noise will prevent the receiver from determining precisely which symbol was sent. The example presented in Figure 1-1 is representative of a form of modulation called N-QAM modulation, where N is the number of potential symbols for the transmitter. An example 64-QAM modulation cluster can be found in Figure 1-1 below. Each of the dots represents a different symbol being transmitted. To determine the symbol being transmitted, a vector is drawn from the origin to the desired symbol. Projecting this vector onto the I and Q axes results in the necessary magnitudes for the in-phase and quadrature components of the vector, respectively. This gives us some insight into how transmitters work. By summing an in-phase and quadrature component, two independent power amplifiers can determine an output symbol. More about this will be explained at later points in this work. More detail on the construction of an N-QAM constellation scheme can be found in [3].

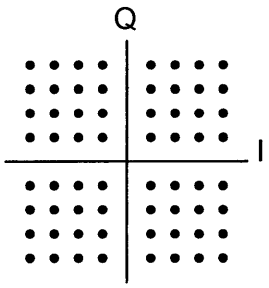


Figure 1-1: 64-QAM Constellation diagram

For the purposes of this work, we will assume a 64-QAM modulation scheme is used. Each communication standard has a unique probability density function (PDF) governing the probability of sending each symbol. Often, chips are designed to be efficient with respect to a particular standard, such as WLAN or HSUPA, but it is difficult to optimize for different standards. Naturally, a system which is more efficient around the more commonly transmitted symbols is likely be more efficient overall. We would like our system to be efficient regardless of the PDF's of the symbols for the communication standard. One way to achieve this goal is to maintain a high efficiency over all symbols. This case is ideal,

but virtually impossible to achieve in practice. An alternative is to provide a method in which the efficiency can be adjusted over certain ranges of symbols to ensure the system is adaptable to multiple standards, and efficient at each of them. This is one of the strengths of the system we propose, and will be discussed in depth later in this work.

1.1.3 Transmitter Architectures

In this section, I will provide a brief overview on two of the more common RF transmitter architectures, the polar architecture and the LINC (linear amplification with nonlinear components) architecture. Communication standards which allow for high data rate typically use a variable-envelope modulation, which implies that the power amplifier (PA) in the transmitter needs to be linear. Linear PAs are typically inefficient, so we'd like to use switching PAs instead. Switching PAs introduce a large amount of nonlinearity into the system, but afford more efficiency. Both the polar and LINC architectures allow for the use of switching PAs while producing a linear output. However, each has a set of drawbacks which make them undesirable for our project.

Polar Architecture

A simple block diagram of a polar transmitter is shown in Figure 1-2(a) below. Under this architecture, the incoming signal is split into polar components (with reference to the QAM modulation diagram), phase and amplitude. The amplitude is used to control the supply to the PA, while the phase acts as the input to the PA. The critical piece here is that the supply to the PA can take on a continuous set of values. This implies that the amplitude stage itself requires a linear power converter to drive the supply. The output of the PA is thus a varying-envelope system modulated by the amplitude path and driven by the phase signal. This system is difficult to implement for high-bandwidth systems due to the fact that the amplitude control path includes the power converter. While this topology isn't directly useful for our needs, it does demonstrate that a system with a modulated supply can accomplish the task, if there is a way of shifting the supply efficiently.

Outphasing Architectures

A different topology was presented in [4] and represents an outphasing technique for RF transmission. This architecture allows for the linear amplification of a signal using non-linear components, and is referred to as a LINC architecture. In this system, the incoming signal is split into two outphased vectors of constant amplitude which can be applied to separate PAs and combined through a power combiner. A Wilkinson passive combiner is a typical example of a block which can fulfill this function. Wilkinson combiners are loss-less when two in-phase signals of equal amplitude are applied, and any loss resulting from phase offsets appears across a resistor. A block diagram of a generic LINC system can be seen in Figure 1-2(b) below. Unlike the polar architecture, the amplitude to the PA is held constant. This is acceptable for large signals, but inefficient for small signals. To obtain a small vector output from the system, the two summing vectors end up being outphased from the original vector angle by a large factor. When these two vectors are combined, most of the output power from the PAs gets wasted through the combiner resistive isolation port, making the system inefficient for small outputs. This system doesn't require the linear power converter of the polar system, so it can reach a much higher bandwidth. One modification on this system involves using a multi-level LINC system, in which the supplies to the PAs can change, but only together so that each vector maintains the same amplitude. This system results in efficiency "peaks" at each power supply level, which results in a higher average efficiency [5]. Another similar method using a class-E power amplifier and transmission-line combiner was presented in [6]. Such a technique would offer a unique combination method which would potentially allow using different points on the transmission line to produce correctly combined voltages for an output, but would offer difficulties as the load-sensitive power amplifiers would see a varying combiner load. Regardless, the strength of this architecture lies in the fact that it can be very fast, if inefficient for some output symbols. It appears that the polar architecture and the LINC systems have orthogonal benefits, which we made extensive use of in our system.

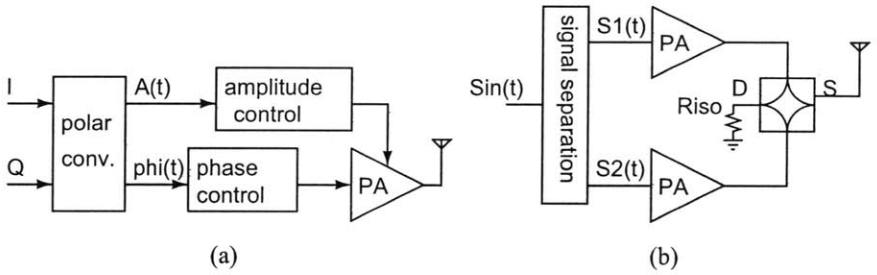


Figure 1-2: Transmitter architectures for (a) polar and (b) LINC systems

1.2 Past Work on AMO

This section serves to introduce the Asymmetric Multilevel Outphasing (AMO) architecture developed in the Dawson laboratory group. This architecture was presented in [7] and a first generation chip was designed, simulated, and fabricated. This chip is currently undergoing testing. A simplified block diagram representing the system architecture can be found below in Figure 1-3.

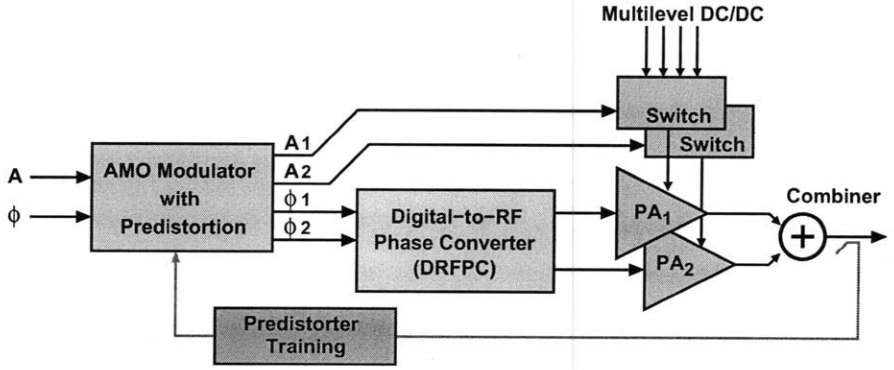
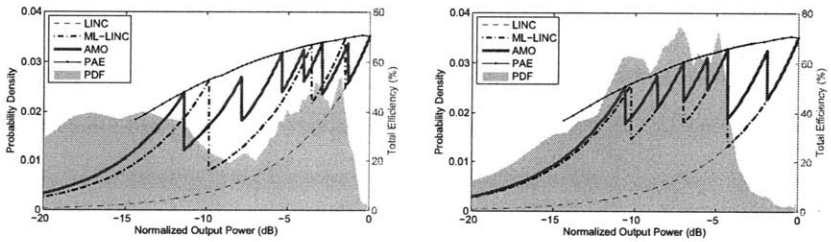


Figure 1-3: Simplified AMO architecture

The AMO architecture is very much a hybrid between the polar and LINC architectures, taking the advantages of each while eliminating their flaws. Using this AMO architecture, we believe that we can eliminate the efficiency versus linearity constraint which has up until now appeared to be fundamental to RF transmitter designs. In a similar manner to the multilevel LINC architecture, the AMO system decomposes the input vector into two outphased vectors and adjusts the power supplies in a manner corresponding to the amplitude of the

input signal. However, unlike the multi-level LINC system, the AMO system allows the PAs to have asymmetric power supplies. This allows the decomposed vectors to have a smaller outphasing angle than would be allowed under the multi-level LINC architecture, which results in less outphasing loss in the power combiner. By adding multiple discrete PA supply choices, the system efficiency diagram results in multiple "peaks" at different output powers with each peak corresponding to the case in which the output amplitude from the combiner is the same as the PA supply amplitude (a theoretical 100% efficiency, disregarding losses in the PA itself). This is equivalent to saying that the outphasing angle is zero at this point. With n switchable supply levels, there exists a total of $\frac{(n)(n-1)}{2}$ of these peaks. However, the control for this system would be difficult with so many choices, and simulations predict that combiner efficiency decreases as non-consecutive voltage levels are used. Thus, a decision was made to limit PA supplies to using consecutive voltage levels, resulting in a total of $(n) + (n-1)$ efficiency peaks, and a much simpler control scheme. In the AMO device already simulated and built, the designers chose to use 4 power supply levels, for a total of 7 efficiency peaks in the system. A plot of simulated efficiencies versus output power can be found in Figure 1-4 below. Note that the locations of the efficiency peaks are variable between the two plots. The gray plot represents the output power PDF for each of the communications standards (using 64-QAM). By adjusting the supply voltages, efficiency peaks can be chosen to optimize overall system efficiency for a given communication standard.



(a) (a) HSUPA standard

(b) (b) WLAN standard

Figure 1-4: Efficiency Plots

The AMO system also uses a digital predistortion lookup table (LUT) to adaptively calibrate itself. When a signal comes in, a pair of output vectors needs to be produced. The

LUT performs this action, producing vectors which will produce the correct output at the output of the highly nonlinear PA system. This predistorter corrects for the static errors (offsets, analog mismatches, nonlinearities) in the AMO architecture. As these errors drift over time, the LUT is periodically run to keep the system calibrated properly. Some of the predistortion techniques used are presented in [8], but this topic will not be the focus of my thesis.

Another avenue of research being implemented in this phase of the program is to use an energy recovery scheme on the output power combiner. In a normal Wilkinson combiner, any outphased power is wasted in an isolation resistor. If we were to couple a rectification system onto this port, energy could be recovered and fed back into the supply, resulting in a greater system efficiency as we are no longer wasting all of the outphasing power. Such a system was demonstrated in [9], and a member of our team is working to implement this in our current design.

1.2.1 Modern AMO

The latest version of the AMO follows the same general topology as the last chip. One difficult specification is that our system needs to be able to output a total of 4W (36 dBm) of output power from the summed PAs. This amount is not achievable from a single set of PA cells. One difference then, is the inclusion of 16 PA cells on this chip. This system is idealized in Figure 1-5 below. We group PAs together in sets of four, using a total of two DRFPC (digital to RF phase converter) modules to drive them with a common signal. We are not constrained to using the same voltage supply on each of the four amplitude paths, however, so we are granted an additional degree of freedom in determining what supply voltage levels to use. This degree of freedom results in more peaks in the efficiency curve, resulting in a higher efficiency overall. This implementation can easily be expanded by using more PA cells in parallel with the cells already existing, and using some variety of power combining structure (tree or otherwise) to combine the outputs.

Two potential chip topologies were considered for this project. In the first, the PAs and switching network are done on one chip, while the DRFPC and control are done on another.

This allows the DRFPC to use a faster technology node, while allowing the PA to utilize a special process to allow for better efficiency. The difficulty with this approach is that the two chips need to communicate at 45 GHz. To do this, each chip has to effectively build a power amplifier to drive a 50Ω load. This method is inefficient, and while it allows for higher block efficiencies, the interconnects suffer. The other alternative is to use a single process, and put all blocks on one chip. This method reduces the importance of interconnects, but degrades the performance of individual blocks. This thesis explores both methods, by using two different processes. In the first process, I explore switching and time alignment in a 130-nm SiGe process, which we intend for PA use only.

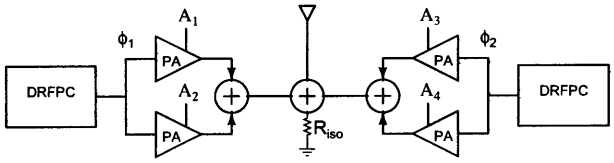


Figure 1-5: AMO architecture with power combining

1.3 Project Overview

As presented in the previous section, the AMO architecture will allow us to improve system efficiency while maintaining output linearity as we increase the overall bandwidth of the system. Our proposed program is to increase the carrier frequency of the signal into mm-wave frequencies. This particular project will focus on creating a system with a symbol rate of 2 GSPS (base bandwidth of 2 GHz), on a carrier frequency of 45 GHz. Operating at these frequencies imposes severe constraints on the system.

Our system proposes to have an output power of 36 dBm (4W) at this frequency, with a peak efficiency of 67%, and an average efficiency of 65%. The output power constraint implies that the sum power of the PA systems need to be 4W. To obtain 36 dBm of power at 45 GHz, multiple PA cells are required. Fundamentally, instead of dividing our input signal into two phase signals and two amplitude signals, we divide the input signal into two phase signals and four amplitude signals. Each of these phase and amplitude paths controls

a single set of PA cells, which sum with one another in a tree structure of power combiners. This opens new opportunities, such as using different supply levels on the different PAs in the same phase path to optimize even further, but also requires multiple levels of power combining.

The bulk of the work presented in this paper deals with the design of the power switching network to change supply levels for the PAs. This work was done for a 130-nm BiCMOS SiGe HBT process. These switching networks will be discussed in great depth in Chapter 2. Chapter 3 will present my design for a 45 GHz voltage-controlled oscillator (VCO) in both 130-nm BiCMOS and 65-nm CMOS processes. Chapter 4 will present the techniques and strategies used in laying out these structures in their respective technologies, and will conclude with some results and analysis.

Chapter 2

Power Switching Network Design

2.1 Introduction

As discussed previously, the proposed AMO architecture makes use of switching between several different voltage supply levels to provide power backoff, which in turn increases system efficiency. A diagram indicating system functionality can be found in Figure 2-1 below. Here, two separate amplitude paths are controlled by sets of four power switches, allowing them to change levels independently. In this work, the PA will consist of multiple PA cells, grouped into four independent amplitude paths. Having four amplitude paths while maintaining only two phase paths provides an extra degree of freedom in system optimization. In this chapter, I will highlight several of the issues which emerged when designing these switches. In the first section, I will give some background and core design specifications I had to work around. In the second section, I will describe how these switches were designed and built in a 130-nm Silicon-Germanium (SiGe) BiCMOS process. The third and final section will propose the topology of a feedback network which could be used to keep the switches time-aligned.

