Advanced Architectures for Efficient mm-Wave CMOS Wireless Transmitters

JiaShu Chen

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2014-42
http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-42.html

May 1, 2014

Advanced Architectures for Efficient mm-Wave CMOS Wireless Transmitters

by

Jiashu Chen

B.E. (City University of Hong Kong) 2007

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Engineeing - Electrical Engineeing and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Ali M. Niknejad, Chair
Professor Robert G. Meyer
Professor Paul K. Wright

Fall 2013

The dissertation of Jiashu Chen is approved.

Chair Date

Date

Date

University of California, Berkeley
Fall 2013

Advanced Architectures for Efficient mm-Wave CMOS Wireless Transmitters

by

Jiashu Chen

Abstract

Advanced Architectures for Efficient mm-Wave CMOS Wireless Transmitters

by

Jiashu Chen

Doctor of Philosophy in Engineeing - Electrical Engineeing and Computer Sciences

University of California, Berkeley

Professor Ali M. Niknejad, Chair

With fast growing consumer demand for high speed mobile data capacity, wireless spectrum has become increasingly precious. This drives the evolution of the personal wireless communication, with new standards developed to improve the spectral efficiency. However, the available spectrum below 10GHz is very limited and packing more bits per second into the same bandwidth requires larger energy consumption as well as more stringent radio and MODEM performance. As a result, such an approach is not sustainable for meeting the future demand. A natural path is to move into higher frequency bands which have larger spectrum bandwidth but less commercial usage. Recent years have witnessed vast technology development on V-band (60GHz) Wireless Personal Area Networks (WPAN) and E-band (80GHz) point-to-point cellular backhauls. Meanwhile, the advancement of low-cost CMOS technologies enables researchers to significantly improve the integration level of high speed mm-wave radios with traditional analog and digital circuitry. However, current mm-wave radio transmitters suffer from short communication distance and low energy efficiency. This is mainly caused by the reduced performance of the CMOS transmitters employing traditional Power Amplifiers (PAs) that suffer from low transistor breakdown voltage, low power gain and poor back-off characteristics. This dissertation investigates the challenges of designing efficient mm-wave transmitters for both long range and short range applications, and proposes concepts and techniques that can potentially break the barriers imposed by the low cost digital CMOS process. The scope of investigation and proposal extends from the architecture level down to the transistor level. Specifically, on-chip and spatial power combining techniques are analyzed and implemented to achieve larger transmitter Equivalent Isotropically Radiated Power (EIRP). To enhance the average efficiency for modulated signals with high Peak-to-Average-Power-Ratio (PAPR), a direct digital-to-RF conversion architecture is proposed and implemented, enabling dynamic DC power scaling. Finally, a Quadrature Spatial Combining concept is introduced to eliminate the tradeoff between low insertion loss and high isolation present in a traditional Cartesian architecture with on-chip signal combiners. Prototype chips are fabricated and tested in 65nm CMOS technology to verify the proposed architectures and techniques.

Dedicated to my parents and grandparents, without whom I would never have reached this far...

Contents

List of Figures

List of Tables

Acknowledgements

Time flies. Six years passed since I came to the United States to embark on this journey that I will always remember throughout my life. Like most journeys, this is a colorful one full of stories, during which I've encountered excitement and anxieties, obstacles and accomplishments, praises and criticisms. Yet, I could never confidently reach the finish line without generous help and insightful guidance from many important people during these years.

The best part of this journey was meeting my advisor Prof. Ali M. Niknejad and working in his group. From a microwave background, I was initially lost in the sea of transistors. However, Ali was patient to share his exceptional knowledge in both fields and helped me to bridge the gap. Through his encouragements, I gradually gained confidence. Being a knowledgeable and humble person, Ali taught me not only technical skills but also readiness to learn new things and adapt to changes, a mentality shared by many successful people in this rapidly involving world. I also greatly appreciate him for giving me the freedom to practice in the industry throughout the graduate study, which turns out to be tremendously rewarding.

I am also very grateful to Prof. Elad Alon who has taught me three IC design courses and advised me on a number of research projects. Elad is an excellent instructor and his courses introduced me to a variety of topics and helped me develop insights for analyzing circuits. Although I am not his student, he is always ready to mentor, and I enjoyed discussing with him. I would also like to thank Prof. Robert Meyer and Prof. Paul Wright for serving as my qualification exam committee. I had an opportunity to work with Prof. Meyer during one of my internship, and I was very grateful to him not only for his advice but also for what he had shown to me as a true scientist.

I was never alone during this journey. I was so happy to meet my fellow classmates who came in the same year as me: Lingkai Kong, Wenting Zhou, Paul Liu, Yida Duan, Lu Ye, Hanh-Phuc Le, Chintan Takkar, Maryam Tabesh, Jung-Dong Park, Rikky Muller, Mervin John, Ping-Chen Huang, Namseog Kim, Kyoohyun Noh. We've been through all the challenges together and I wish you all the best. My life at Berkeley Wireless Research Center was enriched by the company of many other graduate students: Yue Lu, Wen Li, Jun-Chau Chien, Siva Thyagarajan, Charles Wu, Shingwon Kang, Kuangmo Jung, John Crossley, Steven Callendar, Jaehwa Kwak, Zhiming Deng, Amin Arbabian, Debo Chowdhury, Bagher Afshar and Ehsan Adabi.

Studying at BWRC has brought me the precious access to the semiconductor industry. Throughout the last two years, I've been largely involved in the R&D at Tensorcom, a WiGig start-up company. I was exposed to various aspect of a fabless start-up company, from chip design to customer presentation. The technical experience brought many practical elements into my research projects and helped me to reflect deeper on issues that are typically overlooked in an academic work. The wide exposure at the startup was also extremely valuable to my career. I would like thank co-founders Hock Law and Ismail Lakkis for inviting me to participate. Besides, I was lucky to get to know Sohrab Emami and I would like to thank him for allowing me to test my 60GHz PA at SiBEAM, where I was truly

impressed by the remarkable technological achievement of the start-up founded by former BWRC graduates.

Now that I am close to finishing my PhD study, I cannot resist of thinking about my undergraduate advisor Prof. C.H. Chan, who brought me onto this road. I still clearly remember the first day when I arrived at City University of Hong Kong and was lost in the gigantic academic building when I bumped into him. He showed me around the EE department and encouraged me to take full advantage of the school resources. Three years later, I became his FYP student and he recommended me to take the RF integrated circuit topic. He also showed tremendous support for my graduate school application, and without him I could've not come to my dream school.

My graduate study was also made possible by the Fulbright foundation, and I am so proud to be one of the 27 worldwide recipients of the International Fulbright Science and Technology Fellowship in its inaugural year. My thanks go to Kate Leiva and Tom Koerber at the Institute of International Education and Lance Sung at the Hong Kong US Consulate for their support.

Finally and as always, I want to give a special acknowledgement to my parents and grandparents. Nothing can be even compared to the love and support from you, and no words can ever express my gratitude and love to you.

Chapter 1

Introduction

Since the global smartphone revolution, the consumer demand for wireless data capacity has been rocketing. The percentage of Internet traffic coming from mobile devices has increased dramatically from less than 1% in 2009 to more than 20% in 2012, and it's predicted that the mobile data traffic volume will grow 15 times within the next five years (Fig. 1.1) [1]. This increasing data demand constantly drives the evolution of the wireless communications, with new standards being developed to provide higher speed and larger capacity. In less than a decade, the data rate of both WiFi and cellular networks have increased from less than 1Mbps to more than 100Mb/s today. This is mainly achieved by using larger radio bandwidth and obtaining higher spectral efficiency. For example, the 802.11g standard utilizes 20MHz bandwidth with 64QAM modulation scheme whereas the draft 802.11ac standard utilizes 160MHz bandwidth with highest modulation scheme of 256QAM, which results in 16 times data rate improvement. However, the available spectrum below 10GHz is very limited and packing more bits per second into the same bandwidth requires larger energy consumption as well as much more stringent radio and modem performance. As a result, current approaches may not be sustainable for meeting the future demands.

In contrast, the mm-wave frequency band has much larger spectrum bandwidth but very minimum commercial usage. In recent years, much effort has been made to utilize the advantages of the mm-wave frequency band. One example is the development of the 60GHz Wireless Personal Area Networks (WPAN). The 7GHz unlicensed bandwidth provides 10 times speed improvement compared to the current 802.11n standard, and therefore enables various new applications such as wireless HD video streaming, instant data synchronization. Another example is the E-band (71-76GHz, 81-86GHz) cellular backhaul. Compared to current wireless backhaul systems which use congested frequency bands below 38GHz, the E-band backhaul not only improves the data rate by at least 4 times, but also enables spectrum reuse due to the narrow beam feature of the long distance mm-wave transmission.

The ubiquitous deployment of mm-wave wireless communications is made possible by the low cost integrated Complementary Metal Oxide Semiconductor (CMOS) solution. Tra-

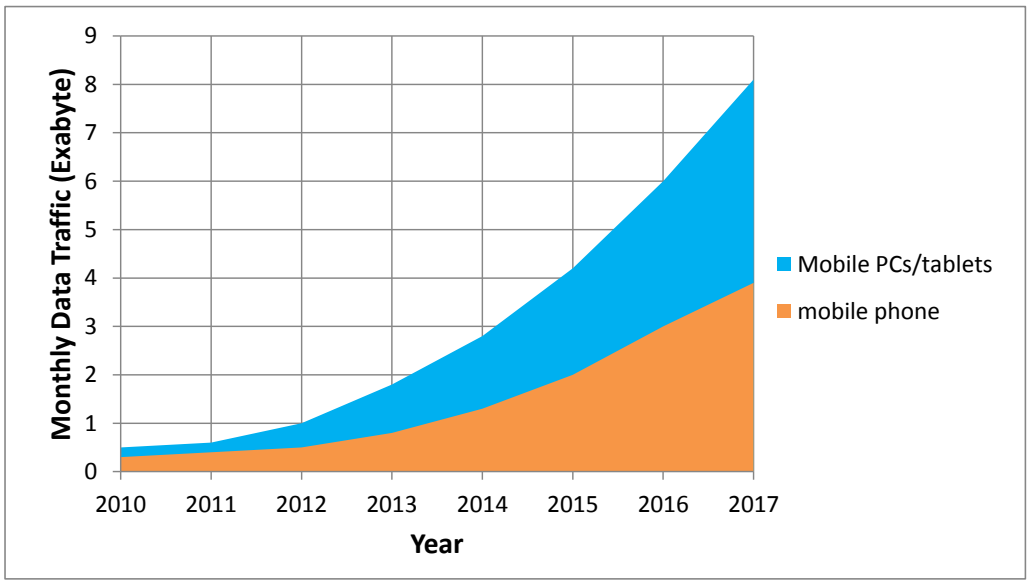


Figure 1.1: Predicted mobile traffic growth

ditionally all the mm-wave radios are implemented by discrete components or compound semiconductor based Monolithic Microwave Integrated Circuits (MMICs), which are not only expensive but also limited in capabilities. Due to process scaling, CMOS technology which was mainly used for digital computations and certain low frequency analog circuits, is now capable of operating at speeds in excess of 100GHz. According to ITRS roadmap for RF CMOS technology as shown in Fig. 1.2, the maximum transit frequency (f_T) will reach 1THz by the end of this decade [2]. Taking full advantages of the increasing transistor speed, researchers are now able to design circuit blocks and complete transceivers in the mm-wave frequency domain using bulk CMOS. In the past decade, a clear trend is seen that engineers have been aggressively pushing the envelope of mm-wave CMOS design, demonstrating sources and detectors beyond 500GHz and fully integrated transceivers close to 300GHz (Fig. 1.3).

In spite of the glory that the circuit frequency world record is being set every year, the mm-wave CMOS radios today have fairly short communication range. The reasons are two fold. First, the path loss increases with the frequency and therefore larger Equivalent Isotropically Radiated Power (EIRP) is needed at higher frequency to cover the same amount of transmission distance. Second, due to low transistor breakdown voltage and parasitic loss, CMOS transmitters have much inferior power delivery capability at higher frequencies, limiting the achievable output power and EIRP. In addition to short link distance, current CMOS mm-wave radios also have very poor energy efficiency, particularly on the transmitter side. The efficiency of current mm-wave transmitters is only about 20% to 30% of WiFi and celluar transmitters in the sub-10GHz frequency range. Among many factors that cause this phenomenon, low transistor breakdown voltage, low power gain and large passive loss are the major contributors.

What's worse? To obtain better spectrum efficiency and immunity to multipath effect,

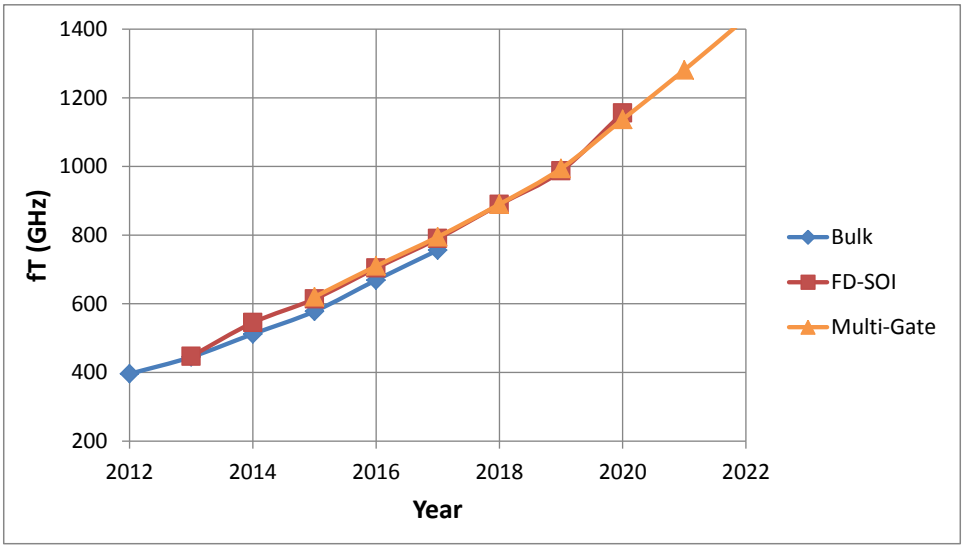


Figure 1.2: ITRS roadmap for RF CMOS

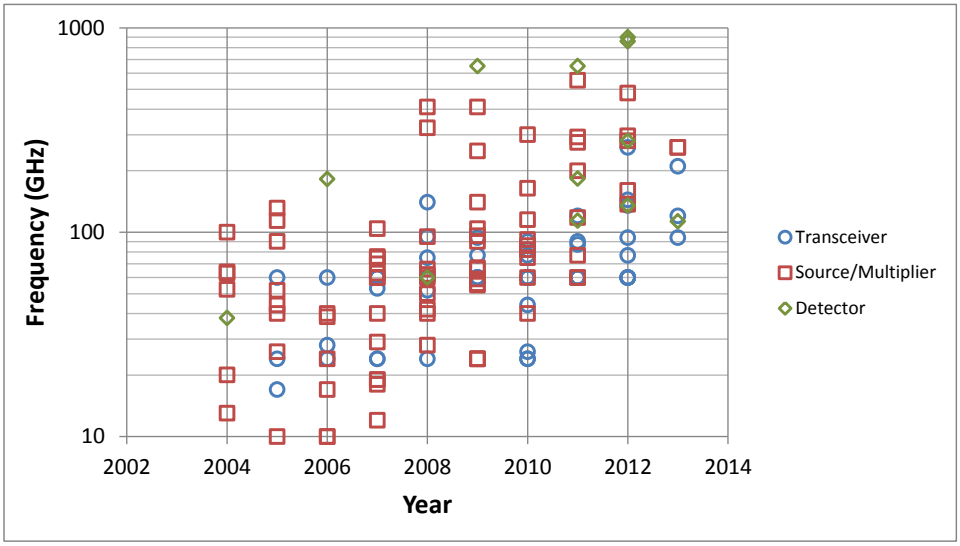


Figure 1.3: Operating frequency of CMOS circuits and systems

modern communication modulation schemes use high order QAM and Orthogonal Frequency Division Multiplexing (OFDM). As a result, the modulated radio signal usually has fairly high Peak-to-Average-Power-Ratio (PAPR), which means the transmitter has to back-off from its peak output power level. For a linear transmitter, which has been the default choice for almost all mm-wave radios, the power efficiency decreases linearly with output power level, therefore the average efficiency diminishes very quickly when the transmitter backs off from its peak momentum. As an example, a typical linear mm-wave transmitter with 10% peak efficiency will only have 2.5% average efficiency at 6 dB back-off when delivering modulated signals. This means for every 100W of power consumed, 97.5W are being wasted.

The same back-off characteristic exist for WiFi and cellular counterparts, however since the peak efficiency is much higher, the average efficiency is also proportionally better. In addition, various average efficiency enhancing techniques have been introduced at both the architecture level and the circuit level for WiFi and cellular transmitters, demonstrating remarkable improvement. In contrast, there is much less similar endeavor at the mm-wave domain, mainly due to the difficulty of adopting existing architectures and techniques.

This dissertation investigates the challenges of designing efficient mm-wave transmitters for both long range and short range applications, and proposes concepts and techniques that can potentially break the barriers imposed by the low cost digital CMOS process. The scope of investigation and proposal extends from the architecture level down to the transistor level. Indeed it's shown that the holistic optimization for a particular application at different design hierarchies is the key to achieving overall energy efficiency. The important contribution of this dissertation is summarized below.

1. The fundamentals of the CMOS process are analyzed from mm-wave radio designers' perspective. Bottlenecks have been identified that prevent the implementation of high performance mm-wave transmitters.

2. The traditional Class-E/F switching amplifier family has been extended with the potential benefits explained.

3. mm-Wave on-chip power combining techniques are analyzed. A compact and low-loss solution is proposed to enhance the output power of single-element transmitters. An optimization procedure is outlined and the theoretical limits are predicted.

4. The general design procedure for mm-wave linear Power Amplifiers (PAs) is documented.

5. The pros and cons of various beamforming transmitter architectures are analyzed. The optimal architecture choice is predicted based on array size, operating frequency and process.

6. An analog baseband phase shifting beamforming transceiver is implemented for short range high data rate links.

7. The direct digital-to-RF conversion architecture is analyzed and compared to traditional efficiency-enhancing architectures. Optimal mm-Wave CMOS implementation is proposed.

8. The concept of Quadrature Spatial Combing is proposed to solve the dilemma between minimizing insertion loss and minimizing undesired load-pull.

The remainder of the dissertation is organized as follows. Chapter 2 discusses the basic features of wireless transmitters, including various architectures and link budget analysis. It also introduces the most important block inside the transmitter: the power amplifier. Different classes of power amplifier topologies are presented, with insights on the design difficulty at mm-wave frequencies. Chapter 3 presents the design procedure for linear transmitters,

with focus on techniques that improve the PA output power. Various on-chip power combiners are discussed, and a compact solution is used to realize a 19dBm 60GHz PA. Chapter 4 focuses on using beamforming techniques to improve the transmitter EIRP. The tradeoff among several different beamforming architectures is presented. An optimal analog baseband phase shifting topology is chosen for an ultra-low power 4-element 60GHz phased-array transmitter covering 2 meters of communication range. In Chapter 5, the main focus shifts to the average efficiency enhancement. It first discusses the existing techniques including envelope tracking, Envelope Elimination and Restoration (EER), and outphasing, as well as the reasons why current architectures are not suitable for mm-wave applications. It then introduces the concept of direct digital to RF conversion as an effective approach for enhancing back-off efficiency of mm-wave transmitters. Optimal circuit level implementation is also presented. In particular, the concept of quadrature spatial combining is introduced as an effective signal combiner for Cartesian transmitters. A WiGig prototype is built based on the proposed transmitter architecture and measurement results are presented. Finally, Chapter 6 summarizes the important findings of the research.

Chapter 2

Wireless Transmitter Basics

2.1 Link Budget Analysis

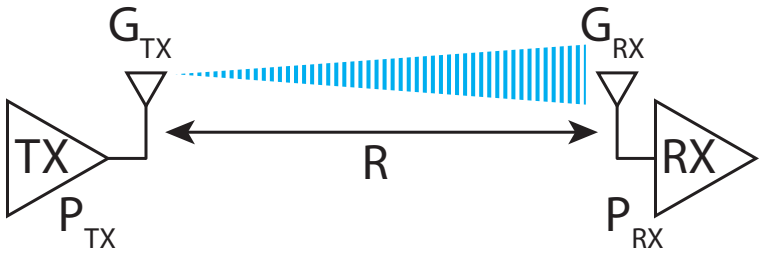


Figure 2.1: Wireless link budget analysis

The wireless radio design usually starts from the link budget analysis, which determines the specifications for individual blocks based on the application requirements such as distance and data rate. Fig. 2.1 shows a wireless link with a transmit to receive antenna distance of R. The transmitter delivers an output power of P_{TX}, and the antenna gains are G_{TX} and G_{RX} for the transmit and receive side respectively. According to the Friis transmission equation, the received signal power at the receiver input is,

$$P_{RX} = P_{TX} \times G_{TX} \times G_{RX} \times (\frac{\lambda}{4\pi R})^2 \tag{2.1}$$

Frequency (GHz)	FSPL (dB)		
	R=2m	R=10m	R=100m
0.9	38	52	72
5	52	66	86
60	74	88	108

Table 2.1: FSPL for different distances and frequencies

or in dB units,

$$P_{RX} = P_{TX} + G_{TX} + G_{RX} - 20\log(\frac{4\pi R}{\lambda}) \tag{2.2}$$

$$= EIRP + G_{RX} - FSPL \tag{2.3}$$

$$EIRP = P_{TX} + G_{TX} \tag{2.4}$$

$$FSPL = 20\log(\frac{4\pi R}{\lambda}) \tag{2.5}$$

where EIRP is the transmitter output power plus the transmitter antenna gain and the last term is the Free Space Path Loss (FSPL). FSPL describes the loss in signal power level due to line-of-sight propagation through free space. It assumes isotropic antennas on both the transmit and receive side. Table. 2.1 lists the FSPL for various link distances at three different carrier frequencies. One can already observe the challenge for mm-wave transmitter design here. The FSPL increase with the carrier frequency, e.g the FSPL at 900MHz for 10 meters is around 52dB, but it increases by 36dB when the carrier frequency goes up to 60GHz. To maintain the same transmission distance, the transmitter EIRP must be increased correspondingly.