2.1.1 Switch Design Specifications

This section serves to highlight some of the design considerations for this power switching network. Primarily, the switches need to operate at a baseband frequency of 2 GHz. In

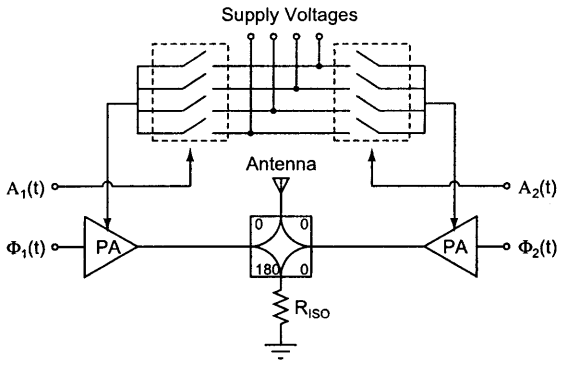


Figure 2-1: Conceptual model of PA supply switching

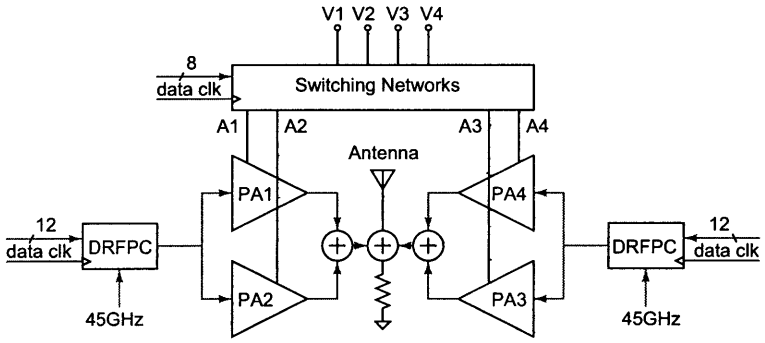


Figure 2-2: AMO model with four amplitude paths

the case of no oversampling, switching at the symbol rate is required by the outphasing nature of the transmitter. Imagine the case where a symbol of large amplitude is followed by a symbol of smaller amplitude. If the power switch doesn't change levels between these symbols, the transmitter will be more efficient during the first symbol than the second, but will still work properly. Now imagine the case where a symbol with smaller amplitude is transmitted first. If the switching network provides a supply level which is too low for the second symbol, then the output will not be a correct value, and we have corrupted our output. This can be bypassed by examining the data ahead, and choosing the supply needed for the largest symbol. A full optimization along these lines is examined in Appendix A. While switching at the symbol rate provides a minimally functional system, it helps to oversample the symbols by some oversampling rate. This allows us to shape the signal spectrum. Unfortunately, the 130-nm process is simply too slow to support oversampling

for a symbol rate of 2 Giga-Symbols per second, but could support oversampling if the symbol rate were reduced. Furthermore, smaller process nodes should yield better results.

A related consideration is the rise time for the switching network. Clearly, if the system takes longer than 500 ps to switch from one level to another, the system is not going to be able to meet the bandwidth specification. One method to determine the maximum allowed rise time is to model the ideal system with a low-pass filter in the amplitude path. By changing the location of the pole in the filter, we are able to approximate the effects of rise time on the spectrum of the system.

These plots were generated by using an idealized model of the AMO system, where the only nonideality was the rise time of the switches. The input signal consisted of a 2 Giga-Symbols per second (GSPS), 64-QAM modulated signal with 10 bits of dynamic range. The signal was oversampled by a factor of 4, and about 2000 symbols were simulated to generate the spectral masks shown. In Figure 2-3 below, we see the ideal spectrum for our system, with no time mismatches and infinitely small rise times in our amplitude path. As expected, the spectrum is smooth outside of our band of interest, and the output is identically an amplified version of the input.

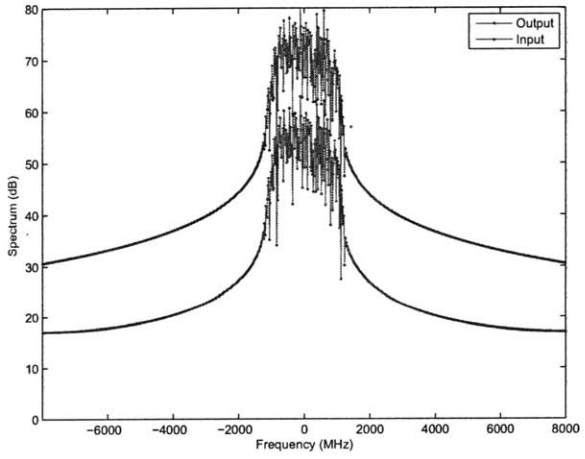
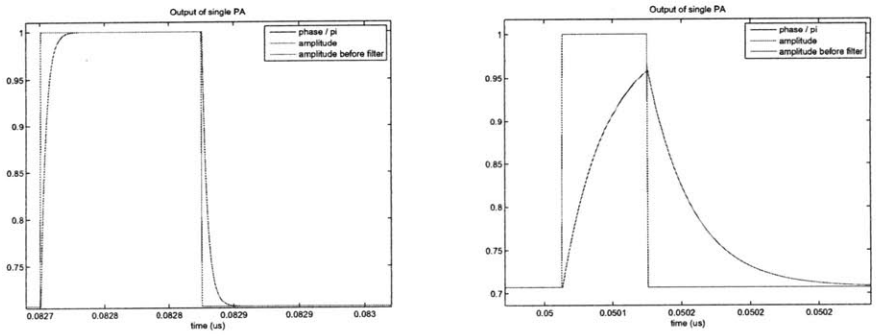


Figure 2-3: Spectral mask for ideal amplitude path switches

By introducing a low pass filter in the system, we can approximate the rise times of our switches. For example, Figure 2-4 below demonstrates the effective result, in the time

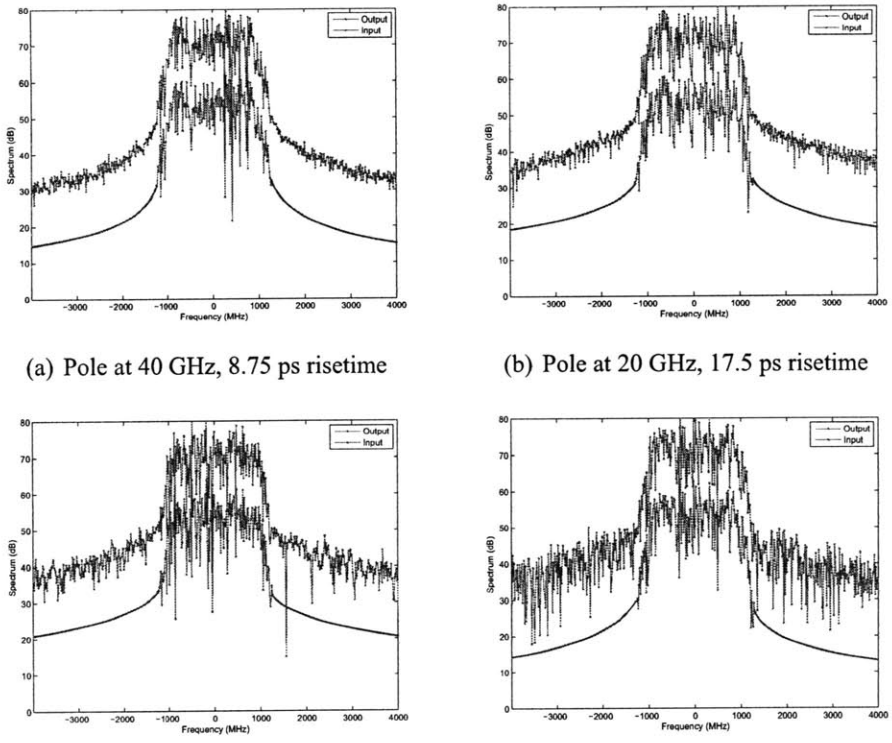
domain, of introducing this low pass filter. As expected, the filter with the lower frequency pole exhibits a longer rise time. Naturally, one might suspect that this longer rise time introduces distortion in the spectrum, which is precisely what occurs. The spectrums for a select group of filters can be seen below in Figure 2-5. In the first three images, the out of band spectrum looks relatively clean. In the fourth image, we can see that the SFDR of the system has been reduced by the distortion in the spectrum (as a result of the upwards spike in the spectrum). This result indicates that for an oversampling factor of 4x, a 70 ps switch rise time is effectively the longest we can allow before our system performance degrades. However, when we change the oversampling rate to a factor of 2x, we find that the risetime becomes much less important in the system architecture. Examining the spectrums presented in Figure 2-6 below, we note there is very little spectral difference between a risetime of 35 ps and a risetime of 70 ps. However, when compared to a 350 ps risetime, we see that there exists a few dB of difference between the two plots. This indicates that the same range of risetimes has less impact on the spectrum of the system when we oversample less. Note that the spectrums with a factor of 2 oversampling have a worse spectrum than the spectrums with a factor of 4 oversampling, it turns out that the oversampling rate plays a more important role than switch risetime here. Given that the system cannot receive signals at a fast enough rate to oversample in the 130-nm process, a target risetime of 100 ps should not affect the spectrum. These simulations and results are a result of work done in [10].



(a) Pole at 40 GHz

(b) Pole at 5 GHz

Figure 2-4: Amplitude switch risetimes



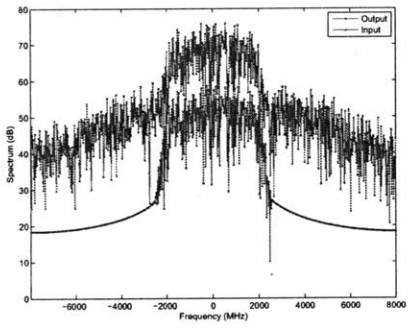
(a) Pole at 40 GHz, 8.75 ps risetime

(b) Pole at 20 GHz, 17.5 ps risetime

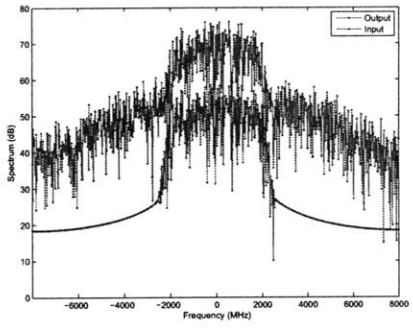
(c) Pole at 10 GHz, 35 ps risetime

(d) Pole at 5 GHz, 70 ps risetime

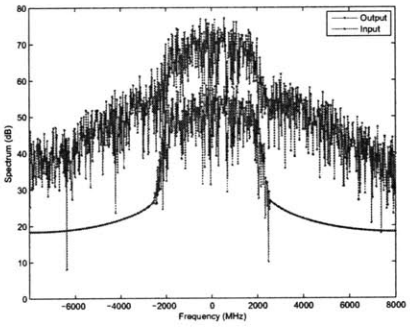
Figure 2-5: Spectral purity due to amplitude switch risetimes, 4x oversampling



(a) Pole at 10 GHz, 35 ps risetime



(b) Pole at 5 GHz, 70 ps risetime



(c) Pole at 1 GHz, 350 ps risetime

Figure 2-6: Spectral purity due to amplitude switch risetimes, 2x oversampling

While discussing switching speed and rise time, we have not yet taken into account the delay in the system. There exists some delay in the system from the time the amplitude signal is put onto the chip to the time the power levels shift. It turns out that this time delay is not important to the system. As long as the delay is relatively well known (and constant), we can put the amplitude signal into the system with that time offset so that the system switches at the appropriate moment. On the other hand, variations between switches related to the mismatches in layout routing and device mismatches will cause efficiency problems, but the feedback system I propose in the last section of this work servos out this error.

The other specifications involved in the design of these power devices are area and power cost. As we make the devices larger, they require more area, and cost more power to switch. These are terms which we would like to minimize to optimize the system.

2.2 Switching in 130-nm SiGe BiCMOS

In this section, I will discuss the choices and tradeoffs I made in designing my switching network. Before I begin however, I would like to take a look at the switching network topology implemented in the previous version of the AMO chip. The last chip implemented a network of the form found in Figure 2-1. A close-up of one of the switches is found in Figure 2-7 below. As the figure demonstrates, the switch is simply a transmission gate between the external supply and the PA supply nodes. The drivers for this circuit are a simple chain of geometrically tapered inverters.

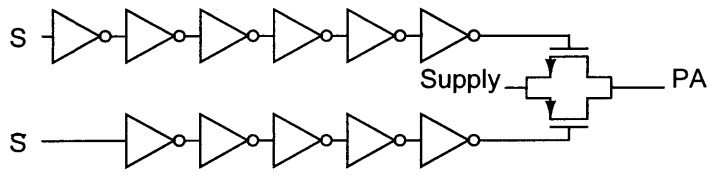


Figure 2-7: Power switch for first AMO

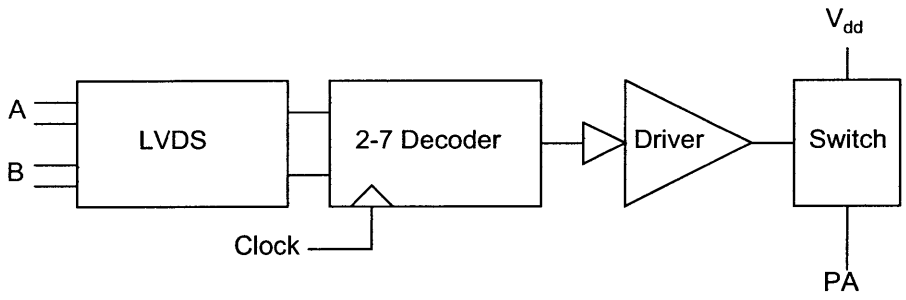


Figure 2-8: Signal chain for power supply switch network

2.2.1 Topology Decision

The signal path for my power switching network is presented in Figure 2-8 above. The differential signal comes onto the chip via a LVDS receiver, which converts it to a single-ended signal. The two amplitude control bits pass through a decoding block, which generates the control signals for the driver blocks. The driver blocks drive the power switches, which in turn modulate the supply voltage for the PA. The remainder of this section will focus on the design decisions involved in each of the above blocks. I will first discuss the power switches themselves, then follow with an analysis of the different driver topologies. I will finish with a brief description of the decoder block. The LVDS receivers were designed by another member of the team.

The first factor in designing my switches was determining which devices to use. Available to me were NPN HBT devices, thick-gate oxide FET's (n-type and p-type), and thin-gate oxide FET's (also n-type and p-type). These devices each have benefits and drawbacks. For example, the NPN devices have very high current gain, making them good current-handling devices. BJT devices however, tend to perform less well as switches than FET devices mainly due to their saturation region. When a BJT enters the saturation region, it ends up building up excess charge which needs to be discharged before the device switches again. As a result, unless the device is kept out of saturation, it has difficulty acting as a switch. Furthermore, when the BJT is kept out of saturation, power is lost across the collector-emitter junction. A similar tradeoff exists between the available FET devices, thick-gate oxide devices can withstand a much higher voltage rail than the thin-gate oxide

devices, at the cost of a lower transconductance.

The key factors for designing a power switch are power loss and current handling capability. A switch needs to be efficient, but also needs to be able to pass the proper amount of current. In other words, the device needs to be wide enough to handle the proper amount of current, but needs to be small enough so that switching it is cost-effective from a power perspective. From these perspectives, a simple n-type FET would provide the best results, as it has a higher electron mobility than a p-type FET, and thus current, for the same drive voltage. This allows us to reduce the size of the switch, resulting in less power lost due to switching the device. However, the n-type FET switch fails for high supply voltage levels, as we cannot overdrive the switch hard enough to pass current without entering breakdown regions for the device. A p-type device however, can operate at any of the supply voltages available on chip. This effectively leaves two switch topology options open to us, which are a purely p-type FET approach, or a transmission gate approach. These will be analyzed further in the next section.