The link specs are ultimately set by the receiver Signal to Noise Ratio (SNR) requirement. The SNR at the receiver output can be expressed as follows,

$$SNR_{RXo} = P_{TX} + G_{TX} + G_{RX} - FSPL - 10\log(kT) - 10\log(BW_{RX}) - NF_{RX} \tag{2.6}$$

where BW_{RX} is the receiver bandwidth and NF_{RX} is the receiver noise figure. From Eq. 2.6 one can see the second challenge for mm-wave transmitters. For the same received SNR, larger data rate links require larger receiver bandwidth and therefore require larger EIRP. Current mm-wave standards demand a radio bandwidth significantly larger than any existing WiFi or cellular standard. For example, the WiGig standard utilizes a RF bandwidth of 1.76GHz, which is 44 times bigger than the 40MHz bandwidth utilized by the 802.11n standard. As a result, mm-wave transmitters need to deliver larger EIRP for the same SNR level.

The minimum SNR requirement depends on the modulation scheme and Bit Error Rate (BER). The BER can be expressed as a function of the energy per bit to noise power spectral

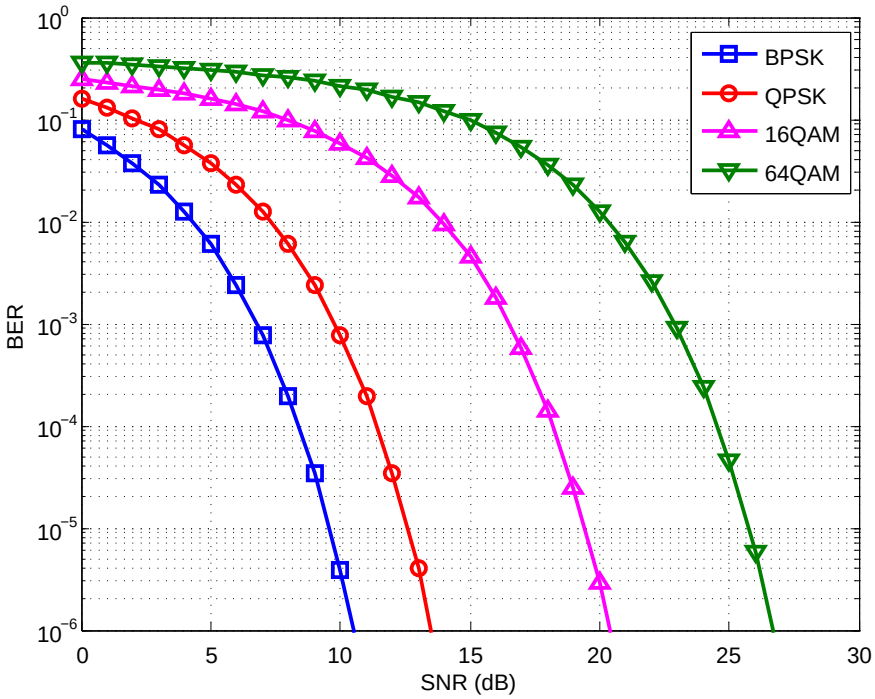


Figure 2.2: BER as a function of SNR

density ratio (E_b/N_0),

$$BER_{BPSK} = Q(\sqrt{\frac{2E_b}{N_0}}) \tag{2.7}$$

$$BER_{QPSK} = Q(\sqrt{\frac{2E_b}{N_0}}) \tag{2.8}$$

$$BER_{M-aray-QAM} = \frac{4}{k}(1 - \frac{1}{\sqrt{M}})Q(\sqrt{\frac{3k}{M-1}\frac{E_b}{N_0}}) \tag{2.9}$$

where M is the number of constellation points and k is the number of bits per symbol. E_b/N_0 can be further expressed in terms of SNR,

$$\frac{E_b}{N_0} = \frac{1}{k}\frac{E_s}{N_0} = \frac{1}{k}\frac{BW_{RX}}{f_s}SNR \tag{2.10}$$

The receiver RF bandwidth is usually designed to be roughly equal to the Nyquist symbol frequency and therefore Eq. 2.10 can be simplified to,

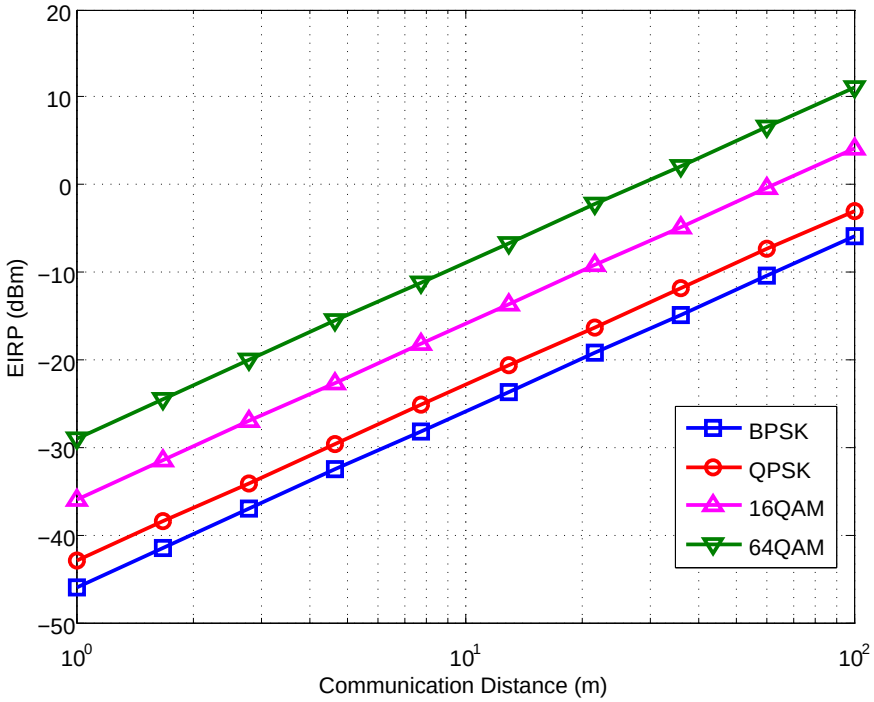
$$\frac{E_b}{N_0} = \frac{1}{k}SNR \tag{2.11}$$

Using this relation, the BERs in Eqn. 2.7-2.9 can be expressed in terms of SNR directly,

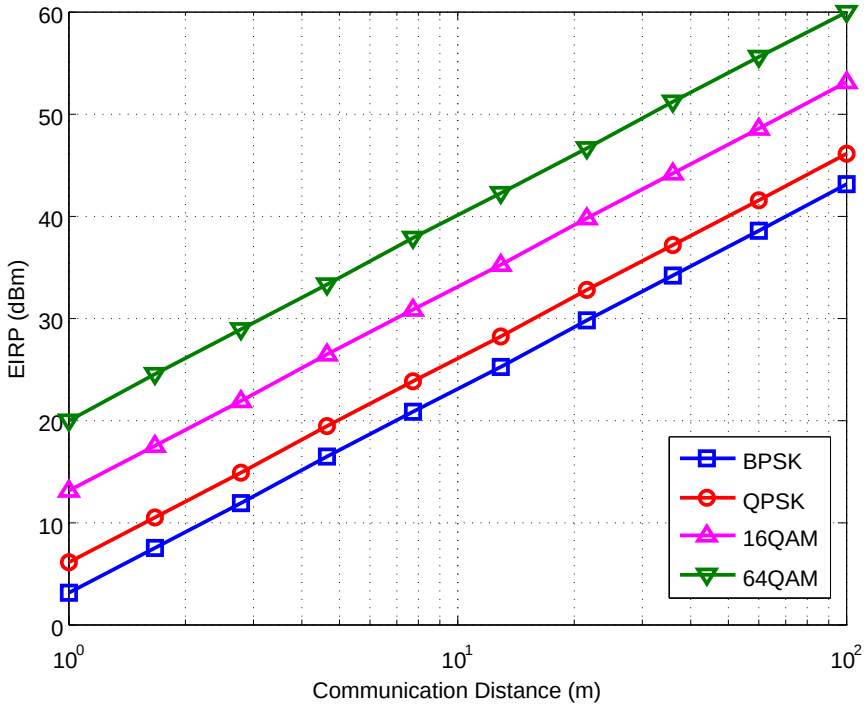
$$BER_{BPSK} = Q(\sqrt{2SNR}) \tag{2.12}$$

$$BER_{QPSK} = Q(\sqrt{SNR}) \tag{2.13}$$

$$BER_{M-aray-QAM} = \frac{4}{k}(1 - \frac{1}{\sqrt{M}})Q(\sqrt{\frac{3}{M-1}SNR}) \tag{2.14}$$



(a) 2.4GHz link with 40MHz RF bandwidth



(b) 60GHz link with 5GHz RF bandwidth

Figure 2.3: Required EIRP as a function of communication distance

Fig. 2.2 plots the BER as a function of SNR for BPSK, QPSK, 16QAM and 64QAM modulation schemes. In modern wireless systems, expected BER at the radio front-end output is around 10^{-3} to 10^{-4} [1], and therefore the minimum received SNR needs to be greater than 7dB for BPSK, 10dB for QPSK, 17dB for 16QAM and 23dB for 64QAM. As a result, with the same receiver, higher order modulation schemes require higher transmitter EIRP. Fig. 2.3 plots the required EIRP as a function of communication distance to achieve a BER of 10^{-3}, and compares a 2.4GHz link with 40MHz RF bandwidth and a 60GHz link with 5GHz bandwidth. In both cases, the receiver has an isotropic antenna ($G_{RX} = 0$dBi) and 5dB noise figure. Clearly, the 60GHz link requires much larger EIRP for covering the same distance, illustrating the aforementioned challenges. Note that the EIRPs shown in Fig. 2.3 represent the average values, which means transmitters need to handle peak EIRPs several dBs higher, depending on the PAPR of the modulation scheme.

2.2 Wireless Transmitter Architectures

Most modern wireless transmitters can be classified into two different categories [2]: the superheterodyne[3] architecture and the direction conversion architecture. Common in both architectures, there are four major sub-systems in the RF front-end: analog baseband, frequency generation, modulation and frequency conversion, and power amplification. The analog baseband first converts the coded baseband I/Q digital bits into continuous time analog signals through Digital-to-Analog-Converters (DACs), and subsequently filters the analog signals to reject unwanted high frequency spectral contents. The filtered I/Q signals are used to modulate a high frequency carrier signal known as the Local Oscillator (LO) signal. The modulated signal is amplified to achieve required power level by the PA. The high frequency LO signal is usually generated by a Phase-Lock-Loop (PLL) taking highly accurate frequency reference from a quartz crystal. In a superheterodyne transmitter (Fig. 2.4), the modulation takes place at an Intermediate Frequency (IF) which is lower than the final RF carrier frequency. Then the modulated signal is up-converted by a mixer. There are two major advantages: first, since the frequency generation block only needs to synthesize a lower frequency, it can achieve lower power consumption and better phase noise performance; second, modulation at IF is more power efficient and linear, and it also reduces the operating frequency of the calibration circuits[4]. However, superheterodyne transmitters need a band-pass filter after the up-conversion mixer since the mixer produces undesired spectral contents 2×IF frequency away from the RF carrier, which is known as the image. Image leaking through the transmitter not only reduces the desired signal power, but also potentially interferes with other communication links operating at the image frequency band.

[1] The errors are being corrected by equalizers and error decoders in the digital MODEM

[2] Except for non-coherence detection based modulation schemes such On-Off Keying (OOK) or Frequency Modulation (FM). These modulation schemes are less spectral efficient and the corresponding transmitter architecture is much simpler. In fact, they existed long before the CMOS radios.

[3] Sometimes simply referred to as hererodyne

[4] This is especially true for mm-wave transmitters where high frequency blocks need proper impedance matching and careful layout.

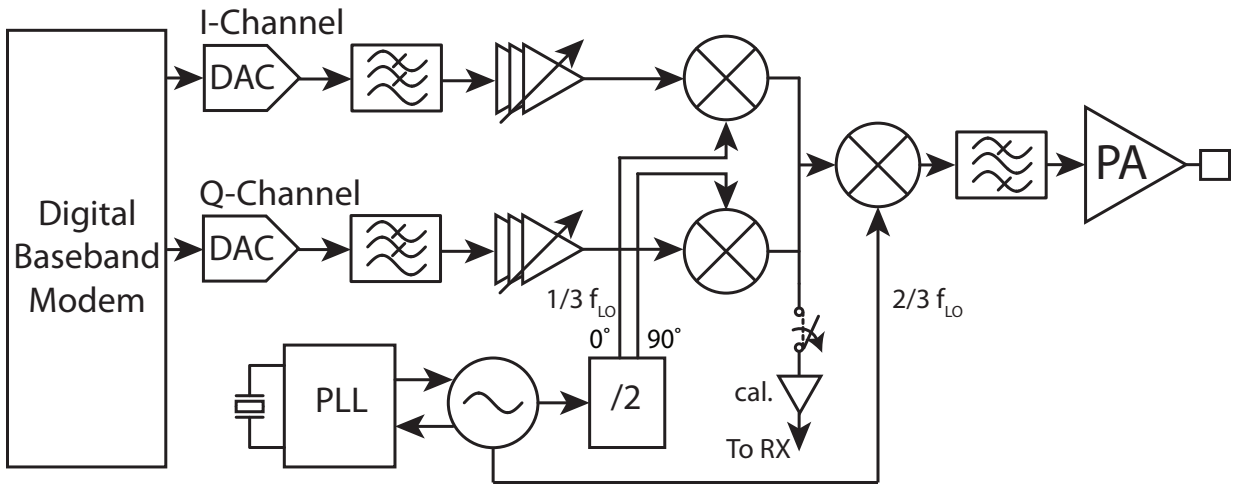


Figure 2.4: A sliding-IF superheterodyne transmitter

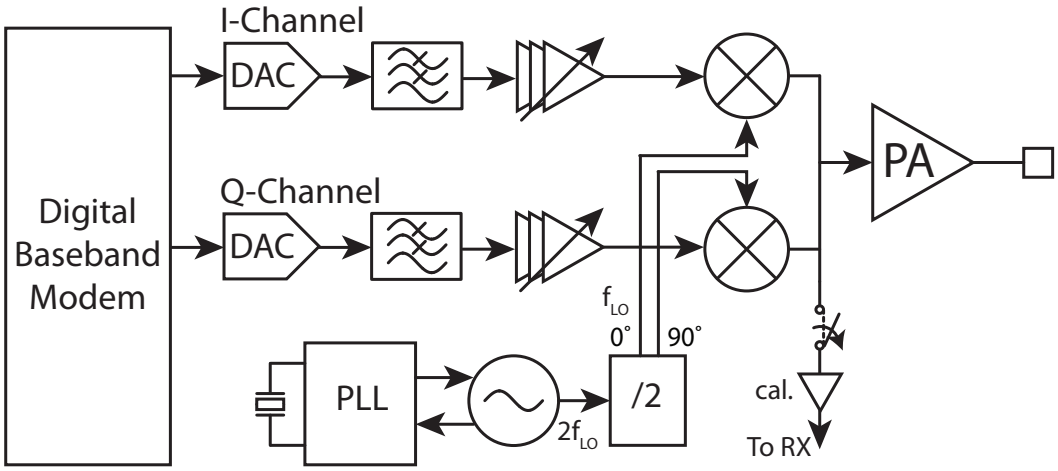


Figure 2.5: A direct conversion transmitter

To eliminate the image problem, direction conversion architecture can be used, in which the modulation happens at the final RF carrier frequency. This means the image is actual the signal itself, and therefore no filtering is required. However, one practical problem of the direction conversion transmitter is the VCO pulling. Since the PA output contains a wideband signal around the carrier and the amplitude level is usually very large (e.g. 20dBm output power with 50Ω load corresponds to 3.3V of swing), the VCO oscillation frequency may be dragged around by the parasitic feedback from the PA. A good layout that isolates the PA from the VCO reduces the pulling effect, however, it's almost impossible to completely block the feedback paths from the PA since they exist everywhere: supply coupling, substrate coupling and even reflection from the package! Pulling at close-in frequencies can be corrected by the PLL, but unfortunately the PLL usually has a much lower bandwidth compared to the data bandwidth, and therefore is not effective in reducing pulling at higher frequencies. A common solution is to synthesize a clock that is multiples of the desired LO frequency, as shown in Fig. 2.5. It's generally much more difficult to pull a VCO running at the multiples of the pulling signal frequency.

2.3 Power Amplifiers (The ABCDEFs)

The most power consuming block in a wireless transmitter is the power amplifier. In a WLAN transmitter, the PA usually contributes over 60% of the total power consumption [3]. As a result, the power efficiency of the PA is directly related to the battery life of mobile devices. To quantify the PA efficiency, several efficiency measures are used. The most widely used measures are the drain efficiency (η_D)[5] and the Power Added Efficiency (PAE). The drain efficiency is defined as the ratio between the useful output power P_{out} and the DC power supplied to the drain of the PA device P_{DC},

$$\eta_D = \frac{P_{out}}{P_{DC}} \tag{2.15}$$

The drain efficiency tells how much power is being dissipated during the DC to AC power conversion. The losses include power dissipation in the transistor as well as in the passive matching network. On the other hand, PAE takes into account of the additional power used to drive the PA device and is defined as the ratio between the added power $P_{out} - P_{in}$ and the DC power P_{DC},

$$PAE = \frac{P_{out} - P_{in}}{P_{DC}} \tag{2.16}$$

The PAE is related to η_D by the power gain of the PA G_P,

$$PAE = \eta_D(1 - \frac{1}{G_P}) \tag{2.17}$$

The significance of the PAE measure can be understood when analyzing a cascaded chain of similar amplifiers. Assume each amplifier has a drain efficiency of η_D and a power gain of

[5]It's also referred to as collector efficiency in bipolar PAs

G_P, the output power and the DC power of the n^{th} stage are,

$$P_{out}(n) = \frac{P_{out}}{G_p^{N-n}} \tag{2.18}$$

$$P_{DC}(n) = \frac{1}{\eta_D}\frac{P_{out}}{G_p^{N-n}} \tag{2.19}$$

Summing the DC power of each stage, the total DC power is,

$$P_{DC} = \sum_{n=1}^{N} P_{DC}(n) = \frac{P_{out}}{\eta_D}\frac{1-\frac{1}{G_P^N}}{1-\frac{1}{G_P}} \tag{2.20}$$

The cascaded PAE is,

$$PAE_{cascaded} = \frac{P_{out}(1-\frac{1}{G_P^N})}{\frac{P_{out}}{\eta_D}\frac{1-\frac{1}{G_P^N}}{1-\frac{1}{G_P}}} = \eta_D(1-\frac{1}{G_P}) = PAE \tag{2.21}$$

Therefore, the cascaded PAE is identical to the single-stage PAE. When N approaches a sufficiently large number, the input power becomes negligible and the PAE represents the total power efficiency of the amplifier chain.

Depending on the input-output relation, power amplifiers are usually categorized as linear amplifiers and non-linear amplifiers. Within each category, multiple classes are defined according to the voltage and current waveforms. The following subsections will describe these different classes.

2.3.1 Linear Classes (Class-A/B/AB)

Linear amplifiers produce an output signal that is an exactly scaled version of the input signal. In other words, the gain of a linear amplifier is constant, independent of the input signal level. All linear amplifiers are transconductance amplifiers, meaning the device operates as a current source with a transconductance gain of g_m. The simplest class that produces a linear behavior is Class-A. In Class-A amplifiers, the transistor is biased at a sufficiently large overdrive voltage so that it's conducting current all the time. The voltage and current waveforms at the drain node are shown in Fig. 2.6. The largest voltage swing obtainable is $V_{dd} - V_{ov}$ and the largest current swing obtainable is the DC bias current I_{dbias}, which is usually set to half of the maximum current I_{max} that the transistor can sink in order to maximize the linear current swing. As a result, the largest power obtainable is,

$$P_{out}^A = \frac{1}{2}V_{sw}I_{sw} = \frac{1}{4}(V_{dd} - V_{ov})I_{max} \tag{2.22}$$

Note that in order to achieve this maximum obtainable output power, an appropriate load impedance is needed to simultaneously maximize the voltage and current swings. This

optimal load impedance can be expressed as a ratio of the voltage swing and the current swing,

$$R_{opt}^{A} = 2\frac{V_{dd} - V_{ov}}{I_{max}} \tag{2.23}$$

When the load impedance is smaller than R_{opt}^{A}, the amplifier becomes current limited, meaning that there isn't enough current swing to generate the maximum voltage swing. Likewise, if the load impedance is larger than R_{opt}^{A}, the amplifier becomes voltage limited, meaning that only a fraction of the current swing is sufficient to saturate the voltage swing. In both cases, the output power decreases from P_{out}^{A}. The peak drain efficiency of Class-A amplifiers is obtained when R_{opt}^{A} is presented,

$$\eta_{D}^{A} = \frac{P_{out}^{A}}{P_{DC}} = \frac{\frac{(V_{dd}-V_{ov})I_{max}}{4}}{\frac{V_{dd}I_{max}}{2}} = \frac{V_{dd} - V_{ov}}{2V_{dd}} \approx 50\% \tag{2.24}$$

Class-A amplifiers have all the attractiveness except for the efficiency. 50% of the DC power is dissipated on the transistor when the product of drain voltage and current is non-zero. In other words, the transistor dissipates power whenever there's overlap between non-zero voltage and current waveforms. Therefore, it's obvious that in order to improve the efficiency, the overlapping period needs to be reduced.

A simple way to achieve this goal is to bias the transistor at the threshould voltage such that the transistor is conducting current during half of the cycle. Such amplifiers are known as Class-B amplifiers, and the drain node voltage and current waveforms are shown in Fig. 2.7. Since the transistor is only conducting current 50% of the time, the amount of V-I overlap is greatly reduced. The current going into the transistor drain becomes a half-wave rectified sine, with a fundamental component of $\frac{I_{max}}{2}$ and a DC value of $\frac{I_{max}}{\pi}$. The output power, optimal load impedance and the drain efficiency of Class-B amplifiers can be found in a similar way,

$$P_{out}^{B} = \frac{1}{2}V_{sw}I_{sw} = \frac{1}{4}(V_{dd} - V_{ov})I_{max} \tag{2.25}$$

$$R_{opt}^{B} = 2\frac{V_{dd} - V_{ov}}{I_{max}} \tag{2.26}$$

$$\eta_{D}^{B} = \frac{P_{out}^{B}}{P_{DC}} = \frac{\frac{(V_{dd}-V_{ov})I_{max}}{4}}{\frac{\pi}{V_{dd}I_{max}}} = \frac{\pi(V_{dd} - V_{ov})}{4V_{dd}} \approx 79\% \tag{2.27}$$

Compared to Class-A amplifiers, Class-B amplifier has much better peak drain efficiency, but maintains the same peak output power. The disadvantage is that the amplifier becomes slightly nonlinear, due to the varying effective bias condition at the input. However, the amplifier is fundamentally considered as a linear class due to the transconductance behavior of the device.

To improve the linearity of Class-B amplifiers, the gate bias of the transistor can be set slightly higher than the threshold voltage so that the transistor is conducting current more than half of the cycle, but still much less than the entire cycle. This sub-category is called Class-AB amplifiers. As a result, the efficiency will lie between Class-A and Class-B

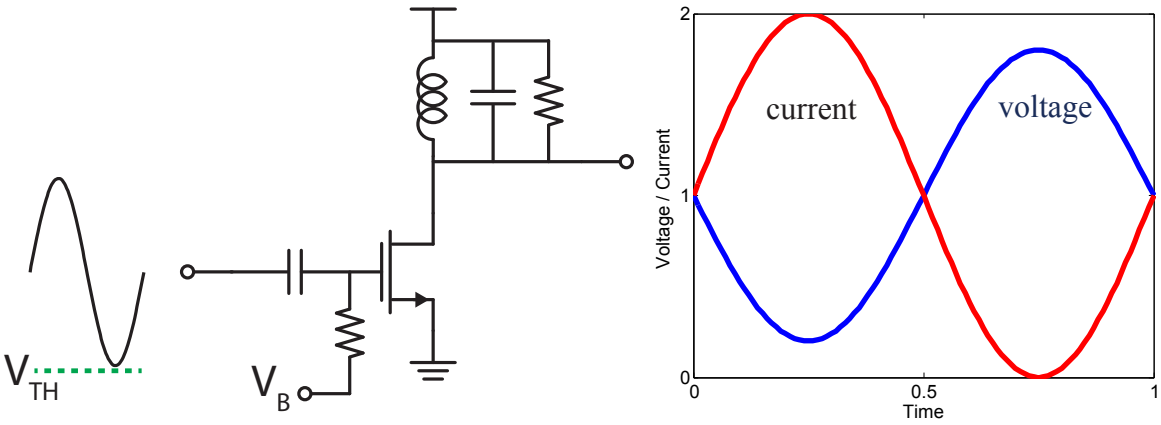


Figure 2.6: Schematics and V-I Waveforms of Class-A amplifiers

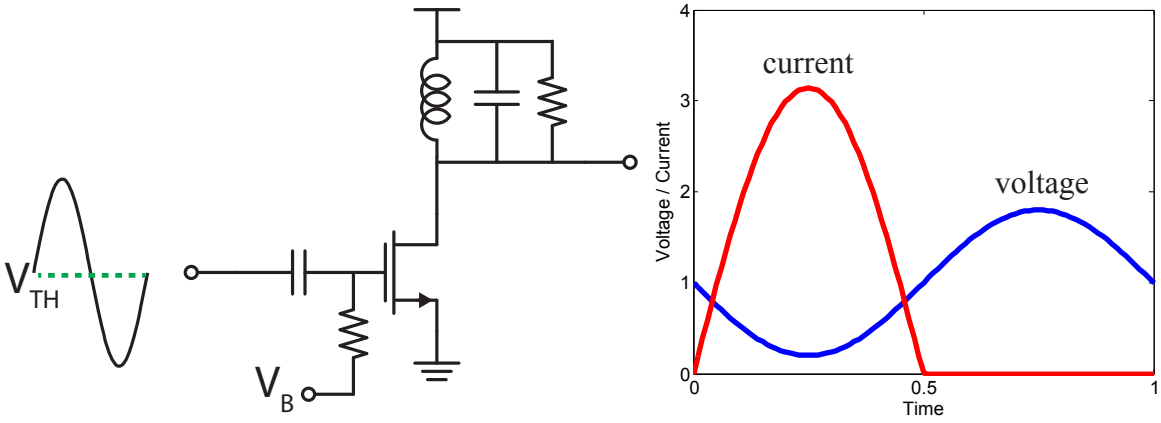


Figure 2.7: Schematics and V-I Waveforms of Class-B amplifiers

amplifiers. To find the power and efficiency of Class-AB amplifiers, the current swing and DC current are first expressed as a function of the conduction angle α, defined as the total number of radians in one cycle during which the transistor is conducting current,

$$I_{sw} = \frac{1}{2\pi}[\frac{\alpha - \sin\alpha}{1-\cos(\alpha/2)}]I_{max} \tag{2.28}$$

$$I_{DC} = \frac{1}{2\pi}[\frac{2\sin(\alpha/2) - \alpha\cos(\alpha/2)}{1-\cos(\alpha/2)}] \tag{2.29}$$

Next, the output power, optimal load impedance and the drain efficiency of Class-AB amplifiers can be found,

$$P_{out}^{AB} = \frac{1}{4\pi}[\frac{\alpha - \sin\alpha}{1-\cos(\alpha/2)}](V_{dd} - V_{ov})I_{max} \tag{2.30}$$

$$R_{opt}^{AB} = 2\pi[\frac{1-\cos(\alpha/2)}{\alpha - \sin\alpha}]\frac{V_{dd} - V_{ov}}{I_{max}} \tag{2.31}$$

$$\eta_D^{AB} = \frac{P_{out}^{AB}}{P_{DC}} = \frac{1}{2}\frac{V_{dd} - V_{ov}}{V_{dd}}\frac{\alpha - \sin\alpha}{2\sin(\alpha/2) - \alpha\cos(\alpha/2)} \tag{2.32}$$

$$\alpha \subset [\pi, 2\pi] \tag{2.33}$$

2.3.2 Non-Linear Classes (Class-C/D/E/F)

Unlike linear amplifiers, non-linear amplifiers lack of amplitude linearity. They usually produce significant amplitude-to-amplitude (AM-AM) and amplitude-to-phase (AM-PM) distortions. As a result, they cannot be directly used for convey amplitude modulated signals. However, most non-linear amplifiers are still linear in phase, and have no phase-to-phase (PM-PM) distortions. Therefore, they are often used for amplifying phase modulated signals such as GMSK signals. The biggest advantage of most non-linear amplifiers is the high drain efficiency. A linear transconductance amplifier can be turned into a non-linear amplifier by biasing the transistor below threshold voltage. Such amplifiers are known as Class-C amplifiers (Fig. 2.8). Class-C amplifiers conduct current less than half of the cycle, and have smaller window of V-I overlap. Therefore, the efficiency of Class-C amplifiers is even higher than that of Class-B amplifiers. The expression for output power, optimal load impedance and efficiency of Class-C amplifiers is the same as Class-AB amplifiers, but the conduction angle is smaller than π.

$$P_{out}^{C} = \frac{1}{4\pi}[\frac{\alpha - \sin\alpha}{1-\cos(\alpha/2)}](V_{dd} - V_{ov})I_{max} \tag{2.34}$$

$$R_{opt}^{C} = 2\pi[\frac{1-\cos(\alpha/2)}{\alpha - \sin\alpha}]\frac{V_{dd} - V_{ov}}{I_{max}} \tag{2.35}$$

$$\eta_D^{C} = \frac{P_{out}^{AB}}{P_{DC}} = \frac{1}{2}\frac{V_{dd} - V_{ov}}{V_{dd}}\frac{\alpha - \sin\alpha}{2\sin(\alpha/2) - \alpha\cos(\alpha/2)} \tag{2.36}$$

$$\alpha \subset [0, \pi] \tag{2.37}$$

In theory, Class-C amplifiers can have drain efficiency as high as 100%. However, this is achieved at zero conduction angel, which means the output power also drops to zero according to Eq. 2.34. Unlike Class-AB amplifiers, Class-C amplifier has a clear tradeoff between output power and drain efficiency.

Due to the low output power level, Class-C amplifiers are seldom used as a stand-alone amplifier unit[6]. More widely used non-linear amplifiers are switching amplifiers. In switching amplifiers, transistors behave like switches, and it's either in triode region or in cut-off region. This is in contrast with transconductance amplifiers in which transistors remain in saturation region.

A switching amplifier can be constructed with an inverter which produces a square voltage waveform. In order to extract the fundamental component, a series LC band-pass filter can be added at the output. Such amplifiers are known as Class-D amplifiers (Fig. 2.9). Since only fundamental frequency current can flow through the filter, each of the two transistors contribute half cycle of the sine current waveform. Since there's no V-I overlap, the theoretical peak drain efficiency is 100%. The maximum voltage and current swings are,

$$V_{sw} = \frac{2}{\pi} V_{dd} \tag{2.38}$$

$$I_{sw} = I_{max} \tag{2.39}$$

Therefore the output power, optimal load impedance and the drain efficiency can be found,

$$P_{out}^{D} = \frac{1}{\pi} V_{dd} I_{max} \tag{2.40}$$

$$R_{opt}^{D} = \frac{V_{sw}}{I_{sw}} = \frac{2}{\pi} \frac{V_{dd}}{I_{max}} \tag{2.41}$$

$$\eta_{D}^{D} = \frac{V_{sw} I_{sw}}{2 V_{DC} I_{DC}} \approx \frac{\frac{1}{\pi} V_{dd} I_{max}}{V_{dd} \frac{1}{\pi} I_{max}} = 100\% \tag{2.42}$$

Note that the peak drain efficiency can only be achieved at very low frequency, where the transistor capacitance is negligible. Unfortunately, the power spent charging the transistor drain node parasitic capacitance increases linearly with frequency. Besides this charging power loss, the parasitic capacitor also smooths the edges of the square waveform, and thereby introduces finite V-I overlap. As a consequence, Class-D amplifiers are seldom used for high frequency designs due to dramatically reduced efficiency.

Class-E amplifiers can absorb the transistor parasitic capacitance into an impedance tuning network while ensure non-overlapping V-I waveforms (Fig. 2.10)[4, 5]. When the switch is off, the current flowing into the drain is zero while the drain voltage is non-zero. The voltage waveform reaches zero right before the switch is turned on, after which the drain voltage remains zero when the transistor sinks current. Such transition behavior is called Zero Voltage Switching (ZVS). ZVS not only ensures non-overlapping V-I waveforms, but also avoids charge loss when the switch turns on. In fact, Class-E amplifiers satisfy not only ZVS but also Zero derivative Voltage Switching (ZdVS), meaning the derivative of the

[6]It's often used in conjunction with other amplifiers for efficiency enhancement, such as Doherty amplifiers

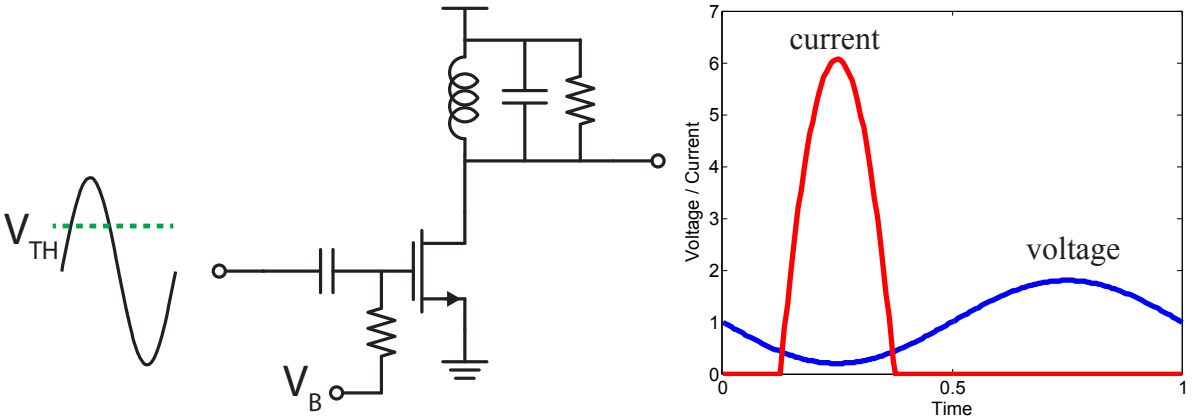


Figure 2.8: Schematics and V-I Waveforms of Class-C amplifiers

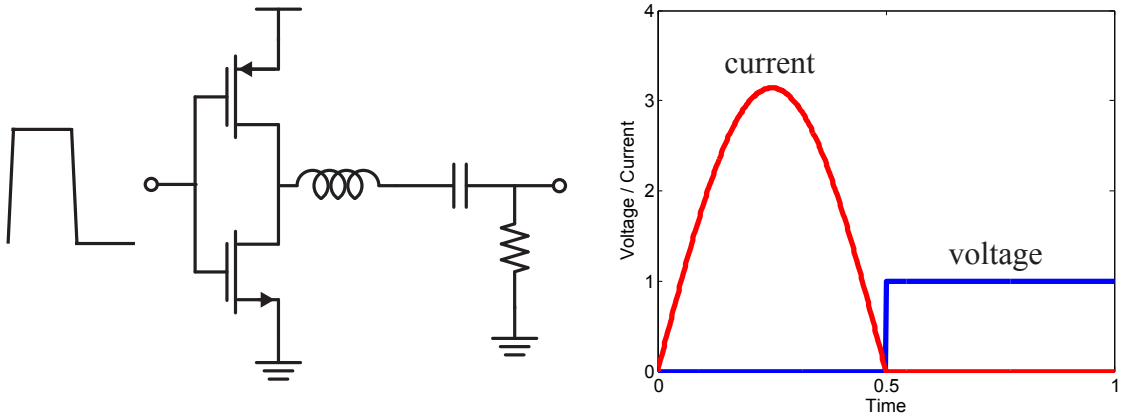


Figure 2.9: Schematics and V-I Waveforms of Class-D amplifiers

voltage waveform at the time instant when the switch turns on is also zero. This property significantly reduces the efficiency sensitivity due to the passive component variations. The typical Class-E implementation is shown in Fig. 2.10. A series LC filter is used so that the load impedance is only seen at the fundamental frequency. As a result, the transistor sees inductive load impedance at the fundamental frequency while it only sees the parasitic drain capacitor at all harmonics. An inductor choke is used to provide the DC current path. The derivation for the required tuning impedances is skipped here, since it's given in multiple literatures [6, 7, 8].

$$P_{out}^{E} \approx V_{DC} I_{DC} = \frac{1}{F_{PI}} V_{dd} I_{max} = \frac{1}{2.86} V_{dd} I_{max} \tag{2.43}$$

$$Z_{C,opt}^{E} = F_{PI} F_{C} \frac{V_{dd}}{I_{max}} = 8.985 \frac{V_{dd}}{I_{max}} \tag{2.44}$$

$$R_{opt}^{E} = 0.1836 Z_{C,opt}^{E} = 1.65 \frac{V_{dd}}{I_{max}} \tag{2.45}$$

$$X_{opt}^{E} = 1.152 R_{opt}^{E} = 1.9 \frac{V_{dd}}{I_{max}} \tag{2.46}$$

$$F_{PI} = 2.86 \tag{2.47}$$

$$F_{C} = \pi \tag{2.48}$$

$$\eta_{D}^{E} \approx 100\% \tag{2.49}$$

The final amplifier class in alphabetical order is Class-F. Class-F amplifiers are constructed based on Class-B amplifiers. In order to further improve the efficiency of Class-B amplifiers beyond 79% while not sacrificing the output power, odd harmonics can be added to the voltage waveform. The addition of the odd harmonics shapes the voltage waveform from a sine to a square, and thus reducing the V-I overlap. By increasing the number of harmonics, the efficiency can approach 100% eventually. To implement this idea, a bank of filters need to be inserted in series at the amplifier output to present Open Circuit (O.C.) impedance to the transistor at odd harmonics to enrich the odd harmonic content in the drain voltage waveform while a low-pass filter is needed at the output to eliminate all the even order harmonics (Fig. 2.11). When sufficient harmonics are added, the waveform resembles that of Class-D amplifiers. The output power, optimal impedance and efficiency of Class-F amplifiers are,

$$P_{out}^{F} = \frac{1}{2} \frac{4}{\pi} V_{dd} \frac{1}{2} I_{max} = \frac{1}{\pi} V_{dd} I_{max} \tag{2.50}$$

$$R_{opt}^{F} = \frac{V_{sw}}{I_{sw}} = \frac{8}{\pi} \frac{V_{dd}}{I_{max}} \tag{2.51}$$

$$\eta_{D}^{F} = \frac{V_{sw} I_{sw}}{2 V_{DC} I_{DC}} \approx \frac{\frac{1}{\pi} V_{dd} I_{max}}{V_{dd} \frac{1}{\pi} I_{max}} = 100\% \tag{2.52}$$

Similar to Class-D amplifiers, the efficiency of Class-F amplifiers is also degraded by the parasitic capacitance at the transistor drain node. In addition, it requires large number of filters for harmonic generation which introduces insertion loss. As a result, the beauty of Class-F amplifiers is usually lost in implementation.

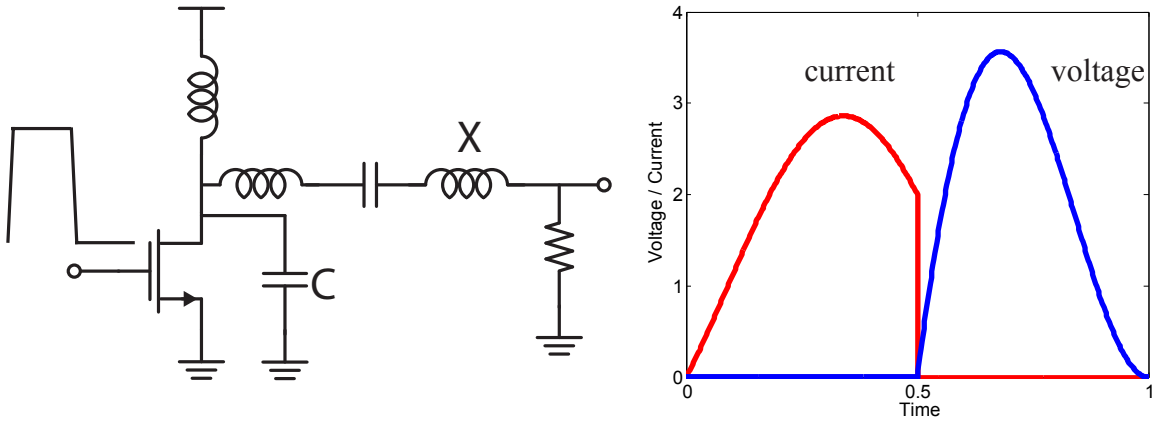


Figure 2.10: Schematics and V-I Waveforms of Class-E amplifiers

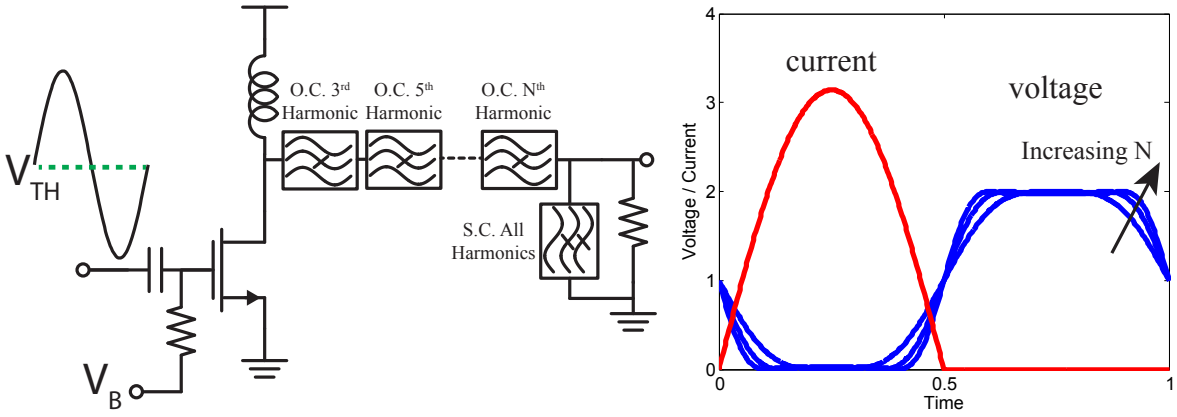


Figure 2.11: Schematics and V-I Waveforms of Class-F amplifiers

2.4 The Extended Class-E/F Family

As discussed in the previous section, Class-E amplifiers have ZVS waveform characteristic. In fact, there are numerous impedance tuning networks which can produce the ZVS waveform. It was not until recently that the relation between different impedance tuning networks was well discovered. In [8], it's summarized that many of the ZVS tuning networks belong to the Class-E/F family. Different members of the family correspond to different harmonic impedances, with Class-E and Class-F^{-1} being the two extremes: Class-E has no harmonic impedance tuning while Class-F^{-1} has impedance tuning at all harmonics. The essential rule of the Class-E/F harmonic tunings in [8] is as follows: the effective load impedance needs to be open circuit at even harmonics but short circuit at odd harmonics. Table. 2.2 shows examples of some Class-E/F tunings in terms of harmonic impedance.