2.2.2 Minimizing Power Loss

As previously stated, one of the most important system parameters is efficiency. To make the system as efficient as possible, it is crucial to minimize losses in the power devices. The power lost in the amplitude path can be modeled with the following equation:

$$P_{lost} = \frac{1}{2} C_{total} V_{dd}^2 f + V_{ds} I_{switch} \tag{2.1}$$

Equation 2.1 was found by realizing that the main sources of power loss in the amplitude path correspond to the losses resulting from switching the devices themselves, as well as the resistive losses from the switch channel. Thus, in this equation, the variable C_{total} corresponds to the total capacitance switched for each symbol. This ends up being the capacitance looking into the gate of the switch itself, as well as the total capacitance from the driver circuitry. V_{dd} corresponds to the maximum supply level in the system, and f represents the sampling frequency at which these devices are switched. The second term effectively represents ohmic loss due to the channel resistance, with V_{ds} representing the

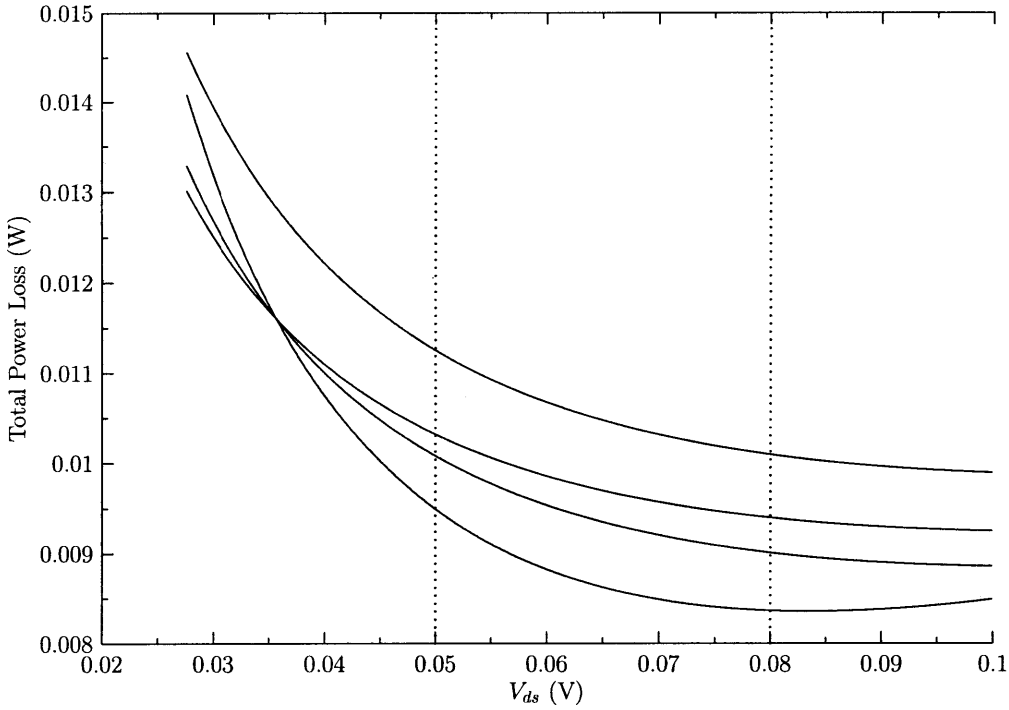
drain to source voltage on the switching device, and I_{switch} representing the current drawn by the PA.

Equation 2.1 indicates a clear optimum for this system. As the switching device grows wider, so does the total capacitance of the system. However, the increased width of the device lowers the channel resistance, resulting in a lowered ohmic loss. Plots of the power loss for the thick-gate p-type FET switch and the transmission gate switch can be found in Figure 2-9 below. Each family of plots represents varying the supply voltage at the top of the switch, as the absolute efficiency of the power devices depend on this parameter. These curves are plotted against V_{ds} to determine the ideal operation regime for the switches. This value can then be used to determine the total width required for the switch devices. As we see in Figure 2-9 the switching losses roll off quickly, while the ohmic losses ramp up slowly. Thus, we are left with a nice wide operating region around 75-100 mV V_{ds}, where we have a worst case power loss of about 20 mW per PA for the transmission gate switch, and a loss of about 15 mW per PA for the thick-gate p-type FET switch. It is important to note that this power loss is fundamental, we cannot escape the laws of physics unless we switch to a process with lower parasitic capacitance. One optimization that can be changed, however, lies in reducing the switching frequency. A complete derivation can be found in Appendix A.

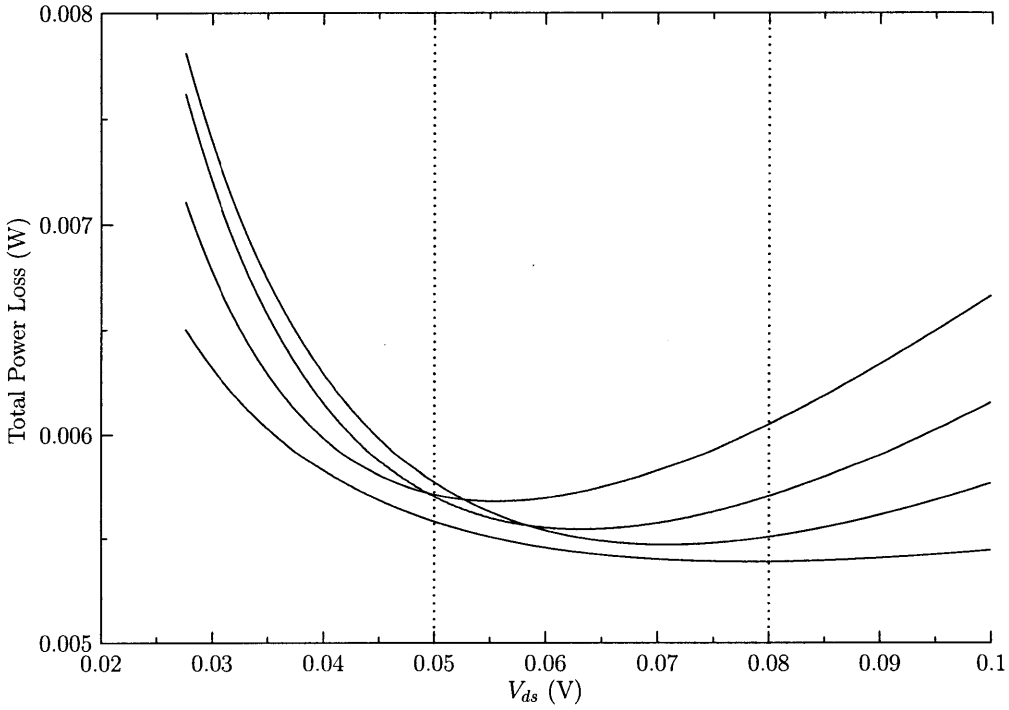
If we assume the PA is ideal, we can model the efficiency of the switches by tracking the power output of the system and power input to the system. The results of this for each type of device are found in Figure 2-10 below. As is demonstrated in the plots, the peak efficiency reaches about 90% for each switch topology, however the thick-gate p-type FET device is much less efficient at low supply voltages. At low supply voltages, the device operates at a lower current density, requiring a larger gate width to pass the requisite current. These efficiency numbers are solely indicative of losses in the switches themselves, they have no reference with the efficiency of a realistic PA. Using a realistic PA would shift the absolute position of the graph, but will leave the relationships between the curves alone.

Once we have determined the ideal operating regime of our switch devices, we need to determine how large they should be to fit within that regime. I used the worst case

(highest) curve of the family plotted in Figure 2-11 below. In these plots, the family of curves represents a parametric sweep of the voltage at top of the switch. Using a similar method, we can get an idea for the total area required for each switch. As is demonstrated in these plots, the total area required for the transmission gate switch is smaller than that of the thick-gate p-type FET one, again due to current density capabilities. All of these factors combined demonstrate that the transmission gate switch is more efficient than a thick-gate p-type FET switch, and is smaller as well. These results informed my decision to make use of a full transmission gate switch on the bottom three switch levels, and a sole p-type FET switch on the top voltage level. It should be noted that these plots demonstrate values for the supply voltages actually settled upon. Under the plots shown, a thick-gate PFET appears to be more efficient than a transmission gate. However, at the time this decision was made, a much lower minimum supply voltage was assumed; this lower supply voltage gave the transmission gate switch an advantage.