It's also shown in [8] that the power, efficiency and gain of a Class-E/F amplifier can be expressed in terms of a set of technology dependent parameters such as transistor resistance and capacitance, defined in Table. 2.3, and also a set of technology independent parameters called the waveform figures of merit. The waveform figures of merit are determined by the combination of the harmonic load impedances and are unique to each tuning class. They are defined as follows,

$$F_V = \frac{V_{pk}}{V_{DC}} \tag{2.53}$$

$$F_I = \frac{I_{RMS}}{I_{DC}} \tag{2.54}$$

$$F_{PI} = \frac{I_{pk}}{I_{DC}} \tag{2.55}$$

$$F_C = \frac{V_{DC} I_{DC}}{V_{DC}^2 / Z_C} \tag{2.56}$$

F_V is the ratio between the peak and DC voltages. F_I is the ratio between the RMS and DC current. F_{PI} is the ratio between the peak and DC current. F_C is the ration between the DC power and the reactive power stored on the drain capacitor. Under the condition that the transistor is sized to be largest possible given the capacitance constraint, the drain efficiency, power gain and PAE can be expressed as follows,

$$\eta_D^{EF} = 1 - (F_I^2 F_C) 2\pi f_0 (\overline{r_{on} c_{out}}) \tag{2.57}$$

$$G^{EF} = \eta_D^{EF} (\frac{\overline{c_{out}} V_{bk}^2}{\overline{p_{in}}}) (\frac{F_C}{F_V^2}) 2\pi f_0 \tag{2.58}$$

$$PAE^{EF} = (1 - \frac{1}{G^{EF}}) \eta_D^{EF} \tag{2.59}$$

Note that here the loss due to the transistor finite on-resistance is taken into account.

The original Class-E/F family in [8] only considers open or short load for harmonic terminations, however, it turns out ZVS waveform can be achieved using arbitrary harmonic

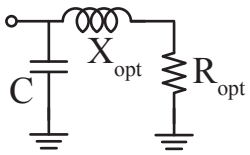
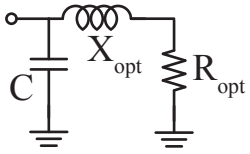
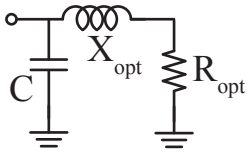
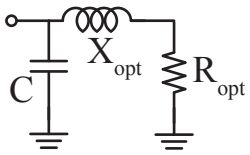
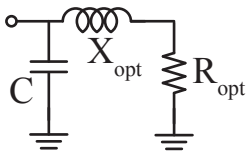
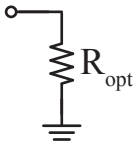
Tuning	f_0	$2f_0$	$3f_0$	$4f_0$	$5f_0$	$6f_0$	$7f_0$
E							
E/F_2		Open					
E/F_3			Short				
$E/F_{2,3}$		Open	Short				
$E/F_{2,3,4,5}$		Open	Short	Open	Short		
F^{-1}		Open	Short	Open	Short	Open	Short

Table 2.2: The harmonic impedance specifications of several original Class-E/F tunings

V_{bk}	Breakdown voltage (V)
$\overline{r_{on}}$	normalized transistor on-resistance (Ωm)
$\overline{c_{out}}$	normalized transistor output capacitance (F/m)
$\overline{c_{in}}$	normalized transistor input capacitance (F/m)
$\overline{p_{in}}$	normalized input driving power (W/m)

Table 2.3: Technology dependent parameters

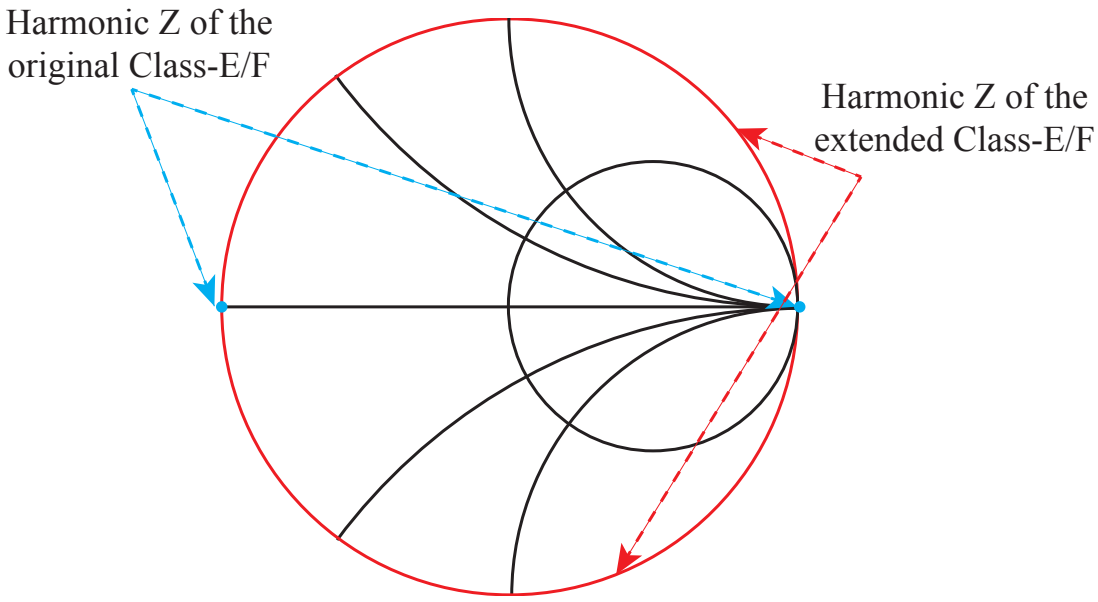


Figure 2.12: Harmonic impedance region of the extended Class-E/F family on the Smith Chart

terminations. Resistive termination at harmonics is undesired since it introduces additional harmonic resistance loss, therefore only purely reactive harmonic impedance is considered. In other words, the harmonic impedance of the extended Class-E/F family includes not only the O.C. and S.C. points on the Smith Chart, but also the entire perimeter, as shown in Fig. 2.12. Table. 2.4 shows several examples of the extended Class-E/F tunings with their corresponding naming conventions.

The extended Class-E/F tunings provide much larger design flexibility for performance optimization, especially when making tradeoffs between gain and drain efficiency. In order to see the impact of harmonic impedance on the amplifier performance, consider the Class-E/F$_{X2}$ tuning with inductive second harmonic termination of X_2. X_2 is the second harmonic load reactance (in parallel with the transistor drain capacitor, as shown in Table. 2.4) normalized to the fundamental impedance of the drain capacitor ($\frac{1}{\omega_0 C}$). Fig. 2.13 plots the V-I waveform of the Class-E/F$_{X2}$ with various X_2 values. The amplifier starts as a Class-E am-

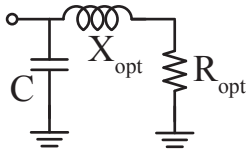
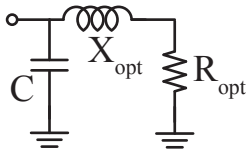
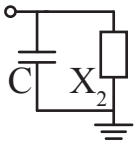
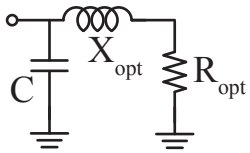
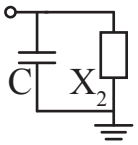
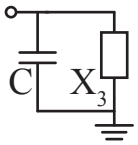
Tuning	f_0	$2f_0$	$3f_0$	$4f_0$	$5f_0$	$6f_0$	$7f_0$
E							
E/F_{X2}							
$E/F_{X2,X3}$							

Table 2.4: The harmonic impedance specifications of several extended Class-E/F tunings

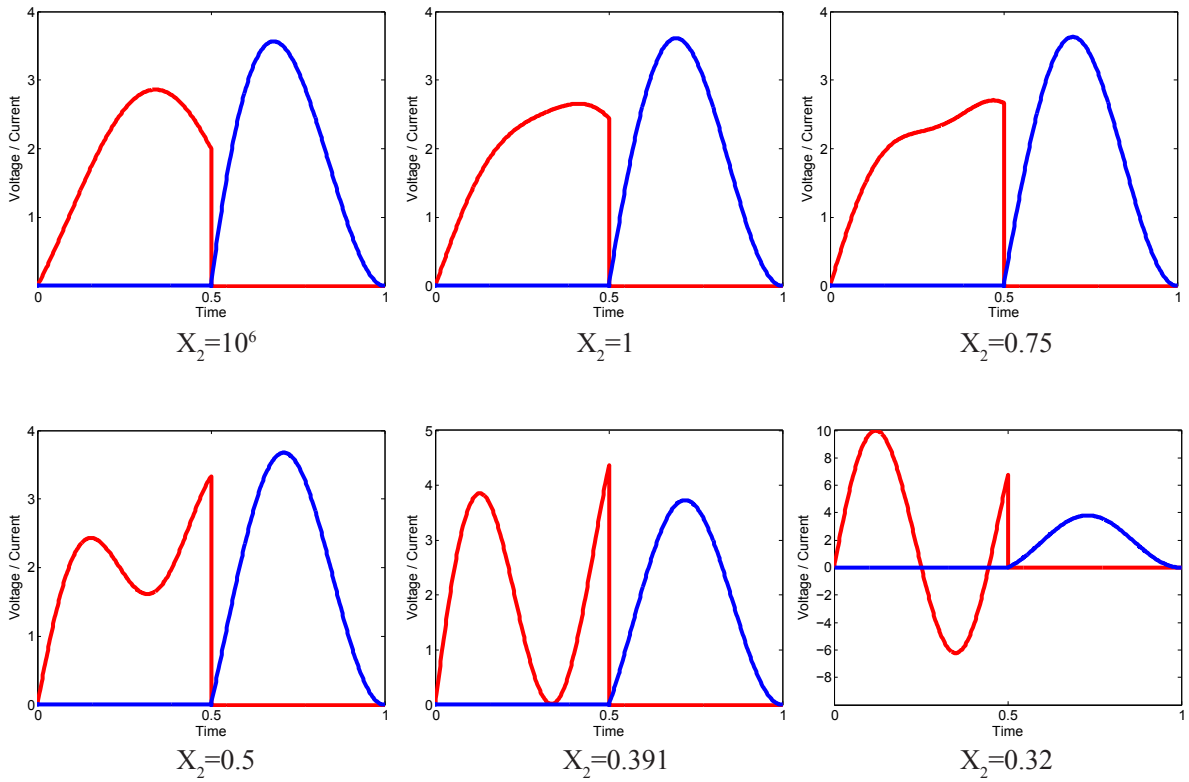


Figure 2.13: V-I Waveforms of the Class-E/F$_{X2}$ tunings with various X_2 values

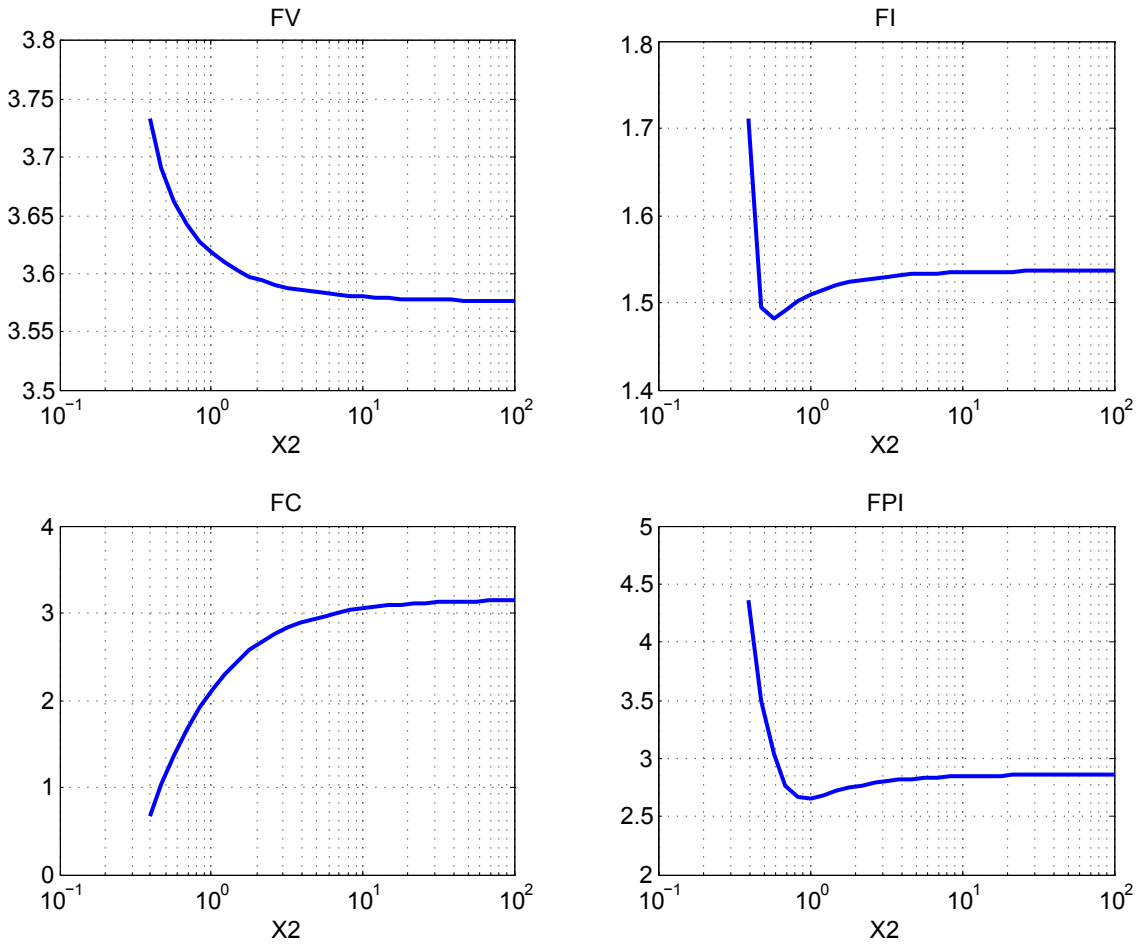
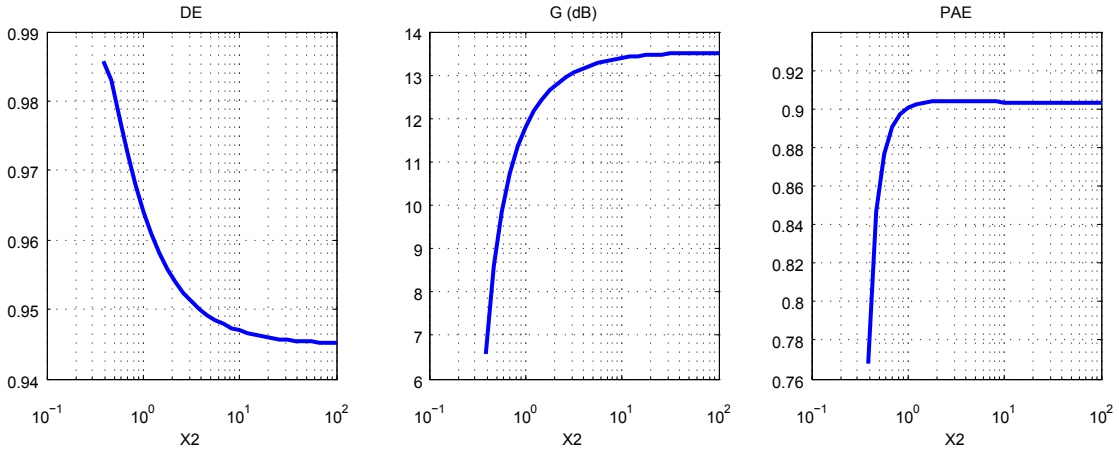
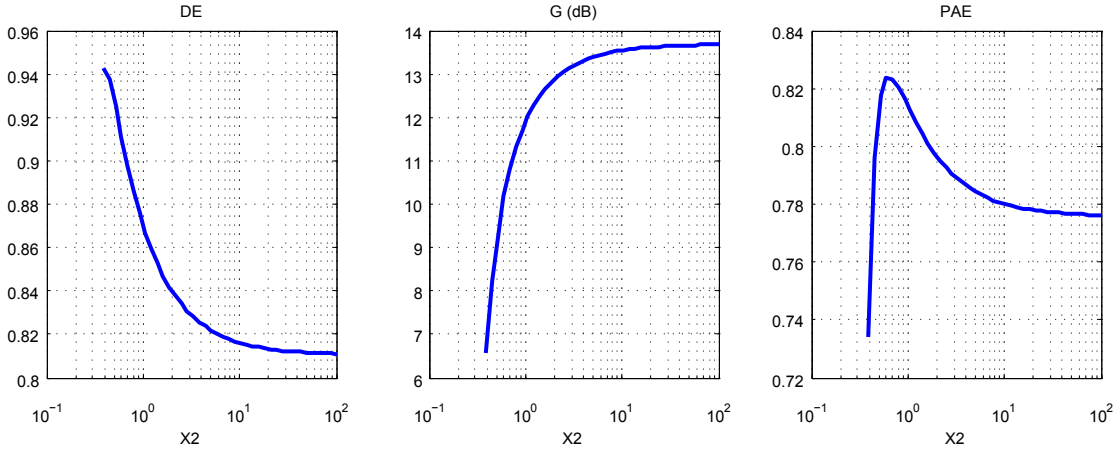


Figure 2.14: Waveform FoM of Class-E/F$_{X2}$ tunings as a function of X_2

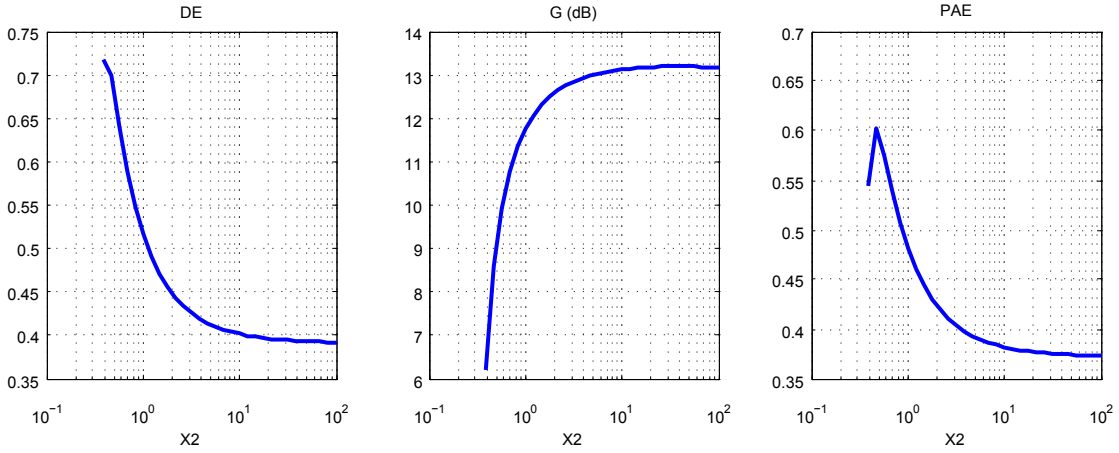
plifier when X_2 is infinity. When X_2 decreases, stronger second harmonic content is present in the current waveform. When X_2 reaches 0.5, the amplifier becomes Class-E/F$_2$, since the load reactance cancels out the reactance of the drain capacitor, presenting an overall O.C. at the second harmonic. Further decreasing X_2 keeps enriching the second harmonic content until the minimum instantaneous current reaches zero. Since it's undesired to have reverse direction current through the transistor, there's a lower limit on the X_2 value. For Class-E/F$_{X2}$, the minimum X_2 is 0.391. Based on V-I waveforms, the four waveform figures of merit can be calculated. Fig. 2.14 plots F_V, F_I, F_C and F_{PI} as a function of X_2. When X_2 decreases from infinity (no second harmonic tuning), F_V increases while F_C decreases, meaning larger peak to average voltage ratio and larger drain capacitance tolerance. Low F_C is desired from drain efficiency point of view since larger transistor size can be used to reduce the on resistance loss. On the other hand, the behavior of both F_I and F_{PI} is non-monotonic. Both F_I and F_{PI} decreases initially until it reaches the minimum value and they start to climb back up. The minimum is obtained when X_2 is 0.5. Since low F_I is desired for



(a) $f_0 = 2.4GHz$



(b) $f_0 = 10GHz$



(c) $f_0 = 60GHz$

Figure 2.15: η_D, G_p and PAE of Class-E/F$_{X2}$ amplifiers as a function of X_2

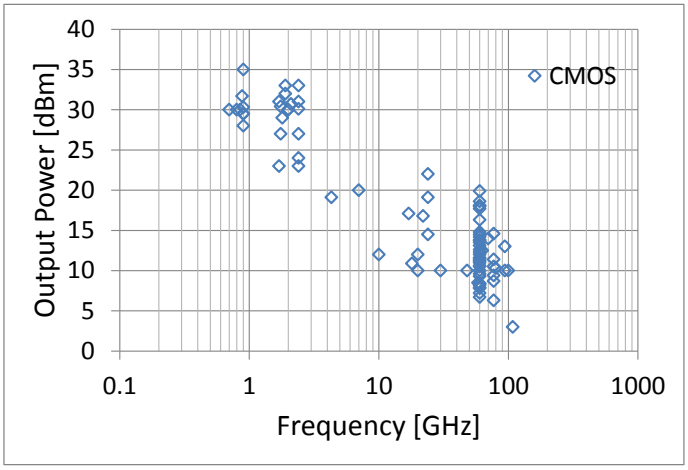
efficiency, $X_2 = 0.5$ gives the best F_I value. Note that with $X_2 = 0.5$, the tuning becomes Class-E/F$_2$.

Based on a typical 65nm CMOS technology, the drain efficiency, power gain and PAE of Class-E/F$_{X2}$ amplifiers can be plotted as a function of X_2 for different operating frequencies, as shown in Fig. 2.15. There's a clear tradeoff between drain efficiency and power gain: decreasing X_2 decreases the product $F_I^2 F_C$ and therefore improves the drain efficiency; but due to decreased F_C, or equivalently increased transistor size, larger input power is needed which leads to reduced power gain. The optimal PAE point depends on not only technology parameters, but also the operating frequency. At 2.4GHz, the optimal PAE is obtained when X_2 is roughly 2. At 10GHz, the optimal PAE is obtained when X_2 is around 0.6. Finally at 60GHz, the optimal PAE is obtained when X_2 is around 0.45. Intuitively, the decreasing optimal X_2 value is caused by the fact that η_D drops with increasing frequency, and therefore larger transistor sizes are needed to reduce the on resistance loss, which means lower F_C value is desired.

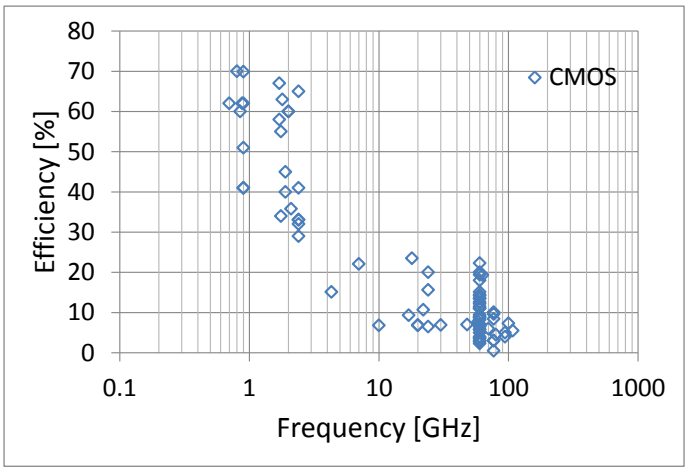
2.5 High Frequency Challenges

CMOS technology scaling has dramatically improved the transistor speed, as a result, significant progress has been achieved in integrating many of the mm-wave circuit blocks. However, mm-wave power amplifiers still face many design challenges. Fig. 2.16 shows a performance survey of the state-of-the-art CMOS PA over a wide frequency range. A clear trend can be observed here: both output power and efficiency drop with increasing operating frequency. This trend can be explained by multiple reasons. Since the product of transistor f_t and breakdown voltage is kept roughly constant during technology scaling, the breakdown voltage reduces as f_t improves. The thin oxide breakdown voltage is around 15MV/cm, and with oxide thickness in the order of 1nm, the breakdown voltage is usually around 1V. This means all the mm-wave devices need to operate from a very low supply voltage, which directly affect the PA output power since power is proportional to the voltage square. To deliver more power with the same supply voltage, the transistor need to deliver more current and therefore it needs to see a smaller load impedance. This means an impedance transformation network is needed at the PA output. However, the insertion loss of the matching network is inversely proportional to the impedance transformation ratio, and larger desired output power translates into larger insertion loss in the matching network. The result is that with a simple LC impedance transformation network, there is a limit on the obtainable output power [9]. To make things worse, technology scaling not only shrinks the copper metal thickness but also reduce the distance between top metal layer and the silicon substrate, thus reducing the Quality factor (Q factor) of on-chip passives. It's also shown in [9] that the maximum obtainable output power is proportional to the square of the Q factor.

The main factor that contributes to the reduced efficiency at high frequency is the reduced transistor power gain. For Class-A/AB amplifiers, the drain efficiency of mm-wave PA is in fact very comparable to their low frequency counterpart, however, since the Maximum



(a) Output Power



(b) Efficiency

Figure 2.16: A survey of CMOS PA performance

Stable Gain (MSG) drops with increasing frequency, more power is needed to drive the PA. At 60GHz for example, a single stage PA can only provide 5-6dB power gain after including the passive loss. According to Eq. 2.16, the PAE can be much lower than the drain efficiency due to the low power gain. Driver stages are usually used to boost the overall gain of the PA chain, but it's also shown in Eq. 2.21 that cascading gain stages do not affect the PAE[7].

[7]In practice, adding driver stages slightly improves the overall PAE since the drive stages are impedance matched for maximum power gain, and therefore exhibit a different η_D and G_p profile than the last PA stage.

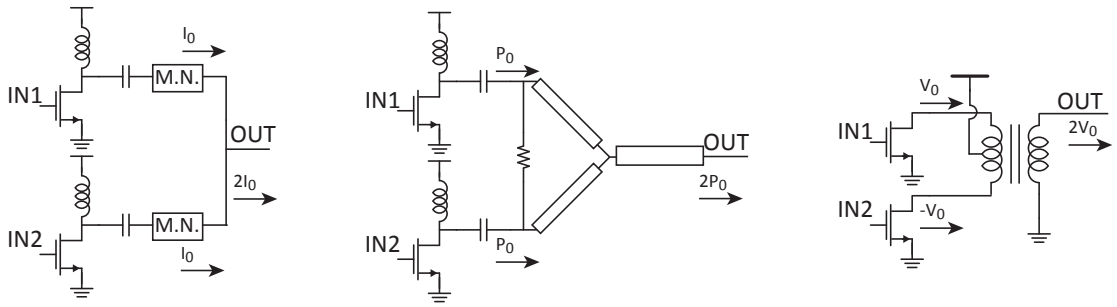
Chapter 3

Linear Transmitter

Most higher order single carrier modulation schemes such as QAM require a linear analog front-end to preserve the relative amplitude information. Even seemingly constant envelope signals such as Nyquist rate BPSK has a non-zero PAPR after low-pass filtering. As a result, linear transmitters are widely used today. To cover longer communication distance, larger PA output power is desired to increase the transmitter EIRP. In addition, Orthogonal Frequency Division Multiplexing (OFDM) technique is widely used in wireless communication today to mitigate the multi-path effect. One downside of such technique is the large signal PAPR. With increased signal PAPR, the peak output power level must be boosted accordingly, presenting a further challenge. This chapter discusses techniques for increasing the output power level of the linear class PAs while maximizing the efficiency.

3.1 Power Combining Techniques

As shown in [9], with a LC type of impedance transformation network, the output power of a single transistor PA cannot be increased indefinitely due to increased insertion loss. Therefore the only way to further increase the output power is to combine power from multiple transistors through a power combiner. A very intuitive idea is to use direct current summing (Fig. 3.1a). The advantage of such an approach is simple in implementation and low insertion loss. However, since each transistor sees N times higher impedance than the load where N is the number of combined branches, additional impedance matching network is needed in order to increase the total output power. As a result, the maximum obtainable output power stays roughly the same as a single transistor PA. The only benefit is that since each transistor is much smaller, transistor internal wiring loss can be reduced. Current summing based mm-wave PAs have shown limited output power, typically below 14dBm at 60GHz [10, 11]. Another well-known approach is Wilkinson combiners. Unlike



(a) Current summing (b) Wilkinsons (c) Transformers

Figure 3.1: On-chip power combining techniques

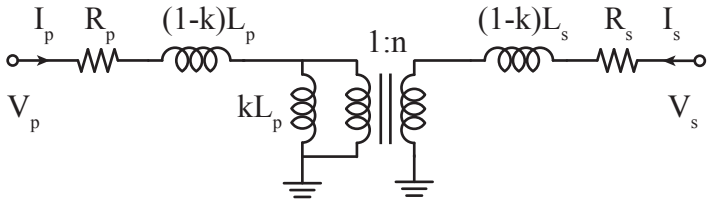


Figure 3.2: Equivalent circuit model of a transformer

direct current summing, the impedance of each Wilkinson input port remains 50Ω, therefore the total output power can be doubled without the need of additional matching network. Combining power from more than two transistors require cascading Wilkinson combiners. This results in a linear-in-dB increase in the passive loss after the PA, which reduces the efficiency and limits the maximum obtainable output power level. An alternative is to use corporate structures [12] which has smaller loss than the cascaded implementation. However, the required characteristic impedance of each transmission line (T-line) increases with the number of branches in the corporate structure, which impose a limit on the maximum number of branches given the finite range of on-chip T-lines impedance. Another disadvantage of Wilkinson combiners is the large size, which is proportional to the wavelength of the carrier signal. Even with meandered T-lines, a single Wilkinson combiner still occupies more than $200\mu m$ by $200\mu m$. As a result, Wilkinson combiner based PAs are typically very bulky [13].

The most popular power combining approach is based on transformers (Fig. 3.1c). On-chip transformers are constructed by two vertically or horizontally coupled inductors. The equivalent circuit model of a transformer is shown in Fig. 3.2. The voltage and current at the primary and secondary ports are related by,

$$\begin{bmatrix} V_p \\ V_s \end{bmatrix} = \begin{bmatrix} j\omega L_p + R_p & -j\omega M \\ j\omega M & -j\omega L_s - R_s \end{bmatrix} \begin{bmatrix} I_p \\ I_s \end{bmatrix} \tag{3.1}$$

$$M = k\sqrt{L_p L_s} \tag{3.2}$$

$$n = \sqrt{\frac{L_s}{L_p}} \tag{3.3}$$

The Q-factors of the primary and secondary inductors are defined as,

$$Q_p = \frac{\omega L_p}{R_p} \tag{3.4}$$

$$Q_s = \frac{\omega L_s}{R_s} \tag{3.5}$$

It's proven in [9, 14] that the maximum efficiency of a transformer only depends on Q_p, Q_s and k,

$$\eta_{max} = \frac{1}{1 + \frac{2}{Q_p Q_s k^2} + 2\sqrt{\frac{1}{Q_p Q_s k^2}(1 + \frac{1}{Q_p Q_s k^2})}} \tag{3.6}$$

This maximum efficiency is obtained when the following condition is met,

$$\omega L_s = \frac{1}{\omega C_{L,series}} \tag{3.7}$$

$$\omega L_p = \frac{\alpha}{1+\alpha^2}\frac{R_L}{n^2} \tag{3.8}$$

$$\alpha = \frac{1}{\sqrt{\frac{1}{Q_s^2} + \frac{Q_p}{Q_s}k^2}} \tag{3.9}$$

Clearly, larger Q-factor and k-factor lead to higher efficiency or equivalently lower insertion loss. One important observation here is that the efficiency is independent of the impedance transformation ratio ($\approx n^2 = \frac{L_s}{L_p}$). This effectively means the tradeoff between output power and matching network loss is eliminated, and higher output power level can be achieved without incurring larger passive loss.

In Fig. 3.1c, the transformer combines the signals from a push-pull pair (pseudo-differential pair) through voltage stacking. This effectively means each transistor sees half of the load impedance, and maximum output power can be delivered by each transistor is twice that of a single transistor PA. Combining the power of two such transistors, the total output power is four times larger. Transformers offer additional advantages such as built-in differential to single-ended signal conversion and convenient DC bias through center tap in a multi-stage design. Mm-Wave PAs based on such transformers have shown output power beyond 12dBm with decent efficiency [15, 16, 17].

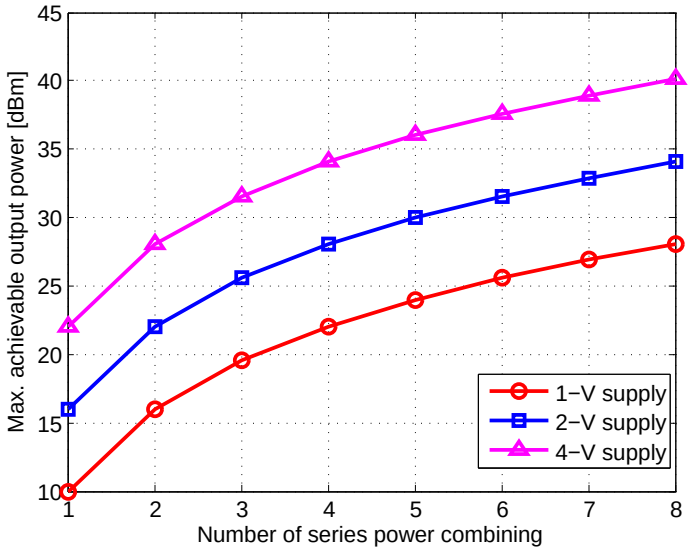


Figure 3.3: Theoretical output power of a DAT based PA as a function of number of combined inputs.

3.2 DAT Combiners

Although the insertion loss of a transformer is in theory independent of the impedance transformation ratio, in real implementation, the k-factor reduces when n increases. Since the efficiency of a transformer is proportional to the k-factor as shown in Eq. 3.6, the insertion loss does gradually increase with the impedance transformation ratio. Therefore, n cannot be increased arbitrarily to increase the power level. The better way to increase the power level without incurring additional transformer loss is to stack voltages from more transistors, and it can be achieved by breaking the transformer primary winding into multiple sections, with each section loaded with a push-pull transistor pair. This technique was introduced in [18], known as Distributed Active Transformer (DAT). The original DAT has four differential pair and combines the power from eight identical transistors. Due to voltage stacking, the load impedance seen by each transistor is one eighth of the load impedance. Therefore each transistor can deliver eight times larger output power than a single transistor PA, and by combing eight such transistors, the total maximum output power enhancement is 64 times [9]. The output power level can be increased by increasing the number of input ports in the primary side, and the combined power is proportional to N^2. Since the transformer geometry remains the same, the insertion loss is kept constant. Fig. 3.3 shows the maximum combined output power as a function of number of input ports with different supply voltages. The output power increases by 6dB when the number of inputs doubles. Although the concept was first developed for microwave PAs, it can be naturally extended to higher frequency. In theory, the output power can be increased indefinitely by increasing the number of combined input ports. In practice though, transistors have certain dimension and the interconnect wires

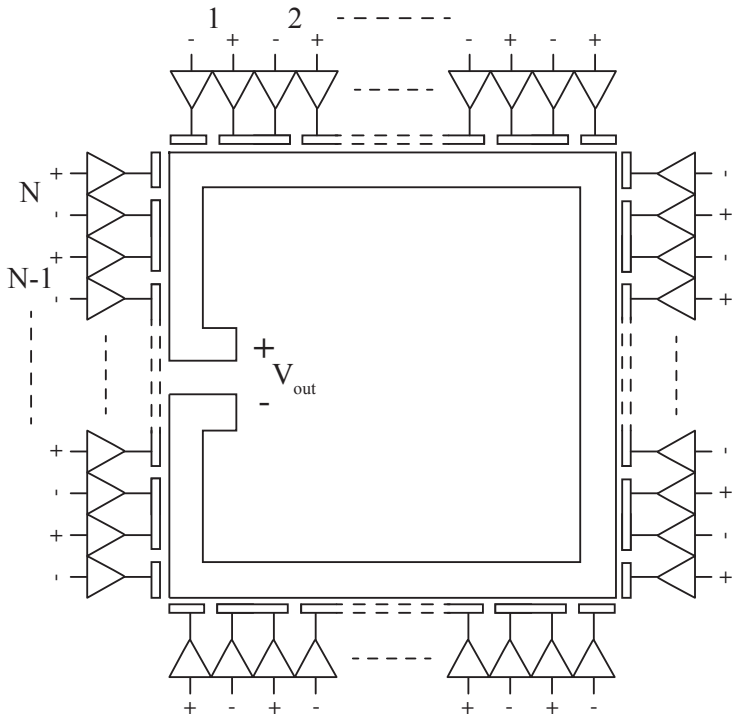


Figure 3.4: Increase the number of inputs to the limit.

also contribute to the total inductance of the primary winding. One can keep increasing the number of input ports until the point that the transistor lateral dimension is comparable to the length of the inductor slab (Fig. 3.4). At this point, there's no room on the primary winding perimeter to accommodate additional transistors. Further increasing the number of input ports does not improve the output power. In fact this limit can be derived based on a particular process (65nm CMOS is used for all the following derivation).

For a single-turn 1:1 transformer driven by a single transistor, the load impedance for the transistor is untransformed (50Ω). Based on Eq. 2.23, the optimal bias current for the transistor is equal to the supply voltage and load impedance ratio,

$$I_{b,opt} = \frac{I_{max}}{2} = \frac{V_{dd}}{R_L} \tag{3.10}$$

V_{ov} is neglected for simplicity. For 1-V supply, $I_{b,opt}$ is 20mA. In a typical 65nm CMOS process, the current-density i_{den} for maximum f_t is around 0.5mA/μm. Therefore, 40μm device size is needed and this corresponds to roughly $40fF$ of output capacitance ($\overline{c_{out}} = 1fF/\mu m$). The inductance needed to resonate out this capacitance is around 176pH at 60GHz. This sets the basic requirement of the PA and the transformer when the number of input is unity. Every time the number of input ports doubles, the load impedance seen by each transistor drops by half and therefore each transistor size needs to be doubled for maximizing the voltage swing. This means the inductance needed for each transistor is halved while the total loop inductance remain constant. Assuming a square shape single-turn

inductor, the inductance can be expressed in terms of the diameter as follows:

$$L = 0.2\pi D \times 1.27(\ln(\frac{2.07}{\beta}) + 0.18 \times \beta + 0.03 \times \beta^2) \tag{3.11}$$

where D is the diameter and β is the filling factor and is equal to 0.0833 for the square shape. Rearranging the terms,

$$D = \frac{L}{0.2\pi \times 1.27(\ln(\frac{2.07}{\beta}) + 0.18 \times \beta + 0.03 \times \beta^2)} \tag{3.12}$$

For N numbers of inputs, the size of each single transistor needs to be,

$$W_0 = N\frac{I_{b,opt}}{i_{den}} \tag{3.13}$$

On the other hand, the maximum lateral dimension allowed on one side of each transistor is,

$$l_{0,max} = \frac{\pi D}{N} \tag{3.14}$$

In a typical 65nm CMOS process, each finger extends about 0.25μm including gate, source and drain area. As a result, the maximum number of finger allowed to fit in $l_{0,max}$ is,

$$N_{finger,max} = \frac{l_{0,max}}{\overline{l_{finger}}} = \frac{l_{0,max}}{0.25\mu m} \tag{3.15}$$

Since the transistor size is set by the number of fingers and the finger width together, one can always increase the finger width to meet the device requirement in (3.13). However, increasing the finger width increases the gate resistance and therefore reduces the transistor gain. This imposes another constraint. The maximum power gain of a transistor is,

$$G_{max} = \frac{1}{4}(\frac{\omega_t}{\omega_0})^2\frac{r_o}{r_g} \tag{3.16}$$

$$= \frac{1}{4}(\frac{\omega_t}{\omega_0})^2(\frac{\overline{r_o}}{\overline{r_g}})(\frac{1}{W_{finger}})^2 \tag{3.17}$$

where $\overline{r_o}$ and $\overline{r_g}$ are the normalized output resistance and input gate resistance. Since r_o is inversely proportional to finger width W_{finger} while r_g is proportional to W_{finger}, the ratio of r_o to r_g is inversely proportional to the square of W_{finger}. In order to maintain a minimum gain of G, the W_{finger} must not exceed a certain value,

$$W_{finger,max} = \frac{1}{4}(\frac{\omega_t}{\omega_0})^2\frac{\overline{r_o}}{\overline{r_g}}\frac{1}{G} \tag{3.18}$$

This sets an upper limit on the size of each transistor. When this upper limit reaches the optimal transistor size obtained in Eq. 3.13, the maximum number of combining has been reached.

$$N_{finger,max} \times W_{finger,max} \geq W_0 \tag{3.19}$$

$$\frac{\pi D}{N}\frac{1}{0.25\mu m}\frac{1}{4}(\frac{\omega_t}{\omega_0})^2\frac{\overline{r_o}}{\overline{r_g}} \geq N\frac{I_{b,opt}}{i_{den}} \tag{3.20}$$

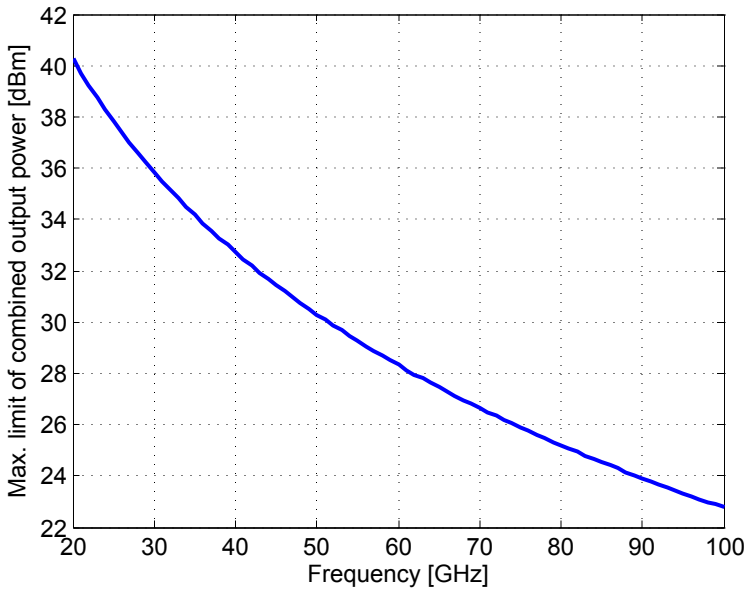


Figure 3.5: Maximum total combined output power

The only unknown here is the number of combined inputs N, therefore it can be solved,

$$N_{max} = \sqrt{\pi D \frac{1}{\overline{l_{finger}}} \frac{1}{4} (\frac{\omega_t}{\omega_0})^2 \frac{\overline{r_o}}{\overline{r_g}} \frac{i_{den}}{I_{b,opt}}} \tag{3.21}$$

Fig. 3.5 and Fig. 3.6 plot the maximum limit on the combined output power and the number of inputs as a function of frequency for a minimum power gain of 3dB. Note that this is the absolute maximum achievable power without considering additional loss from passives. The frequency dependence originates from the fact that larger inductance is needed for tuning out the same capacitance at lower frequency and therefore larger inductor perimeter is allowed to fit more transistors. In practice, different inductor/transformer geometry will lead to slightly different results due to different filling factor β. Fig. 3.7 shows the total tuning inductance as a function of the operating frequency. One thing to point out here is that at lower frequency an explicit capacitor needs to be placed to reduce the tuning inductance. For example, an inductance of 1.6nH is needed for a 20GHz DAT and it's very difficult to design such a large inductor with Self Resonance Frequency (SRF) beyond 20GHz.