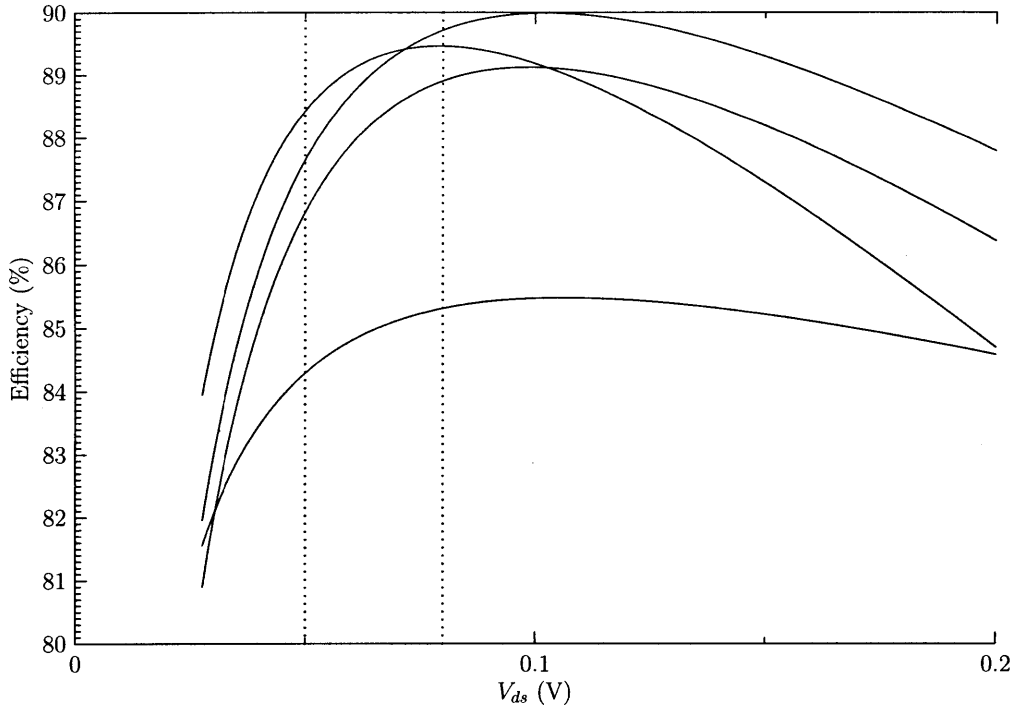


(a) Transmission gate

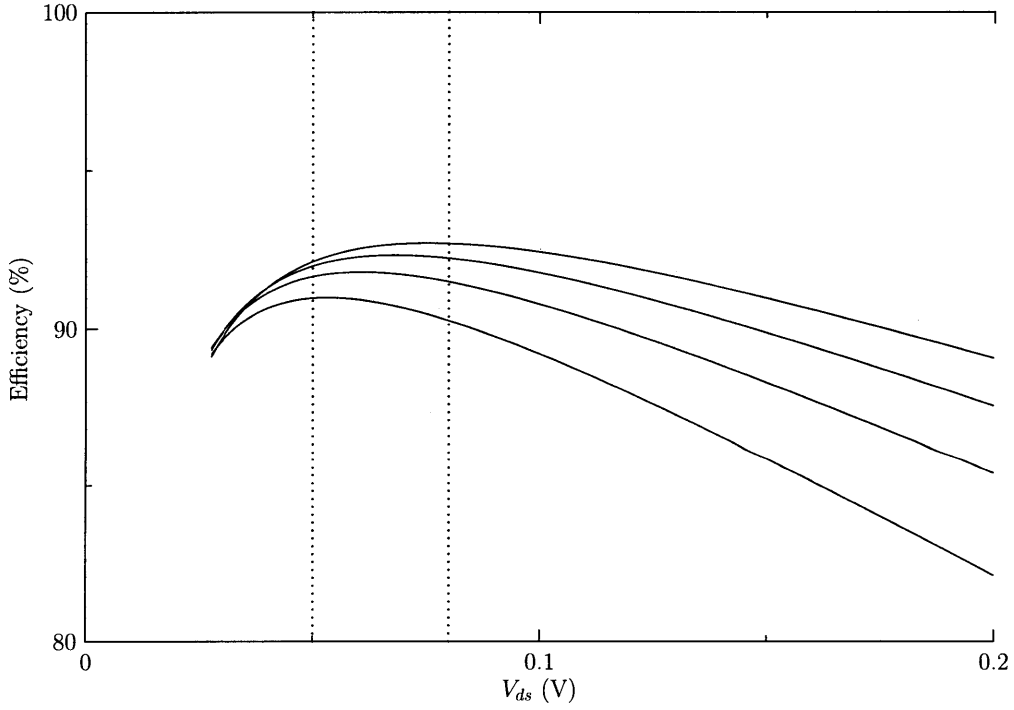


(b) Thick-gate PFET

Figure 2-9: Amplitude switch power loss vs. operating regime

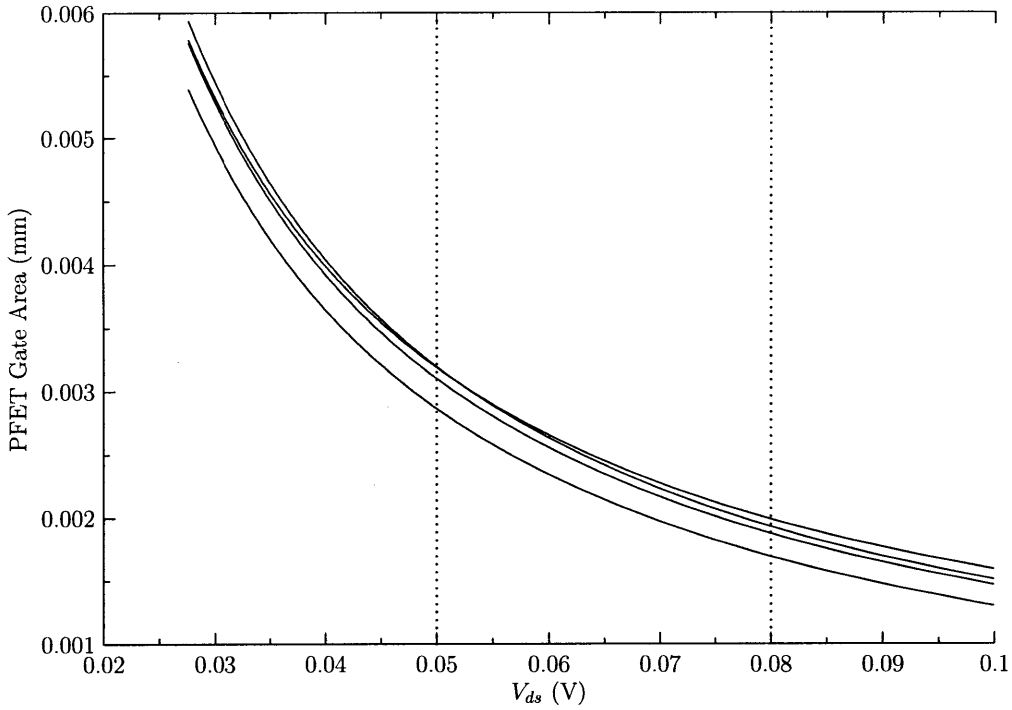


(a) Transmission gate

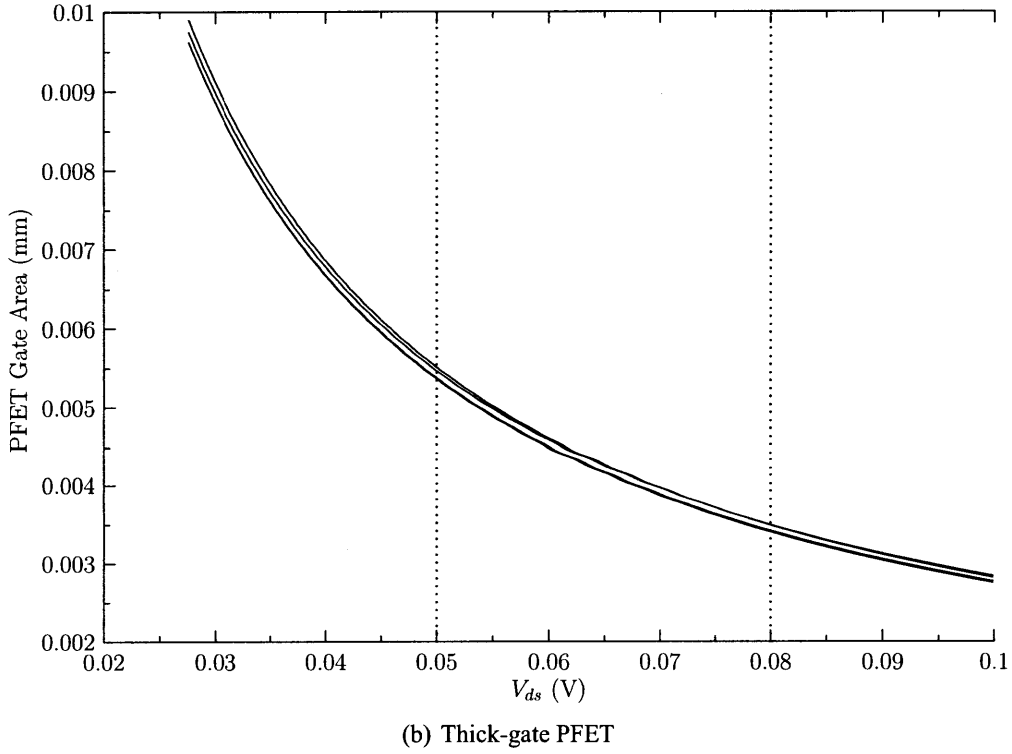


(b) Thick-gate PFET

Figure 2-10: Efficiency plots vs. operating regime



(a) Transmission gate



(b) Thick-gate PFET

Figure 2-11: Total PFET gate width vs. operating regime

2.2.3 Driver Topologies

Once the power switches were designed, a driver needed to be built to take in a minimum-length device signal, and translate it into a driving signal. As already mentioned, the driver itself does not need to be fast with regards to propagation time, only with regards to how fast it can activate the switch. If we model the power device as the capacitance resulting from the gate to source and gate to drain capacitances of the device, then the effective rise time of the power switch input is proportional to how much current the driver can provide to charge up that capacitor.

A few driver topologies were considered for this work. Once more, any bipolar options were eliminated due to difficulties surrounding the saturation of the device. Given that the driver needs to always be at one of the power rails, other options such as a CMOS diamond buffer were eliminated due to biasing constraints near the rails. Two topologies then presented themselves for further inspection, a regenerative feedback network, and an inverter chain. Regenerative drivers, in the form of comparators with positive feedback as presented in Figure 2-12, use cross-coupling to cause the comparator to enter positive feedback. This regenerative effect potentially results in a faster switching speed. The second driver topology uses a geometrically scaled inverter chain to drive the power switch, such as the one demonstrated in Figure 2-7. The driver is geometrically scaled to minimize the propagation time through the driver.

Effectively, what determines the speed of a particular driver is the current slewing onto a node, and the capacitance which needs to be driven. Examining the cross-coupled driver, we see that for each output node, the capacitance on the node is $C_{gs,n}+C_{gs,p}+C_{db,n}+C_{db,p}$. The $C_{gs,n}$ comes from the input capacitance of the next stage, the other three capacitances come from the current stage. For identically sized transistors, the capacitance on the output node of an inverter can be represented as $C_{gs,n}+C_{gs,p}+C_{db,n}+C_{db,p}+[C_{gd,n}+C_{gd,p}]||[C_{db_n}+C_{db_p}]$. In this equation, the C_{gs} terms and the parallel term are derived from the input capacitance of the next stage, the bulk capacitances from the current stage.

Given identically sized transistors, I would expect the cross-coupled driver to be faster, due to smaller capacitance on the node. In the inverter, the C_{gs} terms are for the next

stage, which are bigger than the current stage. In the cross-coupled driver, only one of the C_{gs} terms is from the next stage, so there would likely be a smaller total capacitance on the node. Note that this is a qualitative argument, it is distinctly possible that the cross-coupled stage would have to have larger n-type FET devices. In addition, the regenerative driver stages are fully differential. Given that not all of my power switches require a fully differential driving signal, this results in a waste of area and power for those switches. One additional benefit of the cross-coupled drivers however, is that the output signals are necessarily well aligned. This is in contrast to two independent inverter chains, which are not necessarily time-aligned. Despite the potential advantages of the cross-coupled driver, I chose to use an inverter chain for a few reasons. First, the inverter is a simpler circuit. This proves to be a boon while performing layout, especially given the pitch constraints presented in Chapter 4. Second, the time-alignment advantage is eliminated when using the time-alignment feedback network presented later this chapter. Third, when comparing the capacitance load differences, the switch speed is determined by the speed of the last driver stage. In the last driver stage, the capacitance is dominated by the input capacitance of the switch, so topology differences will result in a potentially different signal propagation time, but not a significantly different switch speed.

Using the inverter topology, the amount of power wasted during each switch is equivalent to $\frac{1}{2}C_{total}V^2$ where C_{total} represents the total capacitance being switched, or the sum of the gate capacitances. Given that these capacitances are scaled geometrically with factor $\frac{1}{\alpha}$, last stage capacitance C and n stages,

$$C_{total} = \sum_{i=0}^{n} C\frac{1}{\alpha}^{i} \cong \frac{C}{1-\frac{1}{\alpha}} \tag{2.2}$$

Larger values of C_{total} result in more power being burned, while a smaller value of α results in a faster switch time (to a point). Thus, there exists a direct tradeoff between power lost in the driver circuit, and transition time of the power device. The design strategy chosen here was to design the driver to meet the upper limit for the transition time, this allows us to still meet spectrum requirements while minimizing power loss in the driver. These results are displayed below in Figure 2-13 and Figure 2-14. For Figure 2-13, I measured the

10%-90% risetimes between different consecutive levels. For Figure 2-14, I simulated the average power lost by a single driver switching at 1 GHz (the maximum switching speed of these drivers). I chose a fanout factor of four for my driver, resulting in both a decent 10%-90% rise time, and a relatively small power loss.

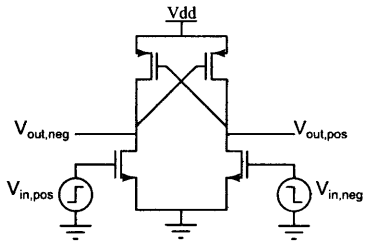


Figure 2-12: Single stage regenerative driver

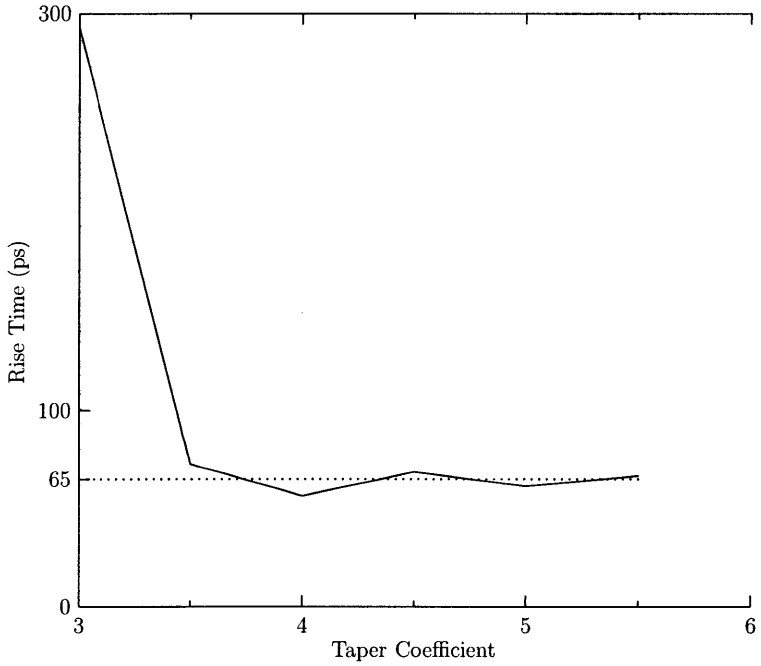


Figure 2-13: 10% - 90 % Voltage risetimes by scaling area

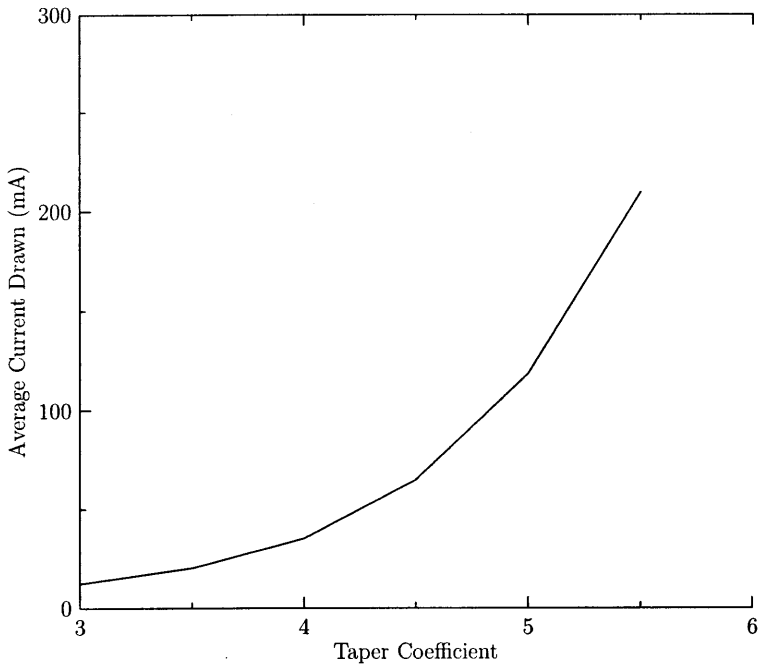


Figure 2-14: Power consumed by scaling area

2.2.4 Decoder Block

The decoder block serves to turn the two control bits into a usable signal for the drivers. Due to an absence of a proper digital library, the design was simple, as presented in Figure 2-15 below. A and B represent the MSB and LSB of the incoming control signal, respectively. The outputs use a D-latch, which both provides a differential signal, as well as allows us to clock the symbols into the system to help reduce timing mismatches. After the D-latch, I placed a basic differential cross-coupled pair of thick-gate FETs. The purpose of this stage was to provide a level shifting of the signal from the thin gate supply to the thick gate supply, ensuring an even turn-on of the driver stages.

2.3 Time-Alignment Feedback Network

To meet the output power requirement, and to allow for outphasing, the AMO chip requires multiple PA cells. These PA cells will differ slightly, through manufacturing mismatches, as well as RC mismatches through path layout differences. As one might imagine, it would

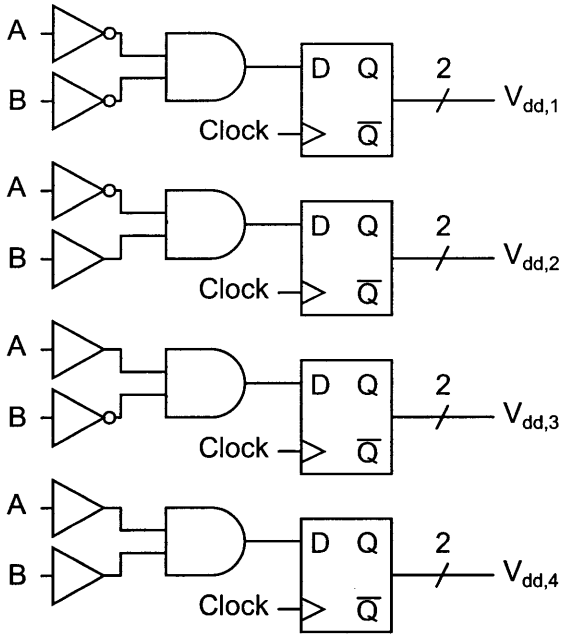


Figure 2-15: 2-8 Digital decoder block

be ideal to have each of the PA cells switch power supplies at the same time. In the event that this doesn't happen, different PA cells will be outputting different symbols, and the combination of these will be a garbled mess. This in turn will increase the noise floor of the system. Given that we cannot control the timing alignments beforehand, it becomes significant to control the timing of the power switches using a feedback network. Analyzing the offsets characteristic of the PA switches, we find that the only offsets present are either manufacturing offsets, or very slowly varying offsets (with temperature, for example). Because of this, we are allowed to have a feedback loop with a low bandwidth, suggesting a periodically calibrated offset. Given that the majority of the offsets are DC offsets, it is possible that a single calibration might suffice for our purposes.

The first AMO attempted a static calibration scheme, but this had some flaws. The previous AMO allowed for a variable RC network using a variable n-well capacitor on the driver. This effectively allowed the user to slow down each driver by adding some additional capacitance to its load. This method works well, but only if the user knows the offsets relative to one another. Unfortunately, this method proved difficult because of

the inability to isolate a single offset from the spectrum performance. The main problem then, lies in the fact that it is difficult to get the information out of the system, and that the user needs to tweak each offset by hand to correct the problem. Ideally, a feedback system would provide both a measurement mechanism, as well as a way to correct the offsets without requiring any user interaction.

A block diagram for a single feedback network is proposed in Figure 2-16 below. Effectively, I utilize the same idea for the feedback actuation mechanism, by adding additional capacitance to a single node, at the penultimate driver stage. The compensation occurs here so that it doesn't add extra capacitance to the final driver stage, which would result in a slower rise time for the switch itself. To measure the error signals, measurements are made off the two driver inverters. The resulting signal is presumably relatively in-phase, so comparing the two signals with a simple XOR gate is not going to work. Effectively, the XOR gate is more sensitive to changes when the signals are in quadrature, simply due to the finite rise and fall times of the device. Thus, a delay is instituted in the reference leg of the feedback network. In Figure 2-16, this is the upper switch, as it is controlled by a dominant capacitor. This results in a fixed propagation delay of our driver network, but this is not important as it is removed statically from the digital side of the chip. More importantly, we now have a reference which we know is slower than any of the other PA switching network paths, so we can align each of these paths to the reference one.