3.3 A 60GHz DAT Power Amplifier

Based on the previous analysis, a DAT based 60GHz PA is implemented in 65nm CMOS. The detailed design methodology for both the actives and the passives, as well as the measurement results are discussed in this section.

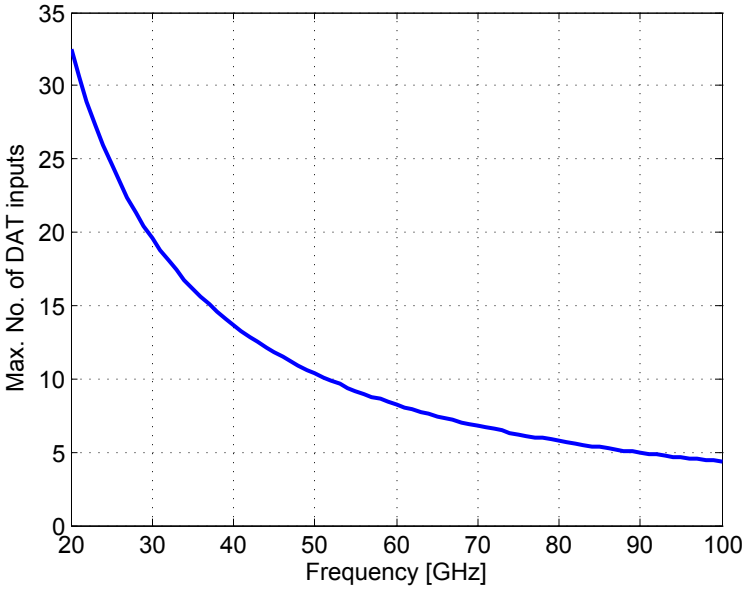


Figure 3.6: Maximum number of DAT inputs

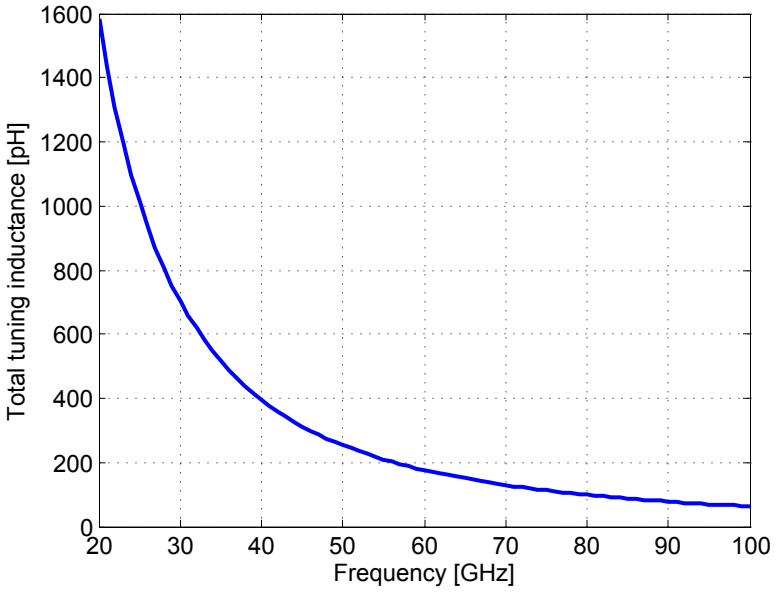


Figure 3.7: Total tuning inductance of the transformer

3.3.1 Design Procedure

In Fig. 3.6, the maximum number of DAT inputs at 60GHz is eight, based on a target power gain of 3dB. Unfortunately such low gain is not very practical since it requires significant driver power and reduces the PAE. Instead the design of the last stage targets a power gain of at least 6dB. This reduces the number of DAT inputs to four. The proposed combiner is shown in Fig. 3.8. The primary side of the transformer combiner has two pairs of differential excitation ports located on the opposite sides of the metal winding while the secondary side is the same as the standard transformer. The two pairs of differential ports are excited in such a way that the AC current circulates in the primary winding and induces the magnetic flux for the secondary winding while the center points of the two half-octagon traces provide convenient DC access. Since the voltage swings are effectively stacked, the impedance seen at each input port is one quarter of the load. Consequently, this structure has a voltage enhancement ratio of 4 and a power enhancement ratio of 16.

Two of the most critical design aspects of the transformer combiner are the insertion loss and the input impedance. As shown in Eq. 3.6, the insertion loss is mainly determined by the k factor and the inductor Q factor. A vertical broadside coupled structure with primary winding on the top and secondary winding on the bottom is used and a coupling factor of 0.87 is achieved at 60GHz. The thin metals in digital CMOS processes usually limit the achievable Q factor of on-chip inductors. In this design, the aluminum capping layer, which is usually used to cover the copper bondpads, is strapped together with the top copper layer to form the primary winding. The aluminum capping layer has a thickness of 1.2μm while the top two copper layers have the same thickness of 0.9μm. By forming a much thicker primary metal winding, the Q factor is improved by 50%, reaching 15.5 at 60GHz. The insertion loss of the combiner is further evaluated as a function of transformer radius and trace width using EM simulators. Fig. 3.9 shows the combiner loss as a function of the transformer dimensions. The optimal insertion loss is achieved when metal trace width is greater than 15μm and the radius is between 25μm to 30μm. However, the design rule limits the maximum trace width to 12μm. Therefore the metal trace width should always to be set to this maximum limit. The lowest insertion loss under this constraint is around 0.65dB.

On the other hand, the input impedance of the transformer must be taken into account when deciding the physical dimensions. Since the input susceptance is a strong function of the transformer radius, the transformer size is determined mainly by the transistor output capacitance. However, the transistor size also depends on the load resistance presented by the transformer. As a result, several iterations are needed to find the optimal size combination of the transistor and the transformer. This iterative process starts with an estimate of the load resistance at each input port, which is roughly one quarter of the load ($R_l = 12.5\Omega$). Optimal bias current $I_{b,opt}$ can be calculated based on R_l. With 1V supply and 200mV V_{ov}, $I_{b,opt}$ is around 64mA. 200mV is chosen to maximize the f_{max} of the transistor. With current density of $0.5mA/\mu m$, the required transistor size is roughly $130\mu m$. A load-pull simulation is performed on the device to determine the optimal load impedance for output power, and a transformer size can be chosen based on the required load impedance. Such a process can be repeated in order to find the optimal active-passive combination. Note that the optimal combination may be different for different goals such as maximizing the

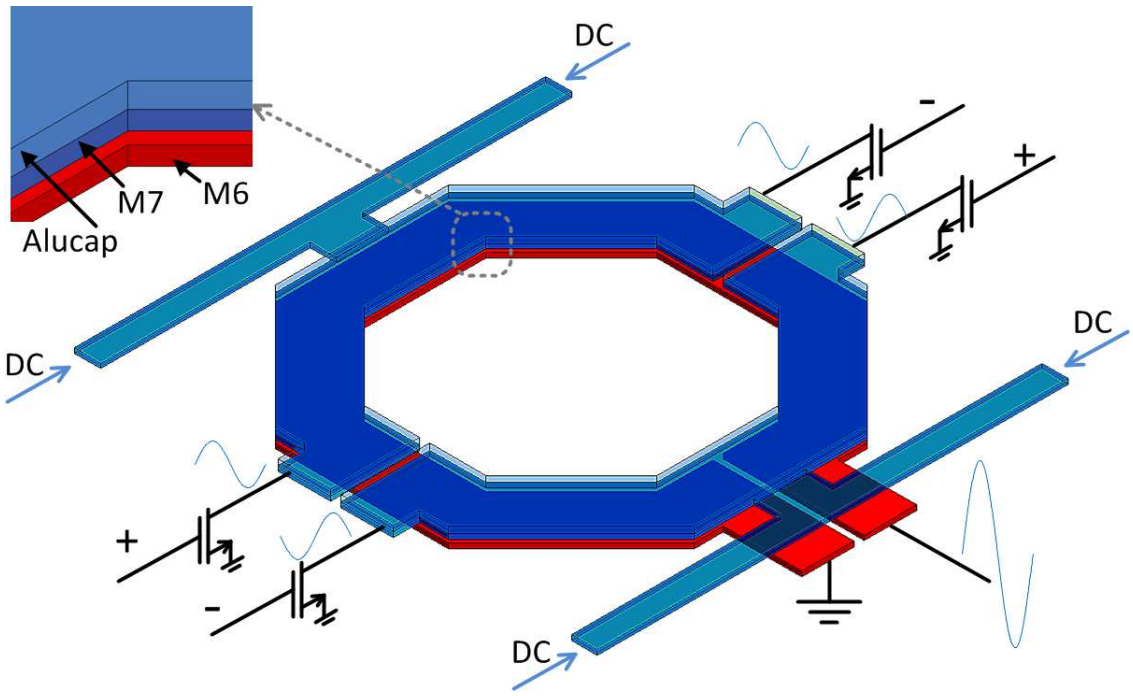


Figure 3.8: Quad-input 60GHz DAT Combiner

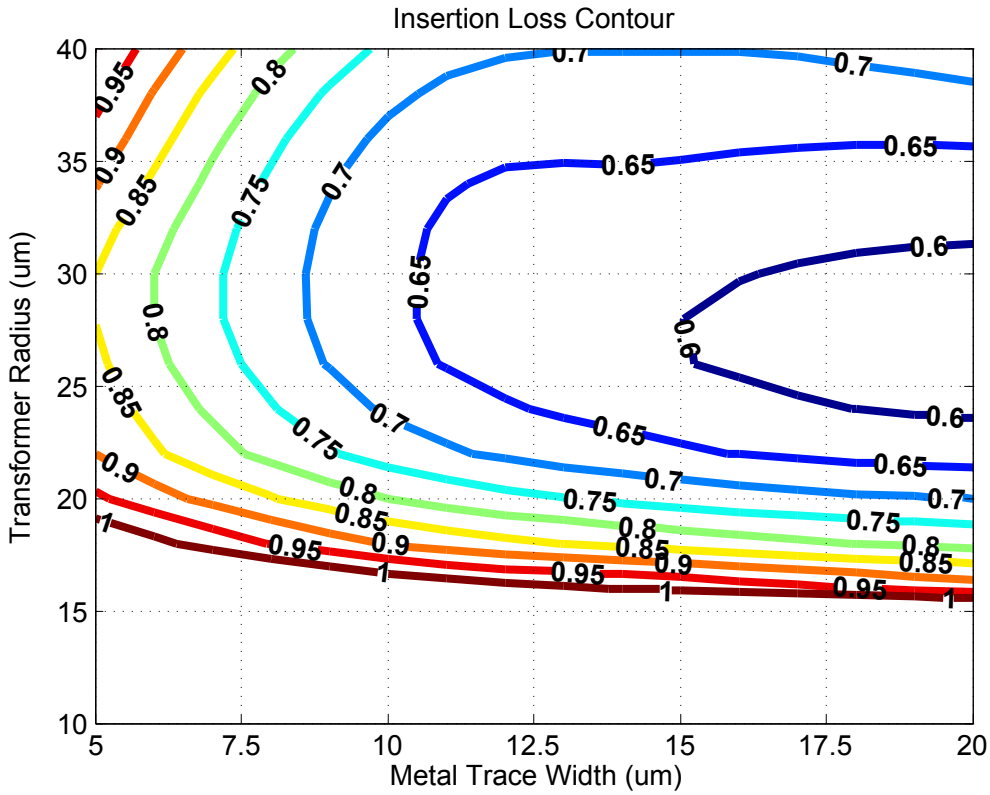


Figure 3.9: Insertion loss of the quad-input DAT combiner

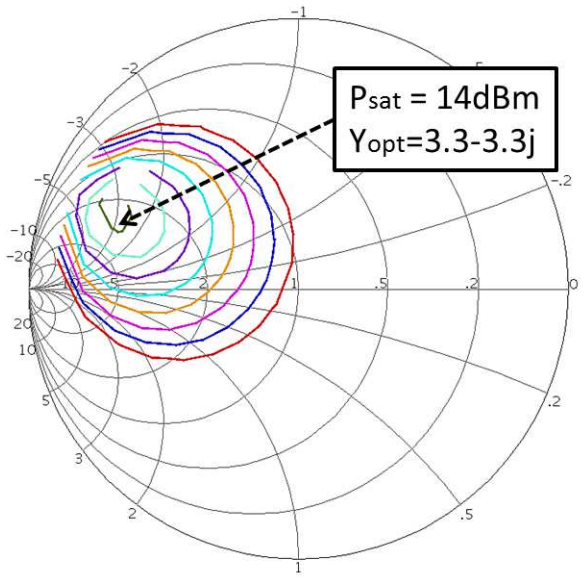


Figure 3.10: Load-pull of a 140μm transistor

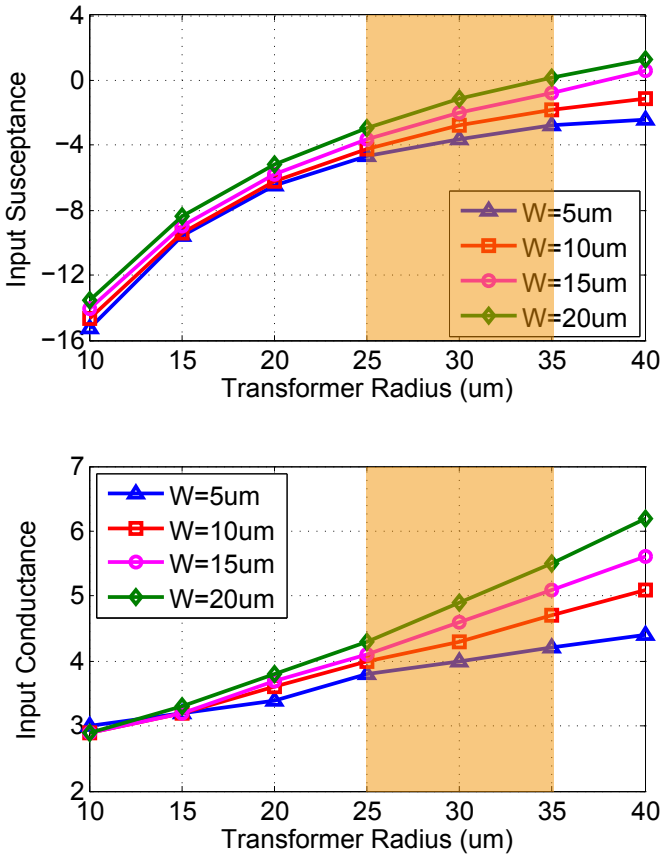


Figure 3.11: Input susceptance and conductance of the quad-input DAT combiner

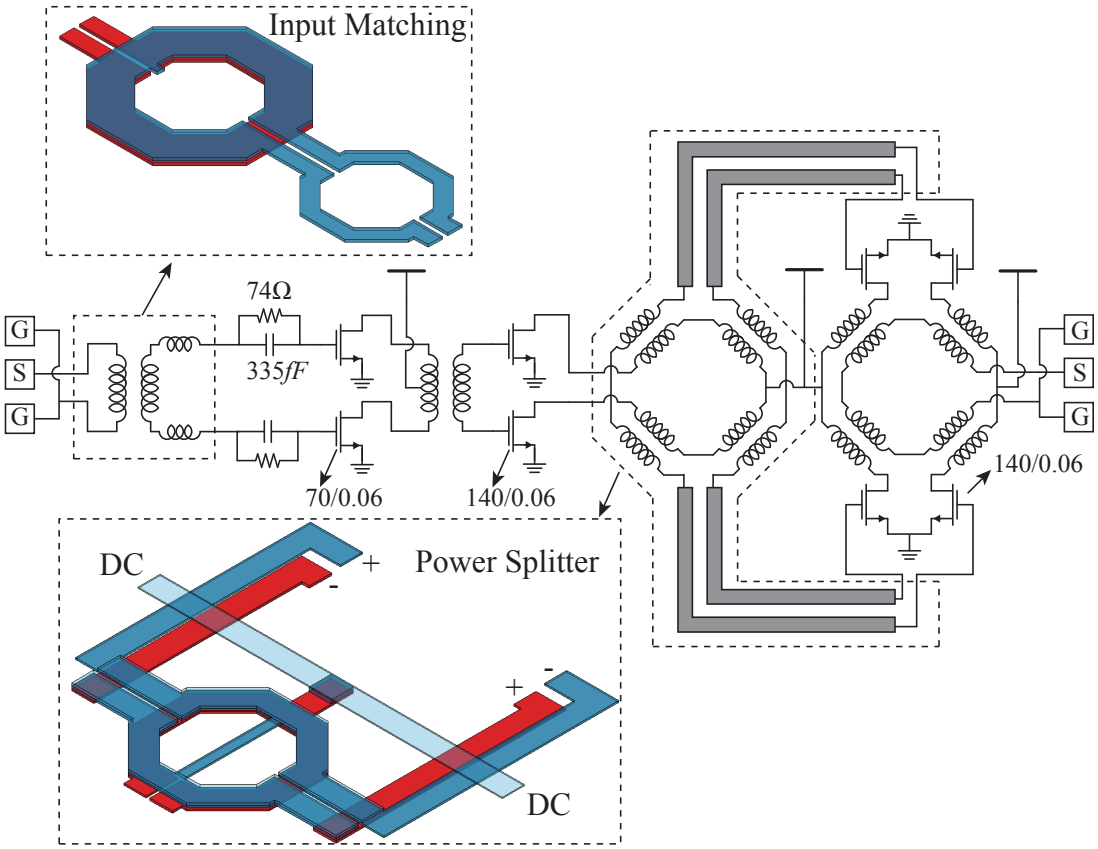


Figure 3.12: Schematic of the three stage 60GHz PA.

output power or maximizing the efficiency. In this design, maximizing output power is used as the main criteria. The final device size is set to 140μm, with the load-pull result shown in Fig. 3.10. The optimal normalized admittance is roughly $3.3 - 3.3j\Omega$. Fig. 3.11 plots the input susceptance and input conductance of the DAT combiner, with the optimal insertion loss region highlighted in yellow. 28μm radius and 12μm trace width are chosen with an insertion loss of 0.63dB. Note that the actual load conductance is set slightly higher than the optimal value from loadpull as a compromise for lower transformer insertion loss.

Characterization of power transistor behaviors at mm-wave frequencies is important to ensure the PA performance. A test chip with a stand-alone 140μm power transistor was fabricated and two port S-parameters were measured up to 110GHz for different bias points. A simple wrap-around model with lumped inductors and resistors are used to achieve better device π-model fitting at mm-wave frequencies.