To make a measurement of the switch time offsets, a fixed-duty cycle of 50% is introduced to each measurement switch, and the measurement switches are used pairwise (for 4 PA networks, we look at the reference and one PA cell, then the reference and another PA cell, and so on). The two signals are run through the XOR phase detector, whose output is a set of pulses, whose length are effectively $2t_{offset} + 2t_{delay}$ for each period. These pulses are run through a low-pass filter, and taken off-chip. The averaged value is then compared to a command value, which is equivalent to $2t_{delay}$ as a nominal t_{delay} is known a priori. The error signal then, is simply equivalent to a scale factor times the offset time of the switch. Using our classical feedback analysis, we know that by adding an integrator into this loop, the error signal gets driven to zero over time, implying the offsets are canceled, as long as we have some stable actuation mechanism. The integrated error signal is run

through an ADC, and taken back on chip, where the DAC output is used to set a bank of binary-weighted capacitors, which effectively adds capacitance to the proper node to slow it down appropriately.

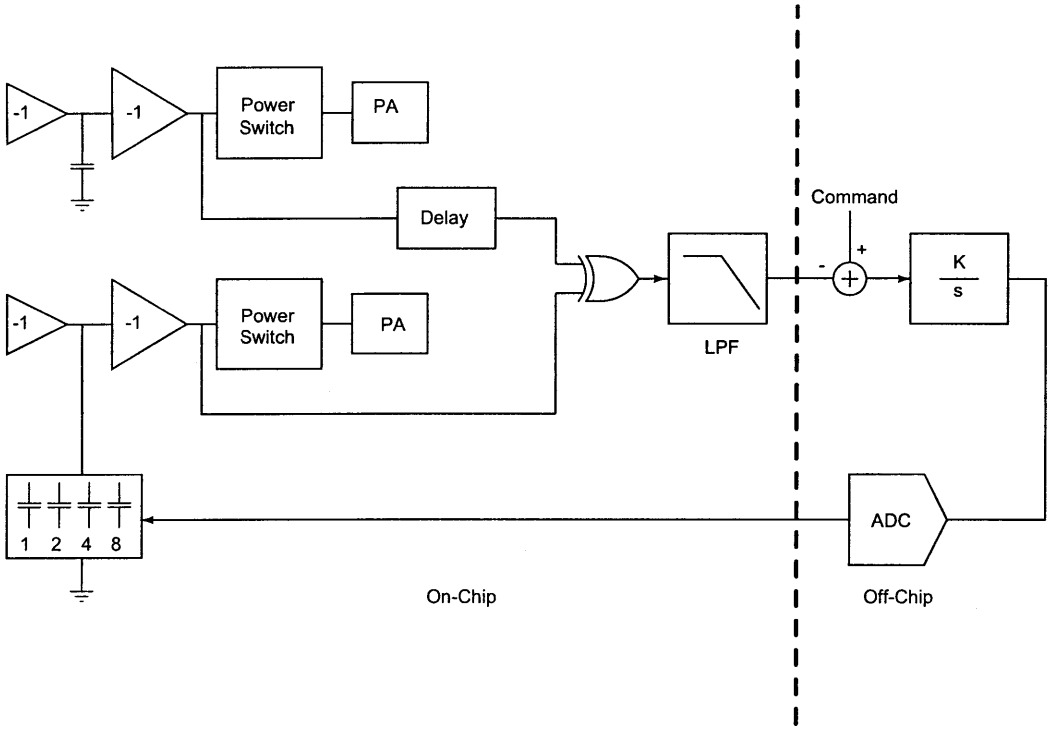


Figure 2-16: Time alignment feedback block diagram

Of course, this feedback network only works if the measurement mechanism has the resolution to measure the error we want to cancel out. Because the measurement block is different between processes, we need to measure the sensitivity of each. Running a Monte Carlo simulation over the process corners in the 130-nm BiCMOS process, we are left with the result in Figure 2-17. We find that the timing mismatch of the measurement system is less than 6 ps, which is a fine enough resolution floor for our purposes. Furthermore, this plot informs us of how much delay we need to add to the dominant leg to ensure that it operates appropriately. This feedback network was not instantiated on chip, primarily due to lack of space in the vertical pitch of the switching cells.

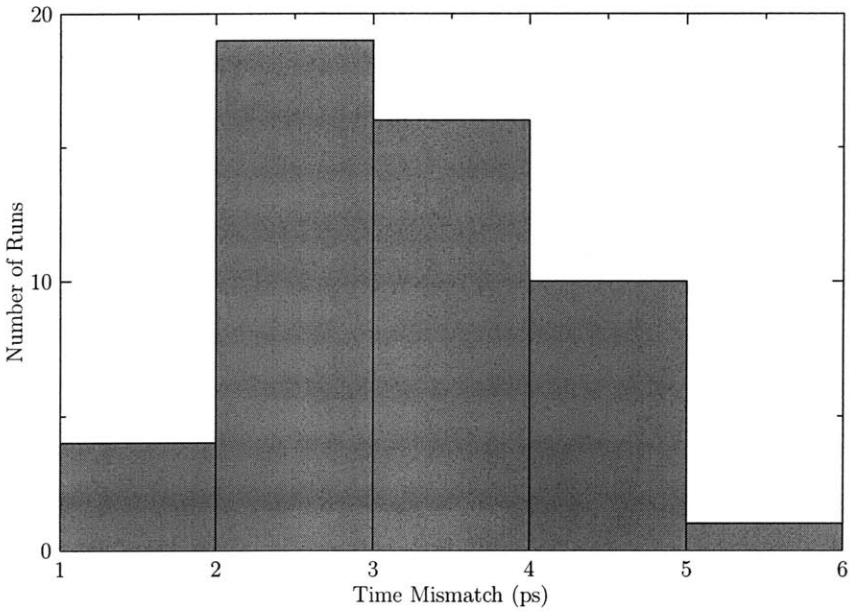


Figure 2-17: Time alignment feedback sensitivity

Chapter 3

VCO Design

3.1 Introduction

The entire system depends on having a tunable system clock which will generate the carrier frequency, and act as inputs to the phase modulator. When designing the chip architecture, we developed two different ideas for clock generation. In the first, the clock is taken on chip directly through a probe station. This requires the clock to be generated off chip, and requires the interconnect to be well modeled. In the second, an on-chip clock generates the control signals. Furthermore, the DRFPC module requires quadrature clock signals, so the single differential clock signal is run through a polyphase filter to yield the required clocks. Voltage controlled oscillator (VCO) designs are presented here for both the 130-nm BiCMOS and 65-nm CMOS processes, with neither actually being fabricated on this chip.

This work did not aim to design a new VCO topology, it simply aimed at building a functional clock. VCO design in this frequency range is relatively well understood, we simply followed the algorithms presented in [11] and [12] in our process. Each of these papers covered both the cross-coupled and differential Colpitts VCO topologies for mm-wave applications.

Before I chose a topology, it was important to determine what my specifications were. I wanted to minimize the phase noise for the system, as the integral of phase noise directly corresponds to the amount of jitter present. Because I wasn't going to have the additional feedback of a PLL to cut down phase noise at low frequencies, I expect the clock to exhibit

slow drifts in frequency, this is acceptable for our current application as we care more about cycle to cycle jitter. Second, we wanted to produce the largest amplitude clock possible, to allow the best driving of the phase modulator switches. This block was intended as a separate module whose power consumption did not add to the power budget of the chip. As such, power consumption was not critical, however minimizing it is good practice.

Two VCO topologies were examined for suitability in the project, a cross-coupled design and a Colpitts design. The Colpitts oscillator initially looked like a better option, primarily due to the fact that it had a relatively high output power (5-6 dBm) by itself. However, 5-6 dBm is simply not enough to drive a 50 Ω load with a decent swing. With this advantage removed, the Colpitts becomes more difficult, primarily due to the large number of inductors required to produce a differential topology. This topology was abandoned due to these difficulties.

My eventual design centered around the cross-coupled VCO shown in Figure 3-1 below. The effective idea here is that the inductor will resonate with whatever capacitances it sees on the output nodes, so we can determine the frequency of oscillation ω_{osc} by:

$$\omega_{osc} = \frac{1}{\sqrt{L(C_{load} + C_{gd} + C_{gs} + C_{var})}} \tag{3.1}$$

Here, C_{load} is the capacitance seen looking into the buffer, C_{var} is a variable capacitance adjusted through the $V_{control}$ handle, and the other capacitances are average parasitics due to the active devices. For a greater tuning range, C_{var} should be the dominant capacitance. The cross-coupling is crucial here, it provides a negative resistance proportional to the inverse of the transconductance of the active devices. For a conventional RLC oscillator, the positive resistance degrades the signal over time, resulting in a oscillation of reducing amplitude. However, if a negative resistance is added to this node, the circuit will oscillate well if the net resistance is negative. This circuit is effectively guaranteed to start up as long as the cross-coupled negative resistance is large enough. Furthermore, current through the oscillator is controlled through the current source on the top of the inductor pair. This current source introduces some noise at the top of the inductor pair, which directly relates to phase noise, but it also provides a way to regulate the current of the system. Furthermore, it

aids in level shifting the output waveform so that it is no longer centered around the supply rail, which would make it useless for our purposes. A more detailed analysis of variants on this topology can be found in [2]. The downside of this topology is that it has a very limited power output, so needs a power buffer following it to be able to drive the input to the polyphase filter. This power buffer needs to be able to have a large power gain, which requires the transistor to be able to conduct a lot of current. The large transistor required for the buffer loads the VCO with a large capacitance which reduces the center frequency and frequency tuning range of the circuit. Following the algorithm presented in [12], I first sized my inductor. Doing a preliminary analysis, I found that I was actually limited in my inductor sizing, due to both matching the inductor resonant peak Q frequency to ω_{osc} as well as due to the large capacitances present on the node. Once the inductor size was chosen, I then sized my cross-coupled pair at the optimum current for minimum phase noise, as presented in [13]. I then designed the buffer, as presented in the next section. Once that was designed, I was able to size the varactor by modelling the non-varactor capacitors on that node using first-order parasitic expressions.

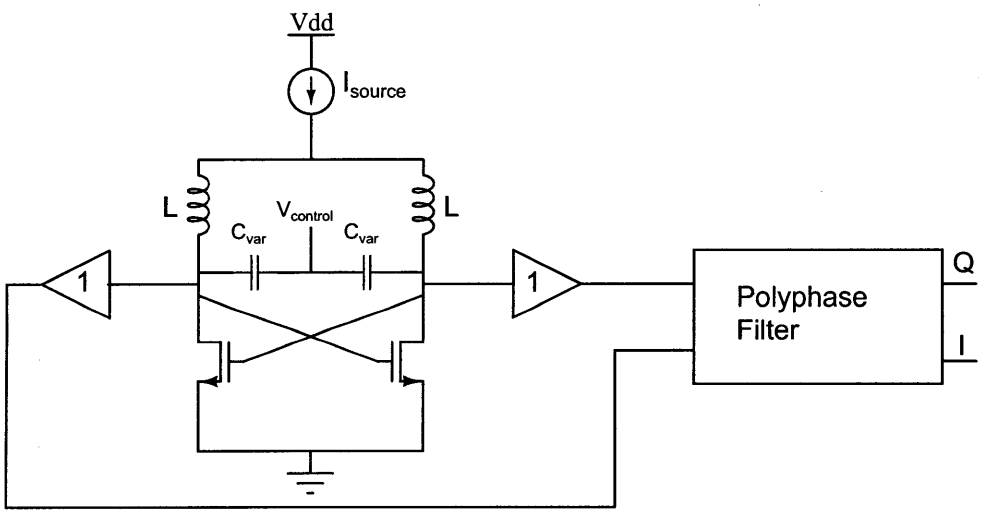


Figure 3-1: Cross-coupled VCO schematic

3.2 Power Buffer Designs

As mentioned in the previous section, the cross-coupled VCO was unable to provide enough power output to drive the polyphase filter with such a large signal. This section aims at exploring different PA topologies for providing high power output from the VCO module.

3.2.1 130-nm BiCMOS Class D Design

The final PA topology used is presented in Figure 3-3 below. This topology represents a fairly standard Class D amplifier as presented in [3], which consists of a set of switches, and a reactive tank for a load. The tank is tuned for operation around 45 GHz. This topology was chosen because it did not sacrifice quiescent current as a linear amplifier would. Furthermore, harmonic distortion proved not to be a problem, and was even filtered out due to the resonant tank.

3.2.2 130-nm BiCMOS Class A Design

A second PA topology explored was that of a Class A amplifier. I originally designed this circuit as a single-stage amplifier, but quickly realized that the gate capacitance the oscillator node would see would be too large to oscillate at 45 GHz. I thus designed the circuit seen in Figure 3-2 below as a two-stage narrowband power amplifier.

I did not optimize the circuit with respect to noise figure. Each stage provides some gain dependent on the impedance of the load at the optimal frequency, which was designed for 45 GHz. The load resistors are added for DC biasing so the output oscillates within an acceptable range. The load capacitors allow me to control the peaking frequency and magnitude of the circuit. The second amplifier is wider than the first, to provide more output power to the load. No impedance matching or level shifting was required for this circuit. One trick I used here was to use thin-gate FET devices for the cross-coupled oscillator, and thick-gate devices for the buffer circuit. This allowed me to use a higher supply rail on the buffer circuit, which allowed me to run the thick-gate FET devices in saturation, so that the output did not distort. It is important to note that while this buffer is narrowband, it still works quite well within a decent frequency range, from about 40-50 GHz, which is

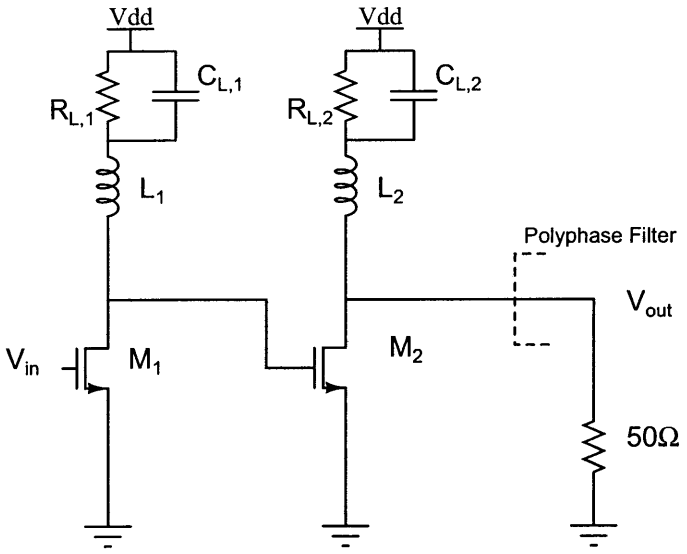


Figure 3-2: 130-nm BiCMOS narrowband power buffer

very important when it comes to tuning the oscillator. I ended up not using this topology due to the cost of the additional quiescent current, and the additional rail due to the use of thick-gate FET devices.