Fig. 3.12 shows the complete schematic of the three-stage PA. A power splitter is needed preceding the output stage in order to feed two differential pairs and the same transformer combiner structure is once more utilized with the inputs and outputs swapped to perform the task. Due to the large mismatch between the input conductance of the output stage and the output conductance of the driver stage, an extra matching network is required to compensate the limited conductance transformation range of 1:1 transformers. To accomplish

this goal, two pairs of differential strip-lines are inserted between the transformer splitter and the output differential pairs. These strip-line pairs also conveniently cover the geometric distance of the transistors separated by the transformers. A similar conductance mismatch at the input of the PA is solved by adding a series differential inductor to the transformer. To ensure the PA stability, a parallel RC combination is added to introduce resistive loss at lower frequencies [15].

3.3.2 CW Measurement Results

The PA was fabricated in a 1V 65nm GP 1P7M digital CMOS process. The measured S-parameters are shown in Fig. 3.13. With 1V supply, the PA achieves a power gain of 20.2dB and a 3-dB bandwidth of 9GHz (56GHz to 65GHz). The amplifier is unconditionally stable over the entire measured frequency range with stability factor greater than unity. The measured S-parameters are compared to the simulation results based on pre-measured transistor S-parameters. It can be observed that the measurement results are very close to the simulation.

The 60GHz power measurement results are shown in Fig. 3.14. With 1V supply voltage, the measured 1dB gain compressed output power P_{1dB} is 15dBm and the saturated output power P_{SAT} is 18.6dBm. The measured peak PAE is 15.1% and the peak drain efficiency is 16.4%. At the saturated output power level, the amplifier still has 11dB of power gain, which significantly relieves the design of the preceding block. The large signal performance is also measured over the IEEE 802.15.3c band. As shown in Fig. 3.15, the PA maintains over 17.8dBm P_{SAT}, 13.8dBm P_{1dB} and 12.6% PAE from 58GHz to 64GHz. A chip micrograph of the PA is shown in Fig. 3.16. Due to the use of the compact transformer as both a power combiner and a splitter, the entire PA only occupies an area of $0.28mm^2$ including the GSG RF pads.

3.3.3 Modulation Measurement Results

The PA is also tested with the IEEE 802.15.3c modulated signal. The Modulation and Coding Scheme (MCS) data is generated by a SiBEAM transmitter, passed through the 60GHz PA, and demodulated by a SiBEAM receiver. Fig. 3.17 and Fig. 3.18 show the PA output spectrum at channel 2 and channel 3 when transmitting a 512 sub-carrier 16-QAM OFDM signal with a channel bandwidth of 1.76GHz at a data rate of 3.8Gb/s. The spectral mask for channel 2 and channel 3 are also shown in red colors. The received constellation maps for channel 2 and channel 3 are plotted in Fig. 3.19 and Fig. 3.20 respectively, with corresponding EVMs of -17.7dB and -17dB. This EVM is achieved at an average transmitted power of 9dBm.

In order to evaluate the necessary power back-off for this type of MCS data, EVM measurements are performed at different average output power levels. Multiple measurements are done at each power level to reduce the impact of measurement error. The EVMs are

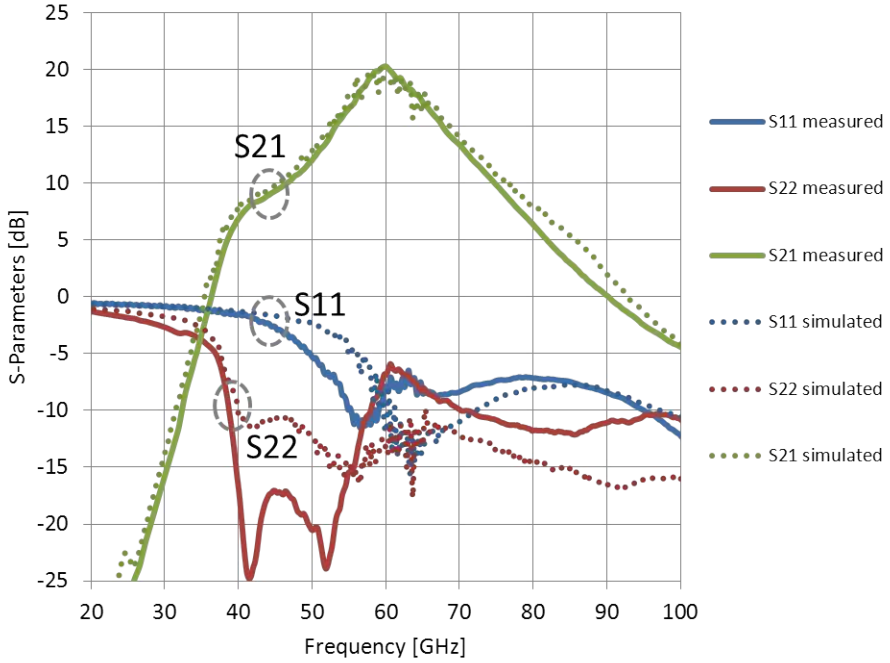


Figure 3.13: S-Parameters of the 60GHz PA.

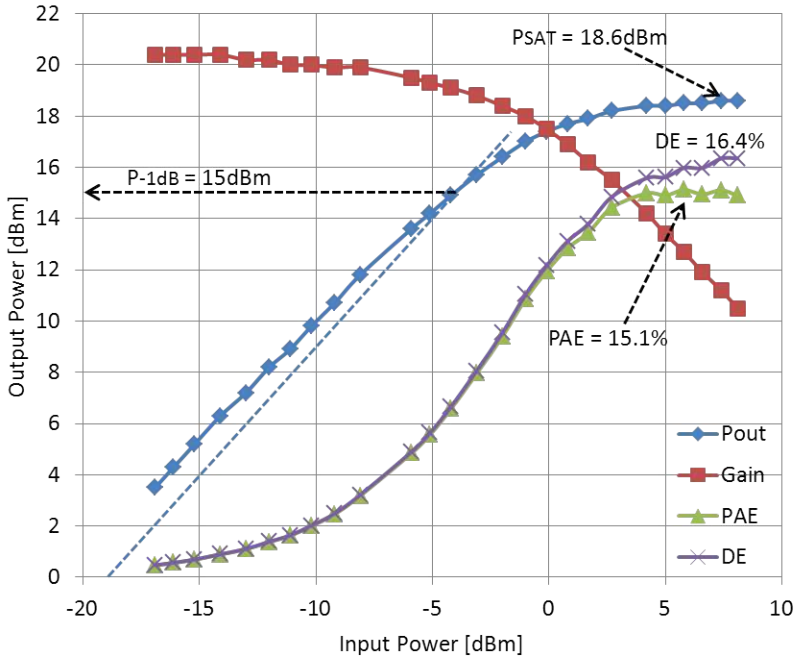


Figure 3.14: Output power, efficiency and gain of the 60GHz PA.

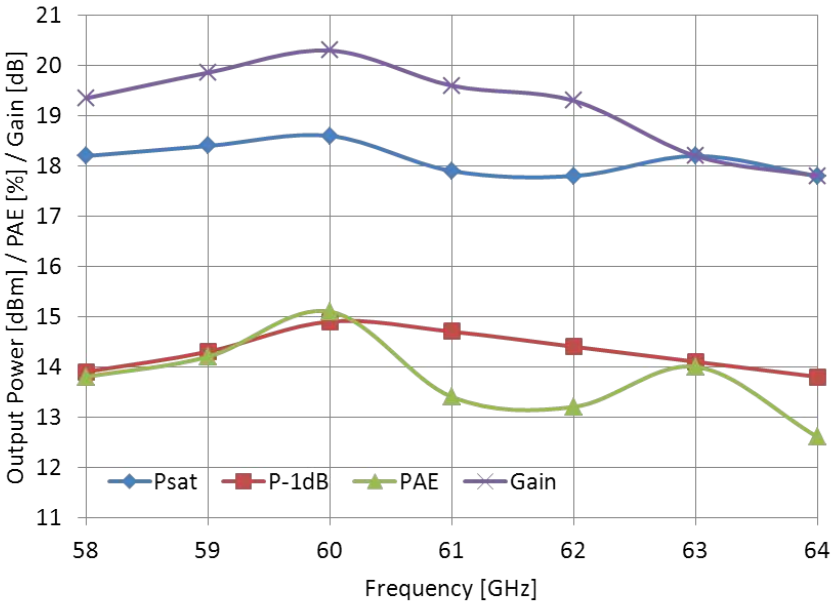


Figure 3.15: 60GHz PA large signal performance across the IEEE 802.15.3c band.

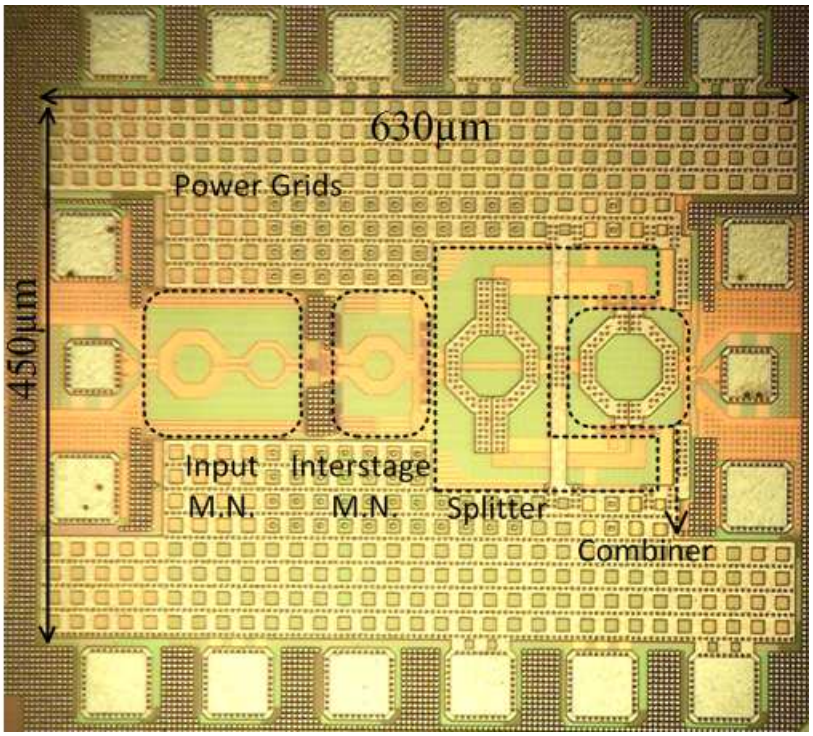


Figure 3.16: Chip micrograph of the 60GHz PA.

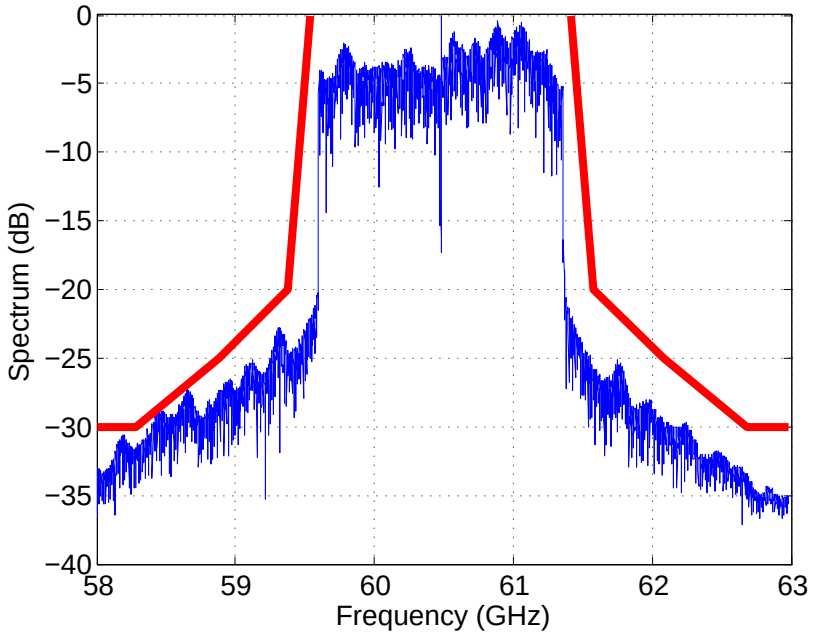


Figure 3.17: PA output spectrum and 802.15.3c spectral mask for Channel 2

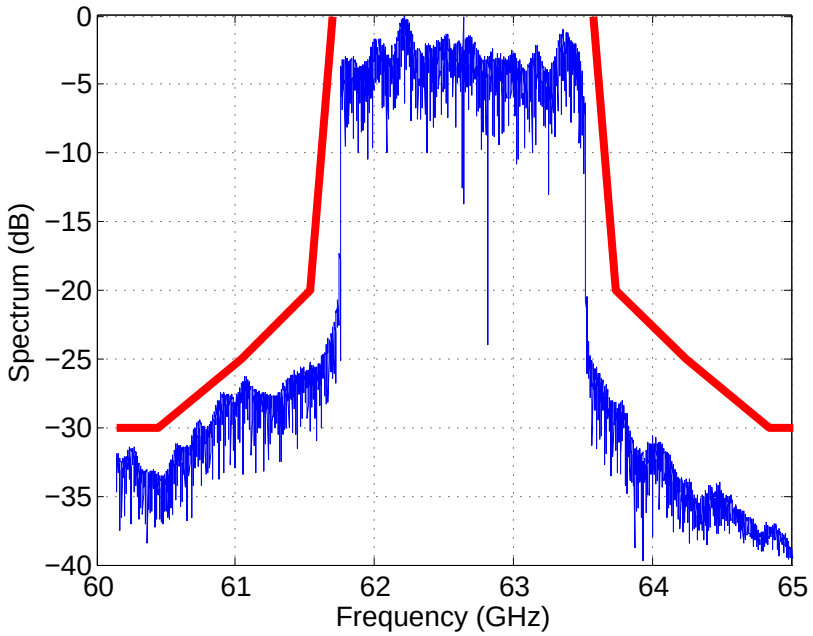


Figure 3.18: PA output spectrum and 802.15.3c spectral mask for Channel 3

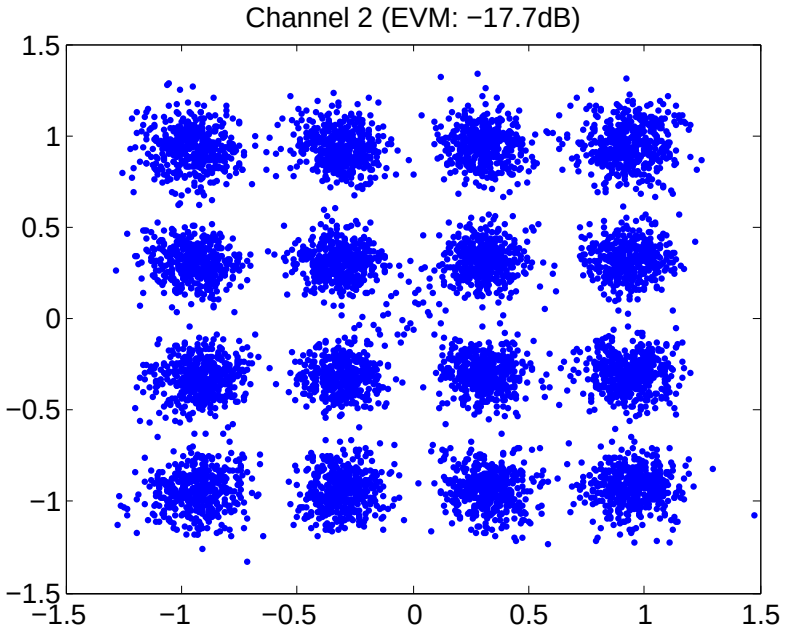


Figure 3.19: Received constellation of the 3.8Gb/s 512 sub-carrier 16-QAM OFDM signal for Channel 2.

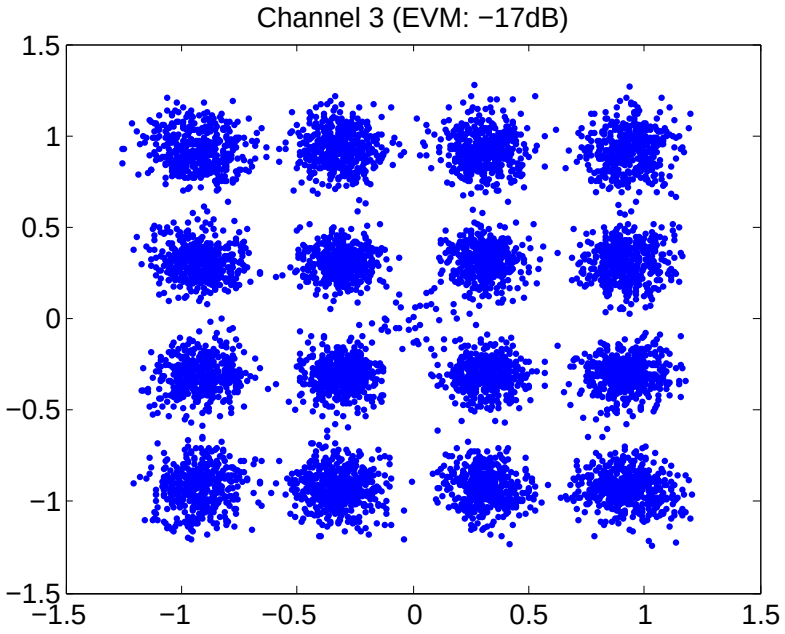


Figure 3.20: Received constellation of the 3.8Gb/s 512 sub-carrier 16-QAM OFDM signal for Channel 3.

plotted as a function of average transmitted power in Fig. 3.21 and Fig. 3.22. To achieve -20 dB EVM, the required back-off from 1dB compression point (15 dBm for channel 2 and 14.5 dBm for channel 3) is around 7.5 dB. The required back-off reduces to 6dB for -17dB EVM. The PA efficiencies at 6dB backoff and 7.5dB backoff are 2% and 1.5% respectively.

Finally, the above EVM measurement has been performed at different ambient temperatures. Heat was applied to the probe station chuck where the silicon die resides. Fig. 3.23 and Fig. 3.24 show the measured EVM curves at several different temperature settings. Clearly, both EVM and average output power degrade when temperature increases. The degradation is relatively small from $25°C$ to $50°C$, and become more significant for temperatures greater than $75°C$. Note that under all the measurement, the transistor gate bias voltage is kept constant, which means the actual bias current reduces with increasing temperature. The impact of temperature on output power can be slightly reduced by using constant current biasing or Proportional To Absolute Temperature (PTAT) current biasing.

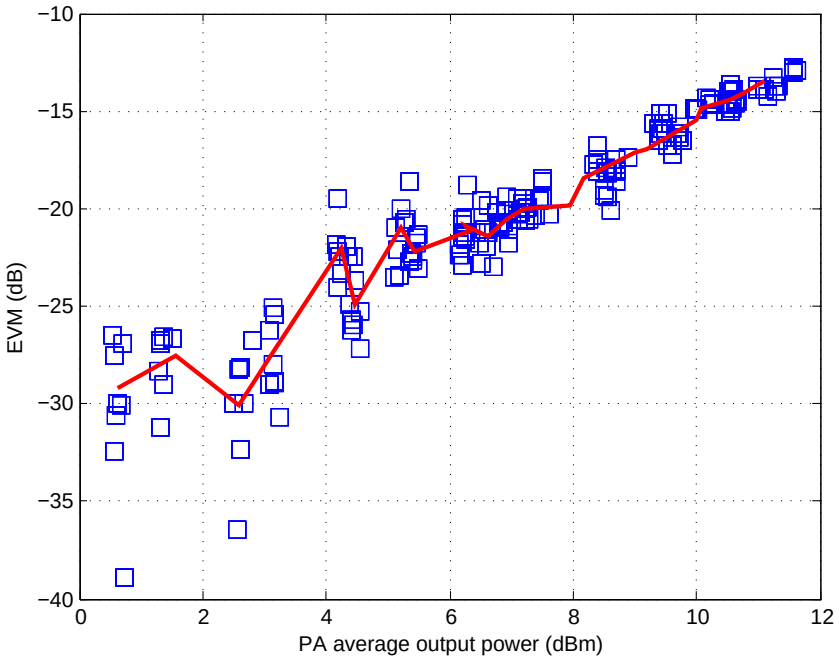


Figure 3.21: Measured PA EVM for channel 2 as a function of average output power

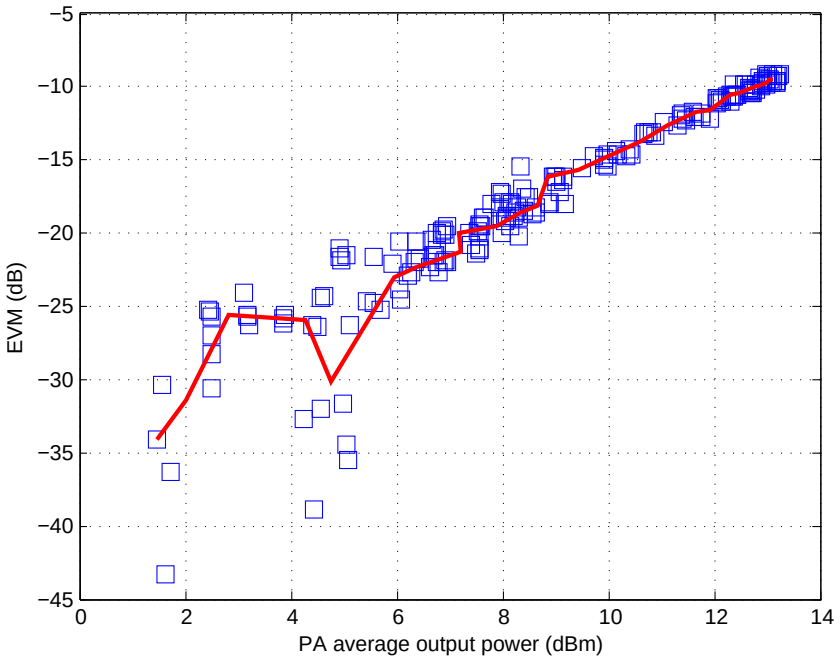


Figure 3.22: Measured PA EVM for channel 3 as a function of average output power

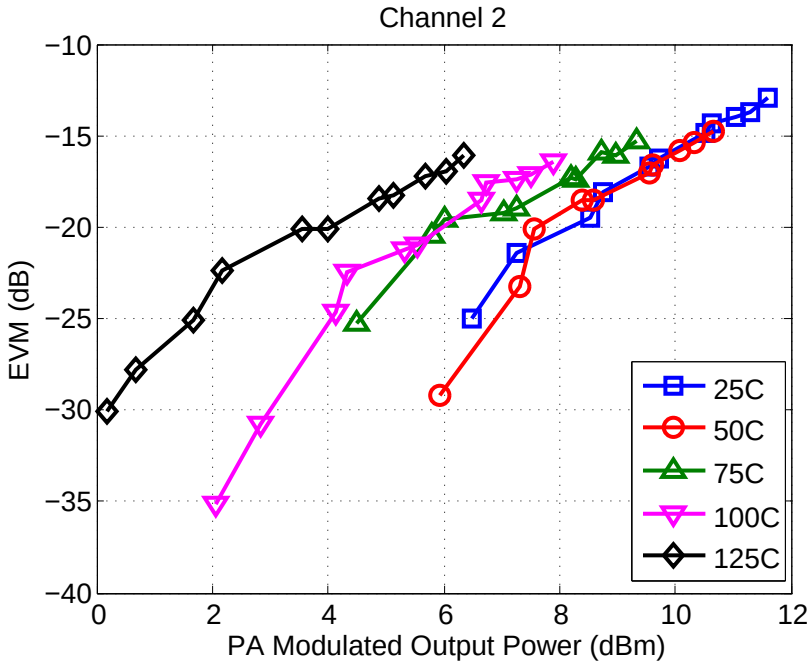


Figure 3.23: Measured PA EVM characteristic for channel 2 at different ambient temperatures.