3.2.3 65-nm CMOS Buffer Design

The amplifier topology chosen to accompany the VCO core in the 65-nm process is presented below in Figure 3-3. This amplifier is once more a Class D RF amplifier, as discussed in [3]. Few differences were found between this circuit and the Class D buffer designed in the 130-nm BiCMOS process.

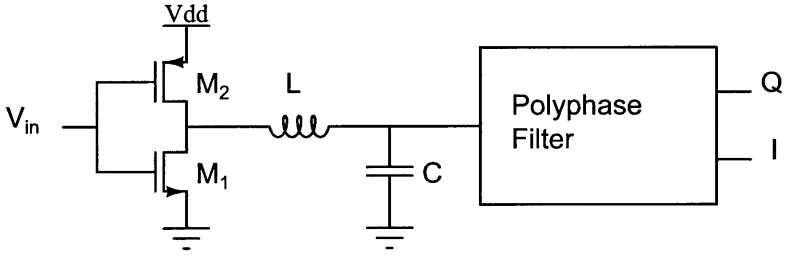


Figure 3-3: Class D power driver

3.2.4 Polyphase Filter

The DRFPC phase modulator system requires two differential clocks which are in quadrature with one another. Rather than generating two independent signals and attempting to phase lock them, this topology drives the polyphase filter with a single set of signals, which emerge from the network as two attenuated differential signals in quadrature with one another. A schematic of this system is presented in Figure 3-4 below. The resistors and capacitors are sized identically, with the RC product at $2f_0\pi$, with f_0 at 45 GHz. A full derivation and analysis for the transfer function of this network can be found in [15]. Effectively, the signal from the input will be attenuated by a factor of $\frac{\sqrt{2}}{2}$ at the designed center frequency, and the input impedance (differentially) is $\frac{R}{RCs+1}$.

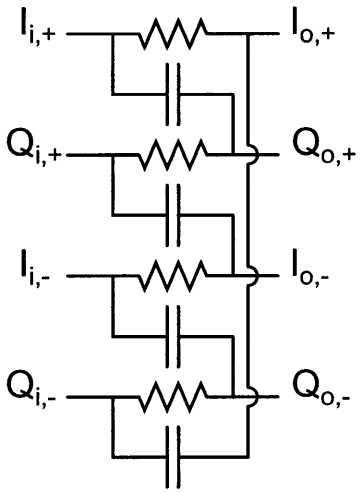


Figure 3-4: Polyphase filter

3.2.5 Simulation Results

This section will evaluate the overall performance of the VCO circuit. As stated previously, the most important considerations are phase noise and signal amplitude. Phase noise plots can be found in Figure 3-5 and Figure 3-8 below. As explained earlier, the phase noise at very low frequencies doesn't matter for our purposes, we don't care about the clock slowly varying with time. More importantly, we can see that the phase noise at higher

frequencies is fairly low. Again, this is not intended to be groundbreaking work, so the phase noise numbers are a little higher than state of the art. We can convert phase noise to phase jitter by effectively integrating the spectrum over the frequency bands of interest using the following formula adapted from [14]:

$$RMSJ_{period} = \frac{1}{2\pi f_c}\sqrt{2\int_{f_1}^{f_2} 10^{\frac{L(f)}{10}} df} \tag{3.2}$$

Where L(f) is the phase noise spectrum, in dBc, f_c is the carrier frequency, and f_1 and f_2 are the frequency bands of interest. We don't have a specific phase jitter specification for this project, but integrating the system from 10GHz to 100 GHz, we find that the phase noise due to this band is approximately 2.5fs in each of the processes, which is an acceptably small portion of the signal period. Given that we care about a smaller frequency band than this, this provides an upper limit on phase jitter which is acceptable for our purposes. It is also interesting to note the shape of the phase noise curve, it appears to be integrating over frequency, but flattens out at about 10 GHz. This makes sense, over frequency the white noise signal of the phase noise will decrease, but will hit a fundamental noise floor at some frequency at which point the system no longer rejects changes in frequency.

130-nm BiCMOS

The 130-nm BiCMOS VCO consumed 22 mW of current with the Class D power amplifier. The 130-nm BiCMOS VCO consumed 105mW of power with the Class A buffer, with the vast majority going to the quiescent current in the buffers. Other results are presented below.

The output of the polyphase filter can be seen below in Figure 3-6. As you can see, the output appears relatively clean. It turns out that the shape of the wave is not exceptionally important for our case, only the zero crossings as the clock is used differentially. Furthermore, the clock signals appear to be closely in quadrature with one another. The DRFPC can tolerate at least 15 degrees of offset error between the quadrature signals, these results fall in bounds of that specification. One output is larger than the other due to the method of driving the polyphase filter with only in-phase signals.

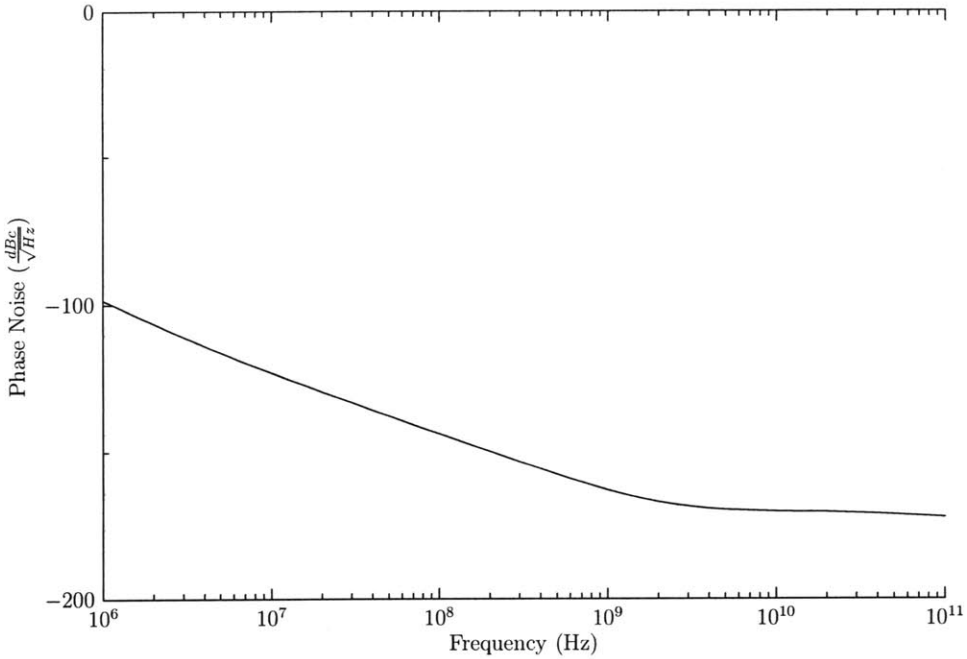


Figure 3-5: VCO phase noise, 130-nm

The last important result is the tuning range for the VCO. Figure 3-7, shown below, compares control voltage to frequency output. It takes essentially an "S" shape to it, which matches directly to the varactor to voltage curve for this process.

65-nm CMOS

The 65-nm CMOS VCO consumed about 21mW of power, mainly due to the removal of a linear amplifier for the buffer. Other results for this VCO are presented below.

The output of the polyphase filter can be seen below in Figure 3-9. The two output signals are very close in quadrature to one another. Past work has shown that the phase modulator can withstand I-Q mismatches of at least fifteen degrees, the error demonstrated here is lower than that.

Once more, the VCO only becomes useful if it can be tuned to the proper frequency. This VCO had a slightly smaller tuning range than the 130-nm VCO, as presented in Figure 3-10 below. This is primarily due to the fact that the driver here was sized larger for a larger output swing, which resulted in a larger load capacitance, and a smaller tuning capacitor.

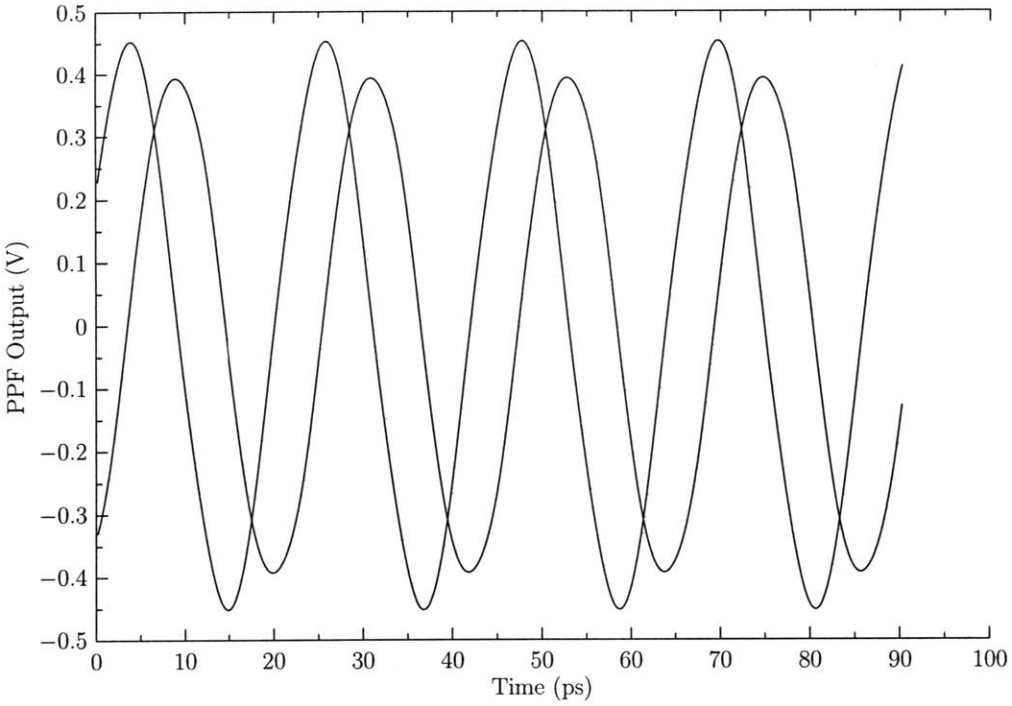


Figure 3-6: VCO output, 130-nm

Regardless, a tuning range of 4 GHz should be enough for our needs. With respect to phase noise, the 65-nm CMOS VCO has worst phase noise performance than the 130-nm BiCMOS VCO. I attribute this to the quality of the components in each circuit, the 130-nm version was able to take advantage of better modeled devices.

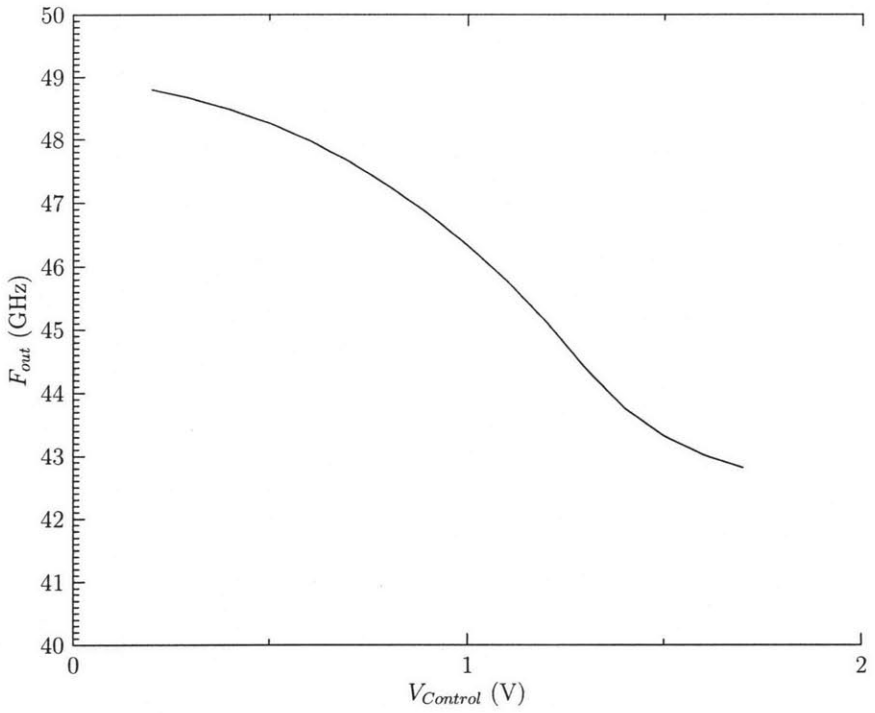


Figure 3-7: Tuning range for VCO, 130-nm

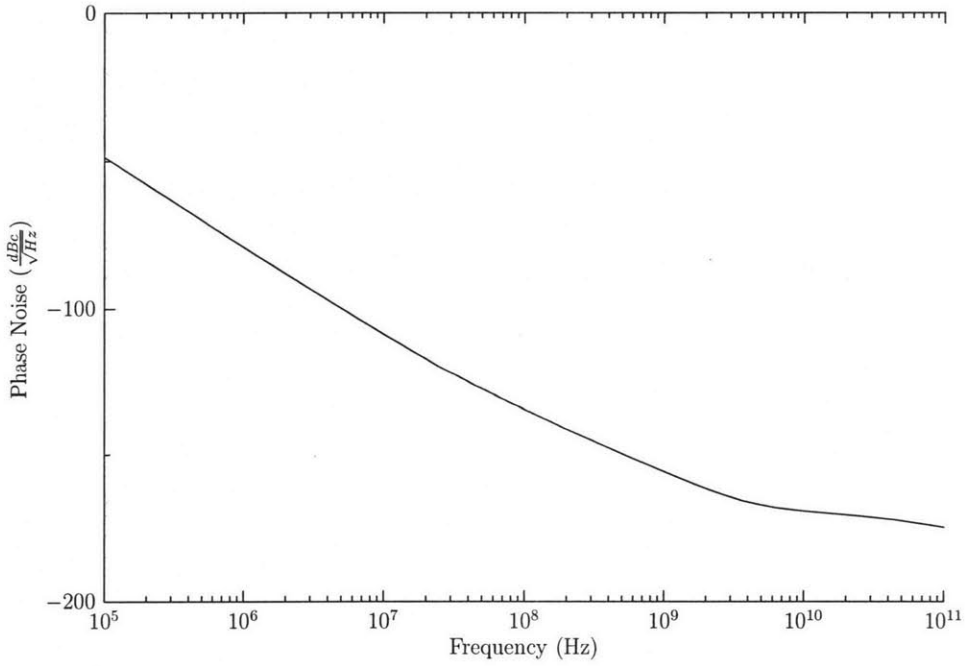


Figure 3-8: VCO phase noise, 65-nm

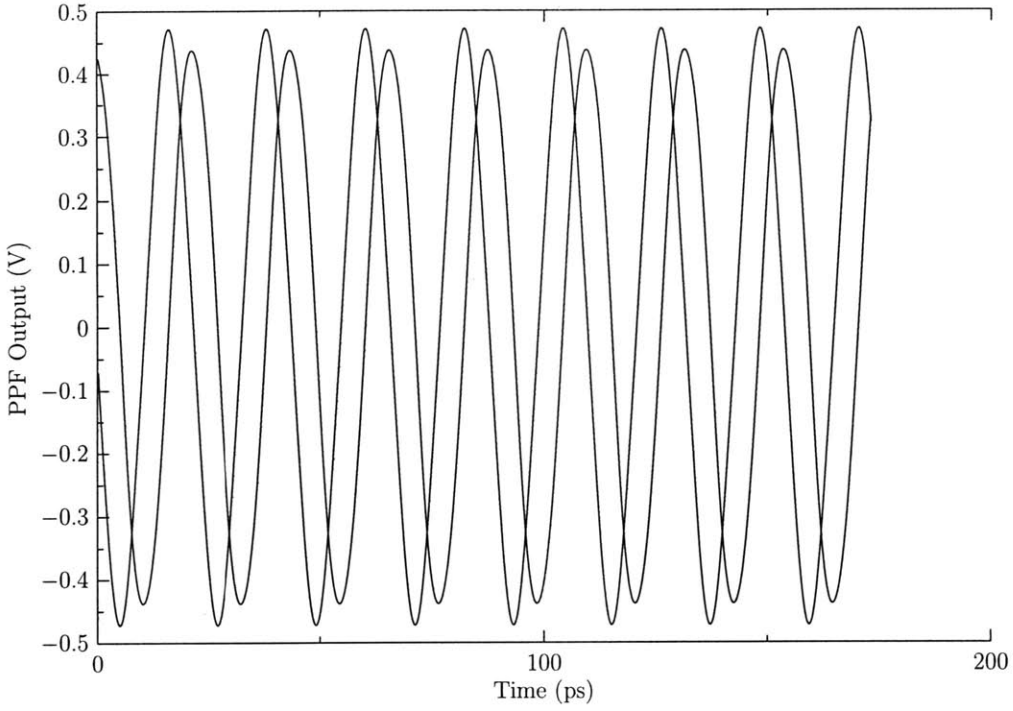


Figure 3-9: VCO output, 65-nm

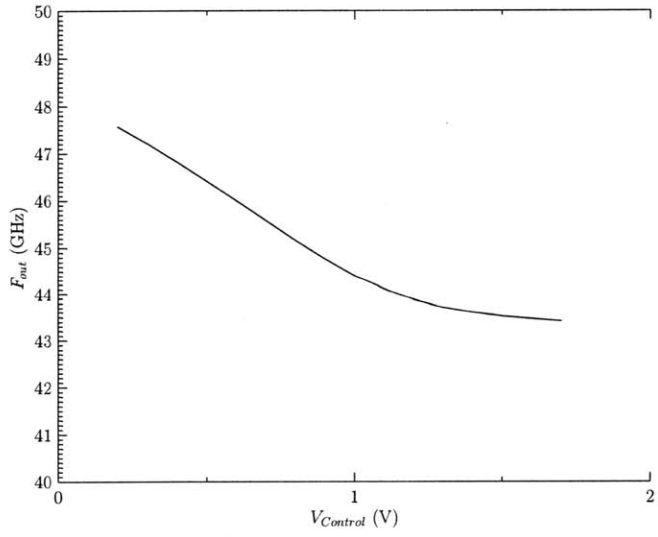


Figure 3-10: Tuning range for VCO, 65-nm

Chapter 4

Layout and Results

4.1 Introduction

This first section of this chapter will serve to explain the design decisions involved in laying out the 130-nm BiCMOS chip. The second section of this chapter will illustrate a comparison between initial simulated results as well as results after parasitic extraction has taken place. The final section will consist of a brief conclusions of my work, and suggestions for further exploration in later revisions of the system.

4.2 Layout

Figure 4-1 below illustrates the generalized floorplan for the 130-nm BiCMOS chip. This chip was limited in that it contained only PAs and switching cells. PA drive signals are brought on-chip through probe pins (not shown), and routed on-chip through transmission lines. Switch drive signals are brought on-chip through normal chip pads (labeled), and pass through LVDS receivers (not shown). The output of the LVDS receiver drives the decoder block for each switching cell (also not shown). The decoder block drives the switch cells, which consist of the drivers, transmission gate switches, and power rail connections. A more detailed sketch of an individual switch cell can be found in Figure 4-2 below. Between each set of 4 switch and PA cells, extra space is allowed for bringing in power supply wires horizontally, allowing us to reduce the total width of each switch cell. This

chip was configured to allow for the use of an on-chip as well as an off-chip combiner to make power efficiency measurements.

The switching cell shown in Figure 4-2 was constrained to a vertical pitch of 215 μm. This constraint emerged from the power combining network. Power combining efficiency was found to be degraded if the cells were spaced further apart. Because of this size of the transistors involved, the cell needs to extend horizontally. Two possible layout configurations came to mind, one of which is shown in Figure 4-2 below. The alternative was to put the switches as close to the PA cell as possible to reduce wiring parasitics at the top of the PA. Due to the dimensions of the cells, however, the length of this configuration proved to be too long for the chip size requested. For reference, the PA connects to the right of this cell on the left side of the chip, on the right side of the chip, we used a mirror image. Because a single cell was so long ($215\mu m$ x $1400\mu m$), the signal routing was a challenge. I modelled the interconnects, and placed a buffer every 500 μm along the signal paths to regenerate the system. Furthermore, the signal path lengths from the decoder block (above the center of a switching cell) need to be matched. I matched the length of each of the seven control signals by weaving the control signal line until it matched. This allowed for better time alignments in the drivers.

In Figure 4-2, I placed each driver chain directly above either an n-type or p-type switch. Together, one of each of these makes a single transmission gate. Each driver requires a separate V_{dd} and ground line, running vertically. Each transmission gate also requires a separate power supply line, denoted as $V_{dd,X}$. Here, the lowest supply rail is $V_{dd,1}$, and the highest is $V_{dd,4}$. Because peak efficiency is slightly more important than average efficiency for our agenda, I placed the highest rail closest to the PA cell in order to reduce losses associated with this level.

Power was routed through the top level metal lines, which have a much higher current carrying capability than the lower metal lines. Each amplitude path of 4 PA cells was provided with its own set of power supply rails.

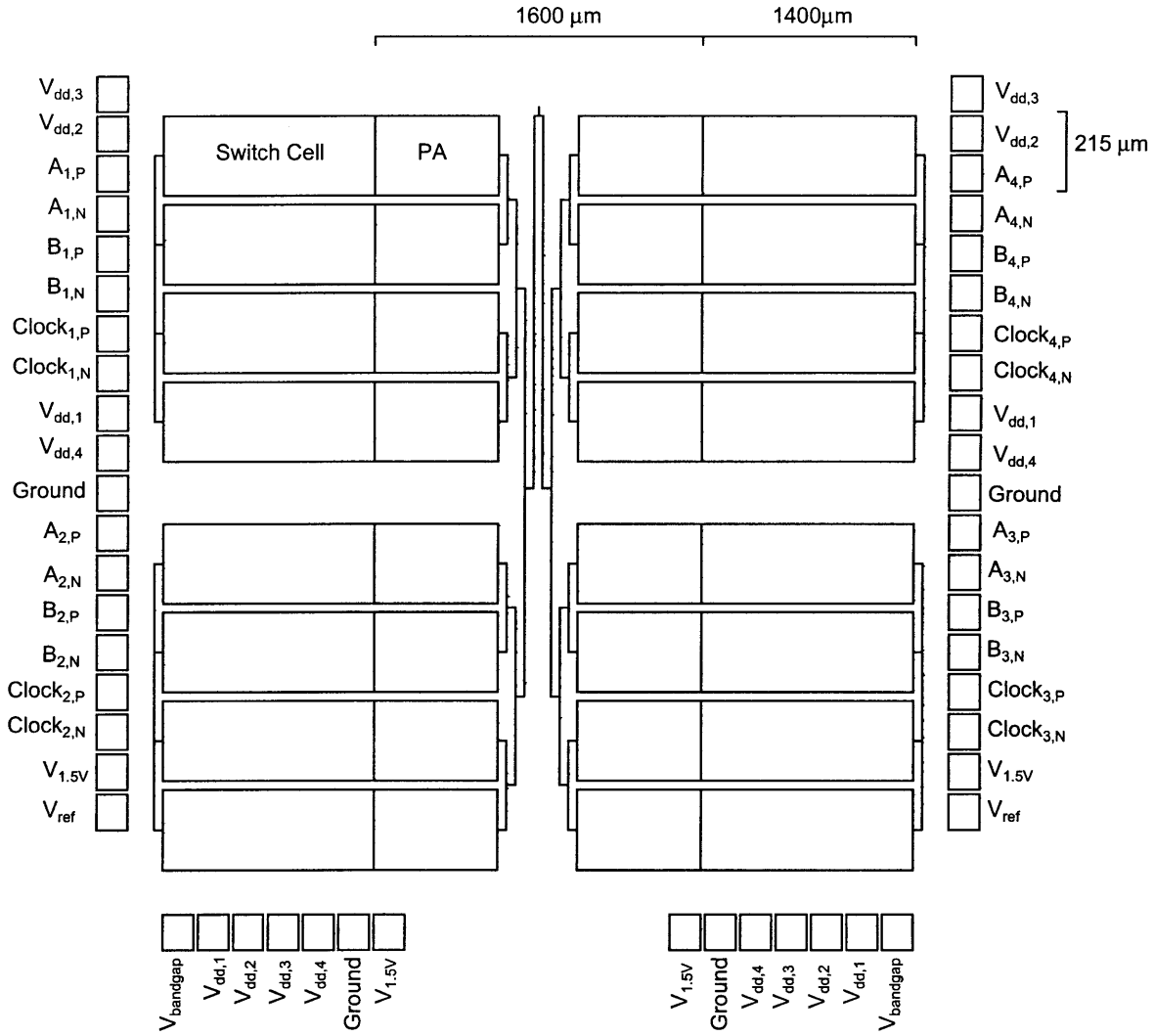


Figure 4-1: 130-nm Chip floorplan

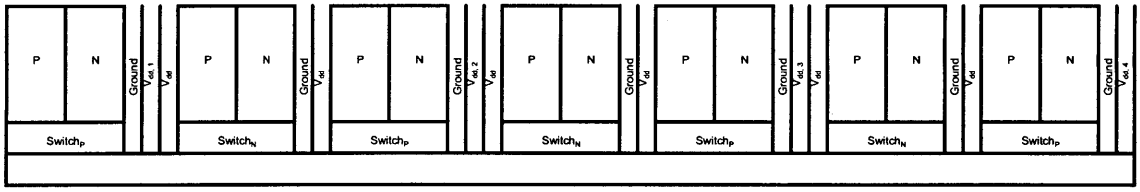


Figure 4-2: Switch cell floorplan

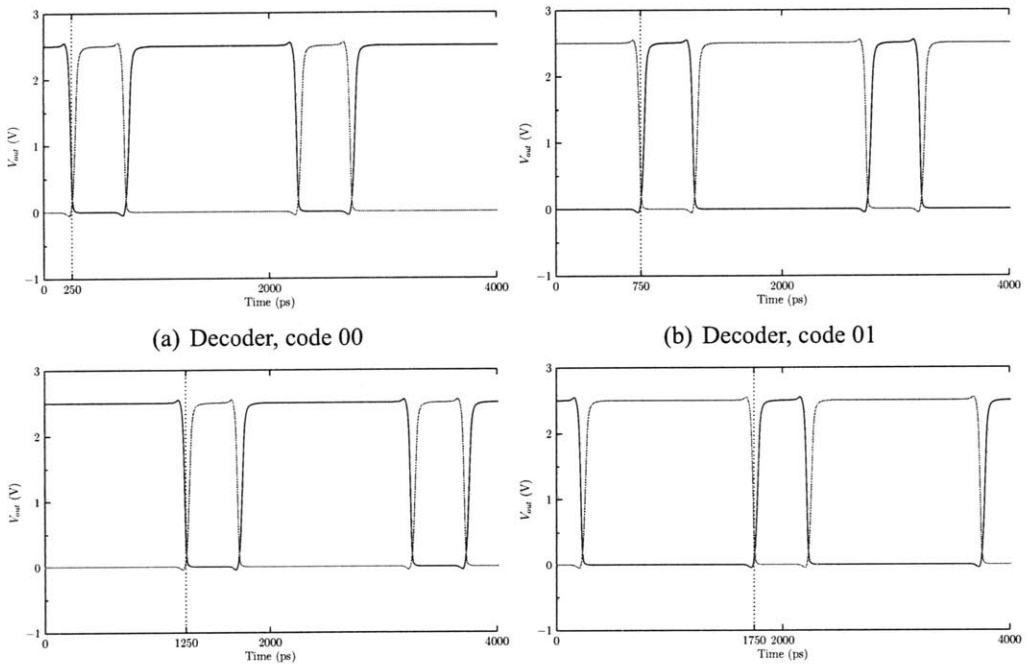
4.3 Simulation Results

This section serves to showcase my simulation results, and to compare my schematic simulations with my post-layout simulations. I will focus on the decoder block first, and then

the main switching block.

4.3.1 Decoder Waveforms

The output of the schematic view of the decoder block can be found in Figure 4-3 below. These plots demonstrate differential outputs on the four separate control voltages, used to turn on the switch drivers. In these plots, I am able to clock in different symbols at a rate of 2 GSPS. After layout, the speed in my circuit dropped, as demonstrated in Figure 4-4 below. I believe this is primarily due to using the lowest metal layer for my logic, the speed might be increased by using smaller wires at a higher metal level. I was able to get the decoder block to function reasonable well at 1.5 GHz by tuning the offset of the clock. Propagation delays were not well matched through this circuit block, so testing should allow for a tunable offset in the clock path.



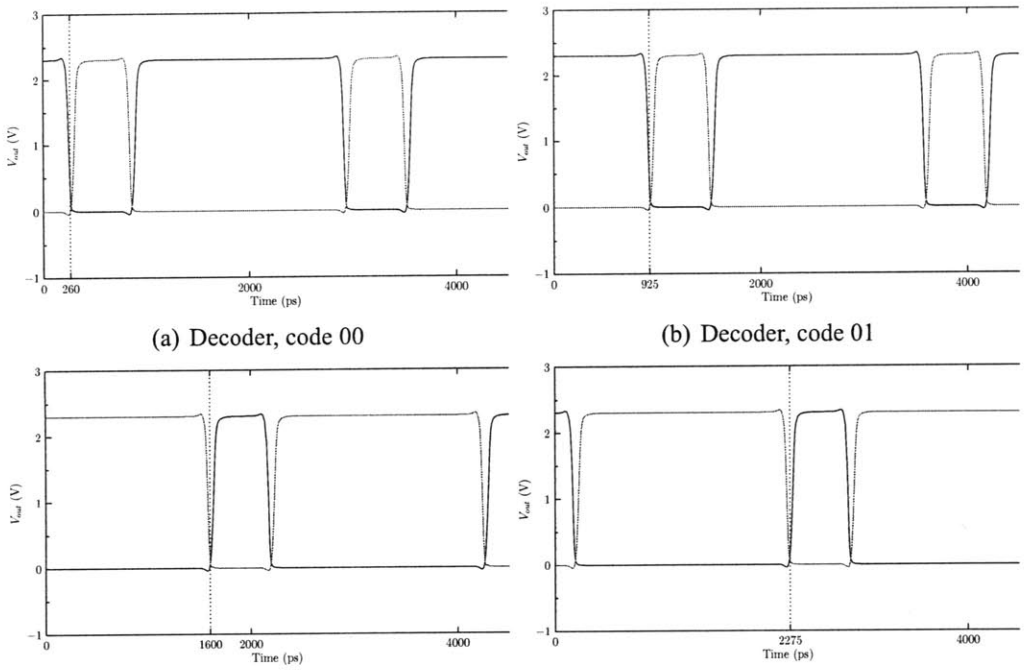
(a) Decoder, code 00

(b) Decoder, code 01

(c) Decoder, code 10

(d) Decoder, code 11

Figure 4-3: Decoder block output waveforms, ideal simulation at 2 GHz



(a) Decoder, code 00

(b) Decoder, code 01

(c) Decoder, code 10

(d) Decoder, code 11

Figure 4-4: Decoder block output waveforms, parasitic extraction simulation at 1.5 GHz

4.3.2 Switch Rise Times

The bulk of my work was ensuring that my switches could handle a 2 GSPS rate. Below, Figure 4-5 shows the supply at the top of the PA as it is switched up and down. I tested the system with voltage levels of 2.4V, 2.1V, 1.8V, and 1.5V, primarily due to even spacing between levels. As Figure 4-5 demonstrates, the schematic view of the system does not seem to have a problem with switching rates, the switches settle within 115 ps. There is some ringing involved, due to the sudden step change into the PA system, but this problem is taken care of when parasitics are involved. Figures 4-6, 4-8, and 4-10 demonstrate the functionality of the switching network after parasitic extractions. As is demonstrated by these plots, the system provides a cleaner sampling at lower speeds, primarily due to increased settling time. The supply voltage settles within about 200 ps for two adjacent levels, this value is increased for larger switches. Effectively, these plots demonstrate that the system is functional for a 2 Giga-Samples per Second, but perhaps not 2 Giga-Symbols per Second. By setting our oversampled rate at 2 GHz, we can avoid switching from the

lowest level to the highest level between periods, the system will have no problem settling into an intermediate value in a single period. Figures 4-7, 4-9, and 4-11 demonstrate the output of the PA while the switches are operating. Clearly, the output signal is amplitude modulated quite nicely. Secondary to the order of switching time is the question of efficiency. These switches see a little more ohmic loss than the ideal switch schematics do, this is due primarily to the added resistivity of the interconnections. Overall however, the parasitic extraction simulations match closely with the ideal simulations with respect to performance.

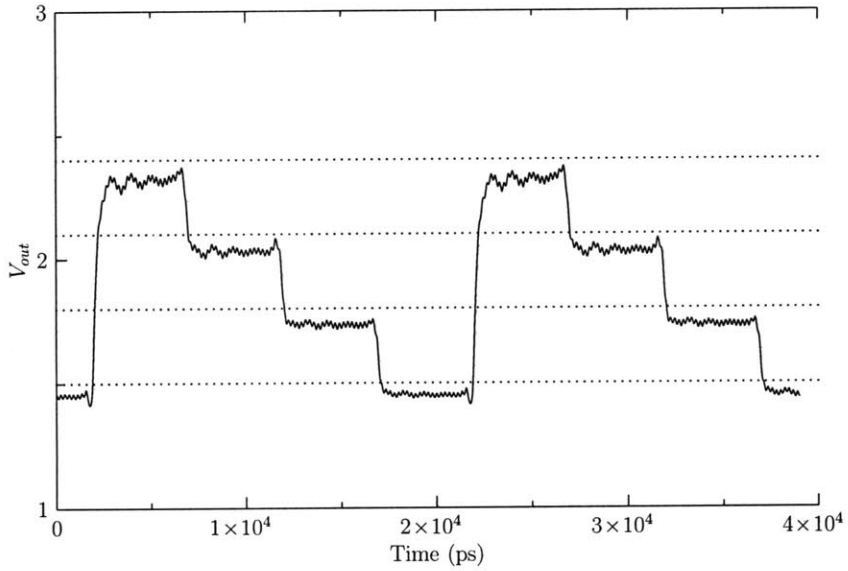


Figure 4-5: Switching cell waveforms, ideal

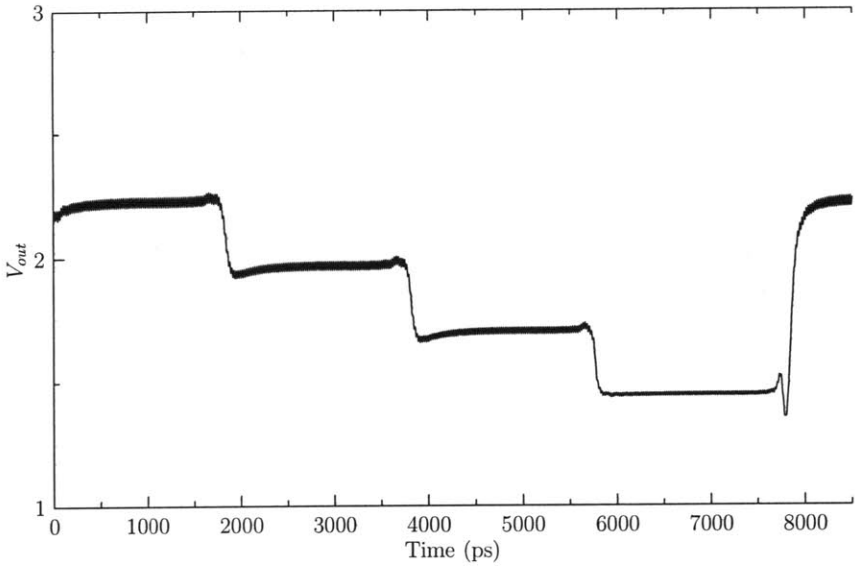


Figure 4-6: Switching cell waveforms, with parasitics, 500 MHz

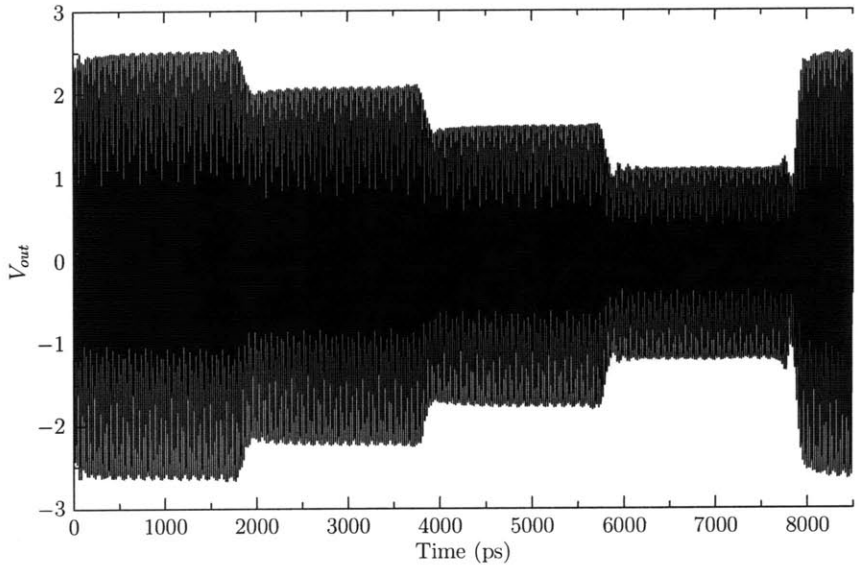


Figure 4-7: Power amplifier output, with parasitics, 500 MHz

4.4 Conclusion

As stated previously, the AMO system requires a set of power switches to modulate the supply voltage for the PA. These switches need to be as efficient as possible while retain-

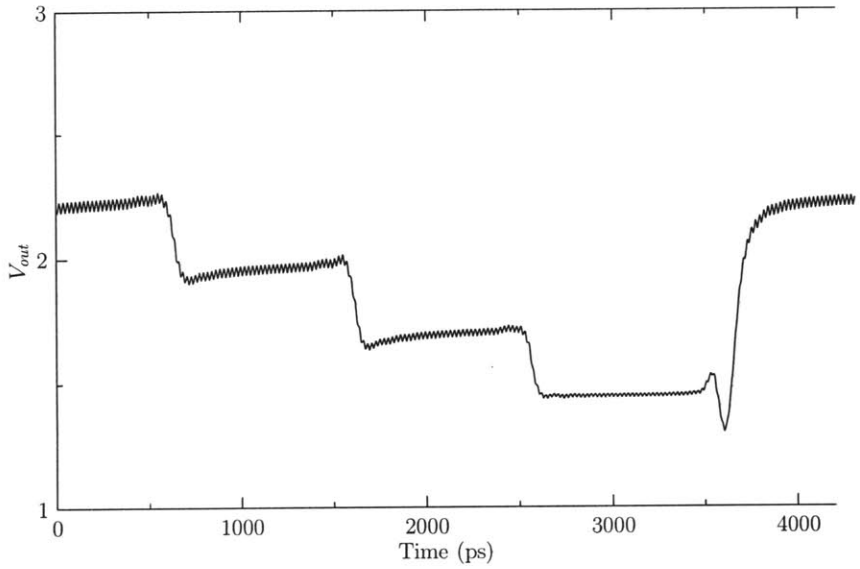


Figure 4-8: Switching cell waveforms, with parasitics, 1 GHz

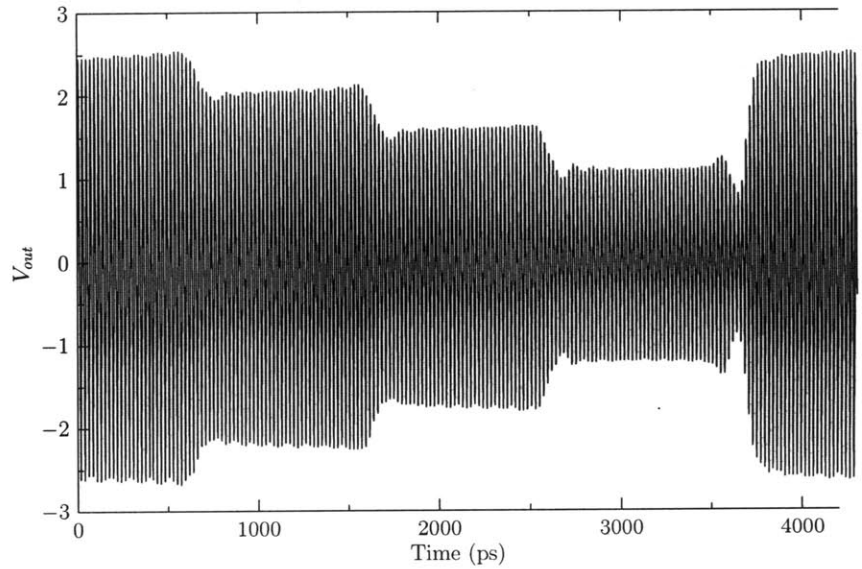


Figure 4-9: Power amplifier output, with parasitics, 1 GHz

ing the ability to switch at a sample rate of 2 GHz. The speed requirement ensures that the switching action does not negatively impact the spectrum of the system, while the efficiency requirement is required to meet the overall system efficiency specification. This work presents a technique for optimizing power switch design for the AMO architecture

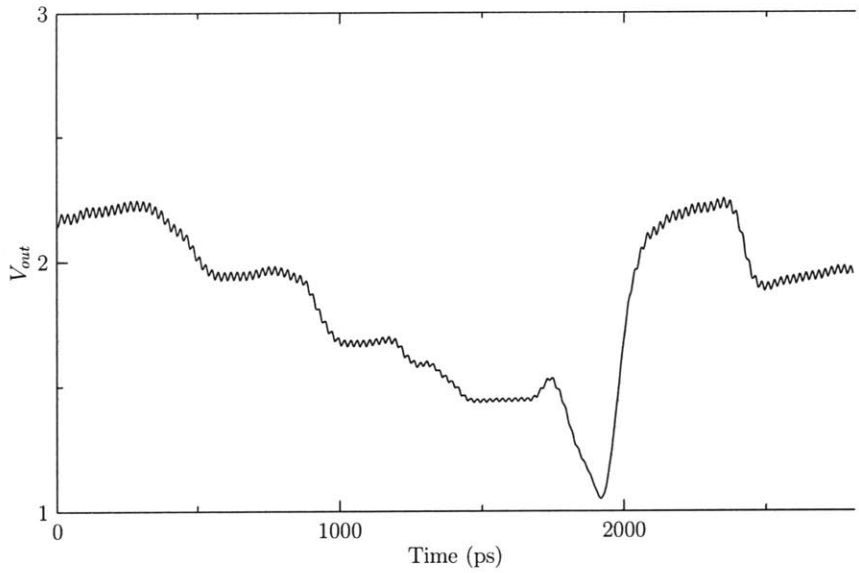


Figure 4-10: Switching cell waveforms, with parasitics, 2 GHz

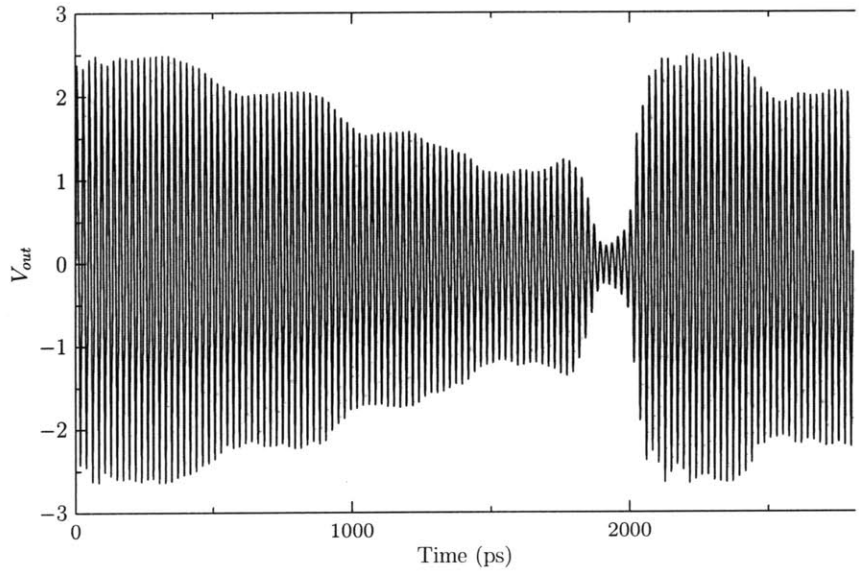


Figure 4-11: Power amplifier output, with parasitics, 2 GHz

by minimizing power lost in the switches. The schematic and post-layout simulations presented here demonstrate that the switches designed in this 130-nm IBM SiGe BiCMOS process meet the switching speed requirements presented above while minimizing power loss.

Future work will need to port this design procedure to a new process node. Smaller process nodes should perform better, primarily due to the reduction of capacitances on the power switching node. Once more, future designers will need to minimize the combination of switching and ohmic power losses while ensuring that switch rise times are met. Furthermore, a smaller process node should allow the reduction of this rise time, allowing for a sample rate faster than 2 GHz. I would suggest that future designs look into the cross-coupled driver topology presented in Chapter 2, it shows distinct promise for keeping the drivers on a transmission gate well aligned. Furthermore, the feedback system presented in Figure 2-16 is not difficult to adapt to other processes, and will provide time alignment between amplitude paths.

Finally, my VCO designs can be adapted to future process nodes as well. I highly recommend the cross-coupled topology followed by a switching driver, this provided the best efficiency and phase noise while still being able to drive a polyphase filter. Further information regarding V_{dd} levels and optimum switching frequencies can be found in Appendix A. While these particular results are process-specific, the algorithms presented here are easily adaptable to future nodes.