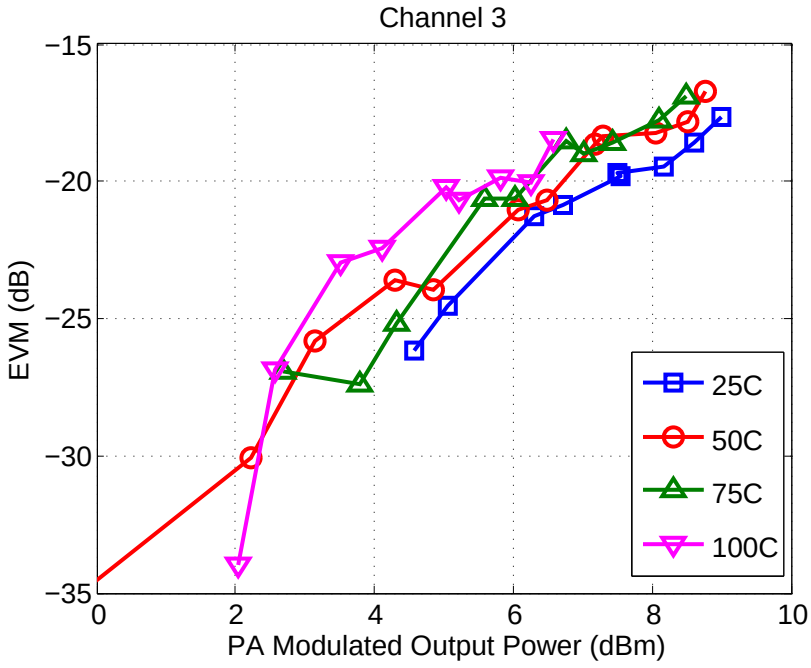


Figure 3.24: Measured PA EVM characteristic for channel 3 at different ambient temperatures.

Chapter 4

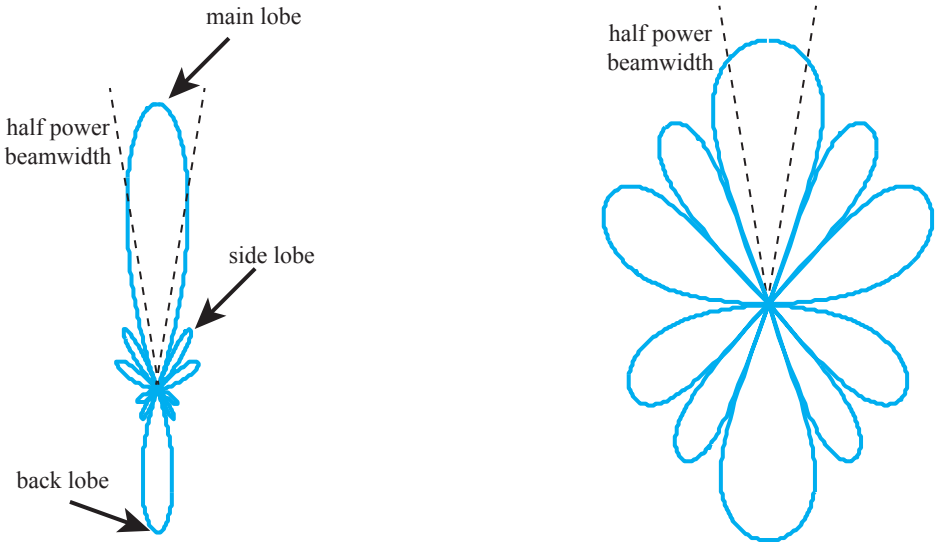
Beamforming Transmitter

As discussed in the previous chapter, the obtainable PA output power drops with frequency due to low supply voltage and passive loss. Even with power combining, the maximum output power is still limited. With DAT combiners, the achievable output power at 60GHz is around 28dBm for a power gain of 3dB. An alternative to increase the transmitter EIRP is to increase the antenna gain. Increased antenna gain means that the radiated energy is more concentrated towards one particular direction, and longer communication distance can be obtained. However, larger antenna gain reduces the signal strength in other directions, and to redirect the energy, the antenna must be physically rotated. Such mechanical movement is very inconvenient and costly. Beamforming transmitters realize an electrically steerable antenna beam by adjusting the electrical signal properties of each antenna element. Beamforming becomes especially attractive in CMOS since the digital computation and control come at an extremely low cost. This chapter discusses the concepts and challenges of CMOS beamforming, as well as a design example of a low power four element transceiver array.

4.1 Antenna Basics

To fully understand beamforming, the definition of various antenna parameters must be explained. The most important parameters that characterize an antenna include radiation pattern, gain, efficiency and polarization.

The antenna radiation pattern is defined as a graphical representation of the radiation properties as a function of space coordinates [19]. Radiation properties include electromagnetic field strength, power density and directivity etc. The radiation pattern can be constructed by plotting the received electric field magnitude at a constant radius as a function of angle. Such a pattern is called an amplitude field pattern. Similarly, a plot of received power density is called a power pattern. Fig. 4.1 shows an example antenna radiation pat-



(a) Linear scale (b) Log scale

Figure 4.1: Antenna radiation pattern in a polar plot

tern. A typical radiation pattern contains several parts that are commonly referred to as main lobe, side lobes and back lobes. A main lobe is a portion of the radiation pattern that contains the maximum energy. A side lobe is a radiation lobe in any direction other than the intended lobe, and a back lobe is a radiation lobe pointing to the opposite direction of the main lobe. The radiation patterns are usually plotted in log scale for better visualization of the side lobes. To quantize the radiation energy concentration, a parameter called beamwidth is used. The beamwidth of a pattern is defined as the angular separation between two identical points on opposite side of the pattern maximum. The most widely used beamwidth is the half-power beamwidth, which is the angle between the two directions in which the radiation intensity is one-half value of the peak. The beamwidth is an important parameter from communication perspective since it determines the link coverage.

The directivity of an antenna is defined as the ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions. The average radiation intensity is equal to the total power radiated by the antenna divided by 4π. In other words, the directivity of a nonisotropic antenna is equal to the ratio of its radiation intensity at a given direction over that of an isotropic antenna.

$$D(\theta,\phi) = \frac{U(\theta,\phi)}{U_0} = \frac{4\pi U(\theta,\phi)}{P_{rad}} \tag{4.1}$$

$$D_{max} = \frac{U_{max}}{U_0} = \frac{4\pi U_{max}}{P_{rad}} \tag{4.2}$$

where $U(\theta,\phi)$ is the radiation intensity of a nonisotropic antenna at a particular solid angle (W/unit solid angle), U_{max} is the maximum radiation intensity of a nonisotropic antenna (W/unit solid angle), U_0 is the radiation intensity of an isotropic antenna (W/unit solid

angle), and P_{rad} is total radiated power (W). Directivity informs how well an antenna can concentrate energy around a particular spatial direction. Maximum directivity is always greater than or equal to unity, and the directivity of an isotropic antenna is always unity. Designing antennas usually involves a tradeoff between beamwidth and directivity since the beamwidth decreases as directivity increases. In reality, an antenna only radiates a portion of the power that it absorbs while the rest of the power is dissipated internally. The internal losses of an antenna include the conduction loss and the dielectric loss. The radiation efficiency is defined as the ratio of the total radiated power to the total absorbed power,

$$\eta_{rad} = \frac{P_{rad}}{P_{in}} \tag{4.3}$$

Taking into account of the radiation efficiency, the antenna gain can be defined,

$$G(\theta, \phi) = \eta_{rad} D(\theta, \phi) \tag{4.4}$$

Substitute Eq. 4.1 into Eq. 4.4,

$$G(\theta, \phi) = \frac{P_{rad}}{P_{in}} \frac{4\pi U(\theta, \phi)}{P_{rad}} \tag{4.5}$$

$$= \frac{4\pi U(\theta, \phi)}{P_{in}} \tag{4.6}$$

In other words, the gain of an antenna is the ratio of the radiation intensity in a given direction to the radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically.

Finally, the polarization of an antenna describes the time-varying orientation of the electric field vector radiated by the antenna. The trace formed by the electric field vector can be a line, a circle or an ellipse, and they correspond to linear polarization, circular polarization and elliptical polarization respectively. The radiated signal from the transmitter antenna can be completely received only if the receiver antenna has the exactly same polarization. Otherwise, energy is lost due to polarization mismatch. In practice, it's almost impossible to avoid polarization mismatch since any reflection results in a change in the wave polarization.

Besides the above essential parameters, there are also several other parameters that need to be considered, such as input impedance and bandwidth. The required input impedance of the antenna is usually 50Ω in order to match the feed line characteristic impedance[1]. In cases where the antenna is attached very closely to the transmitter or receiver, the input impedance can be co-designed with the PA or LNA, and doesn't necessarily have to be 50Ω. The choice of the impedance should be such that maximizing the PA efficiency or minimizing the LNA noise figure. The antenna bandwidth is usually set by the system data rate. Higher date rates usually require larger antenna bandwidth. For example, the WiGig standard has four channels from 58.32GHz to 64.8GHz with 2.16GHz channel bandwidth. As a result, the antenna needs to preserve a bandwidth greater than 8GHz.

[1] 50Ω impedance also easies the antenna stand-alone characterization since most RF instruments use 50Ω ports.

4.2 Beamforming through Antenna Array

In high directivity antenna systems, beam steering is essential to cover larger communication angles. Electrical beam steering can be realized by an antenna array with individual control on the signal transmitted or received in each antenna element. Consider a N-element antenna array shown in Fig. 4.2. To steer the beam θ angle away from the center, the signals that excite the antenna element must be progressively delayed such that they add constructively in the desired transmission direction. Suppose the signal inside the n^{th} element is delayed by $(n-1)\tau$. Due to the antenna spacing d, signals emitted by different antenna elements have different delays to the wavefront. The free space delay difference between the n^{th} antenna and the last (N^{th}) antenna is,

$$t_n = \frac{(N-n)d\sin\theta}{c} = (N-n)t_0 \tag{4.7}$$

$$t_0 = \frac{d\sin\theta}{c} \tag{4.8}$$

The E-field vector contributed by the n^{th} antenna at the wavefront is,

$$E_n = e^{j\omega_0[t-t_n-(n-1)\tau]} \tag{4.9}$$

$$= e^{j\omega_0[t-(N-n)t_0-(n-1)\tau]} \tag{4.10}$$

$$= e^{j\omega_0[t-(N-1)t_0]}e^{j\omega_0[(n-1)(t_0-\tau)]} \tag{4.11}$$

Summing the E-field vectors of all N antenna,

$$E_{array} = \sum_{n=1}^{N} E_n \tag{4.12}$$

$$= e^{j\omega_0[t-(N-1)t_0]}\sum_{n=1}^{N} e^{j\omega_0[(n-1)(t_0-\tau)]} \tag{4.13}$$

$$= e^{j\omega_0[t-(N-1)t_0]}\sum_{n=0}^{N-1} e^{j\omega_0[n(t_0-\tau)]} \tag{4.14}$$

$$= e^{j\omega_0[t-(N-1)t_0]}e^{j\omega_0\frac{(N-1)(t_0-\tau)}{2}}\frac{\sin[\frac{1}{2}N\omega_0(t_0-\tau)]}{\sin[\frac{1}{2}\omega_0(t_0-\tau)]} \tag{4.15}$$

The magnitude of the summed E-field vector is,

$$|E_{array}| = \left|\frac{\sin[\frac{1}{2}N\omega_0(t_0-\tau)]}{\sin[\frac{1}{2}\omega_0(t_0-\tau)]}\right| \tag{4.16}$$

The summed E-field magnitude can be maximized at spatial angle θ by setting $\tau = t_0$,

$$\tau_{opt} = t_0 = \frac{d\sin\theta}{c} \tag{4.17}$$

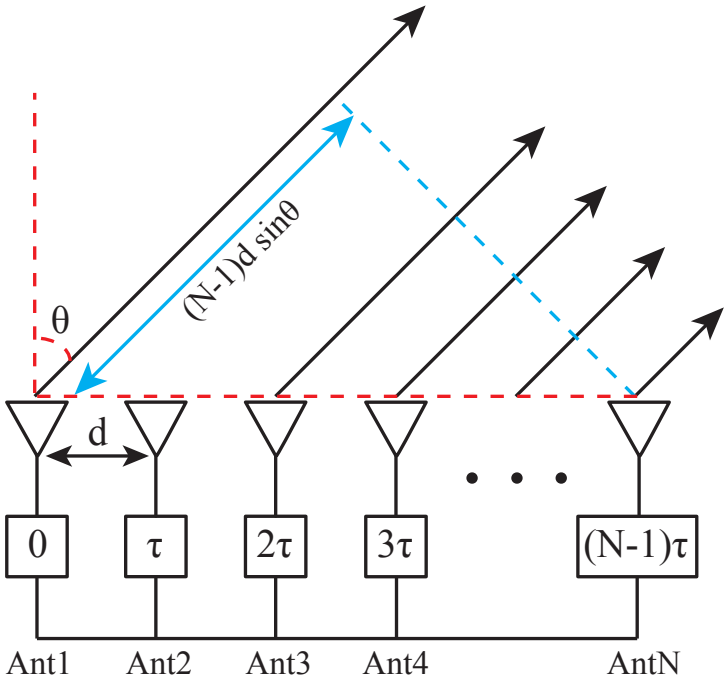


Figure 4.2: N-element timed array.

Under this condition, the maximum summed E-field magnitude is N times larger than the E-field magnitude of a single antenna. Since the power is proportional to the E-field magnitude square, the maximum radiation intensity is N^2 times larger than that of a single antenna. In other words, EIRP of an N-element beamforming array is N^2 times larger than the EIRP of a single antenna. The E-field magnitude can be expressed as a function of spatial angle θ for a fixed delay τ,

$$|E_{array}(\theta)| = \left| \frac{\sin[\frac{1}{2}N\omega_0(\frac{d\sin\theta}{c} - \tau)]}{\sin[\frac{1}{2}\omega_0(\frac{d\sin\theta}{c} - \tau)]} \right| \tag{4.18}$$

The peak directivity of an N-element array can be calculated,

$$D_{array,max} = \frac{U_{max}}{U_0} = \frac{E_{max}^2}{E_0^2} = \frac{\frac{1}{N}N^2}{1} = N \tag{4.19}$$

The factor $\frac{1}{N}$ in the numerator comes from the fact that each element gets one-Nth of the input power. As a result, the peak directivity of an N-element antenna array is N times larger than that of a single antenna.

Programmable time delays are usually difficult to implement in a compact fashion. Instead, most beamforming arrays are actually phased arrays. For narrow band signals, a time delay can be approximated by a phase shift. Fig. 4.3 shows a N-element phased array. In a phased array, the signals that excite the antennas are progressively phase shifted: the signal

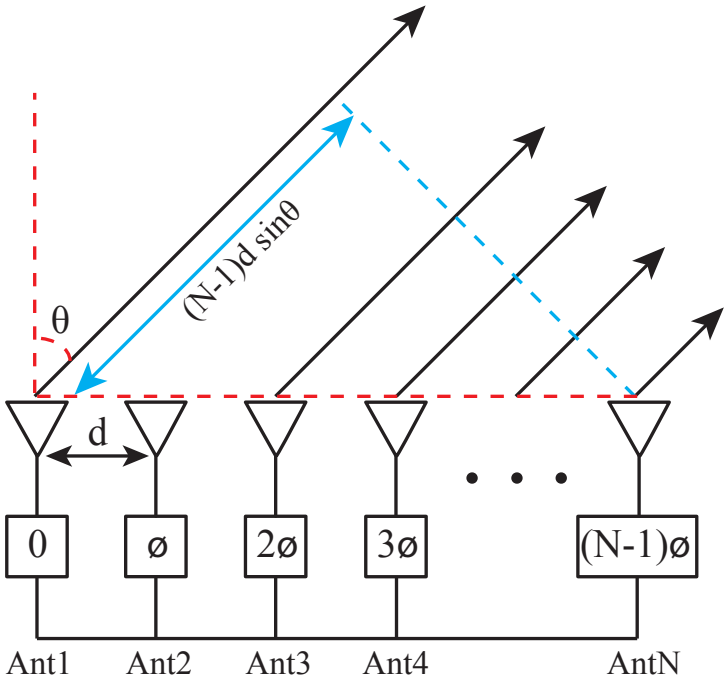


Figure 4.3: N-element phased array.

inside the n^{th} element is phase shifted by $(n-1)\phi$. The E-field vector contributed by the n^{th} antenna is,

$$E_n = e^{j\omega_0[t-t_n-(n-1)\frac{\phi}{\omega_0}]} \tag{4.20}$$

$$= e^{j\omega_0[t-(N-1)t_0]} e^{j\omega_0[(n-1)(t_0-\frac{\phi}{\omega_0})]} \tag{4.21}$$

The sum of all the E-field vectors is,

$$E_{array} = e^{j\omega_0[t-(N-1)t_0]} e^{j\omega_0 \frac{(N-1)(t_0-\frac{\phi}{\omega_0})}{2}} \frac{\sin[\frac{1}{2}N(\omega_0 t_0-\phi)]}{\sin[\frac{1}{2}(\omega_0 t_0-\phi)]} \tag{4.22}$$

The magnitude of the summed E-field vector is,

$$|E_{array}| = \left| \frac{\sin[\frac{1}{2}N(\omega_0 t_0-\phi)]}{\sin[\frac{1}{2}(\omega_0 t_0-\phi)]} \right| \tag{4.23}$$

Similar to Eq. 4.17, the summed E-field magnitude can be maximized at spatial angle θ by setting $\omega_0 t_0 = \phi$,

$$\phi_{opt} = \omega t_0 = \frac{\omega_0 d \sin\theta}{c} \tag{4.24}$$

Finally the E-field radiation pattern can be found for a fixed phase shift ϕ,

$$|E_{array}(\theta)| = \left| \frac{\sin[\frac{1}{2}N(\omega_0 \frac{d\sin\theta}{c}-\phi)]}{\sin[\frac{1}{2}(\omega_0 \frac{d\sin\theta}{c}-\phi)]} \right| \tag{4.25}$$

Fig. 4.4 plots the E-field magnitude pattern for an 8 element array and a 32 element array, with ϕ set to direct the beam at different angles: $-30°$, $0°$ and $30°$. Apparently larger arrays have larger directivity and therefore generate shaper beams.

For the same number of elements, the antenna beam pattern is also affected by the element spacing. Fig. 4.4 assumes half wavelength spacing, which is the maximum limit to avoid large undesired lobes. To illustrate this point, Fig. 4.5 shows the E-field pattern for an 8-element array with different antenna spacing. When the spacing exceeds half wavelength, there are multiple side lobes with the same magnitude as the main lobe. These side lobes are called grating lobes. Grating lobes are generally undesired since they steer energy into unwanted direction in a transmitter while pick up interference from other angles in a receiver. As evident in Fig. 4.5, larger antenna spacings result in larger numbers of grating lobes. The number of grating lobes is roughly equal to the ratio of the spacing to the half wavelength.

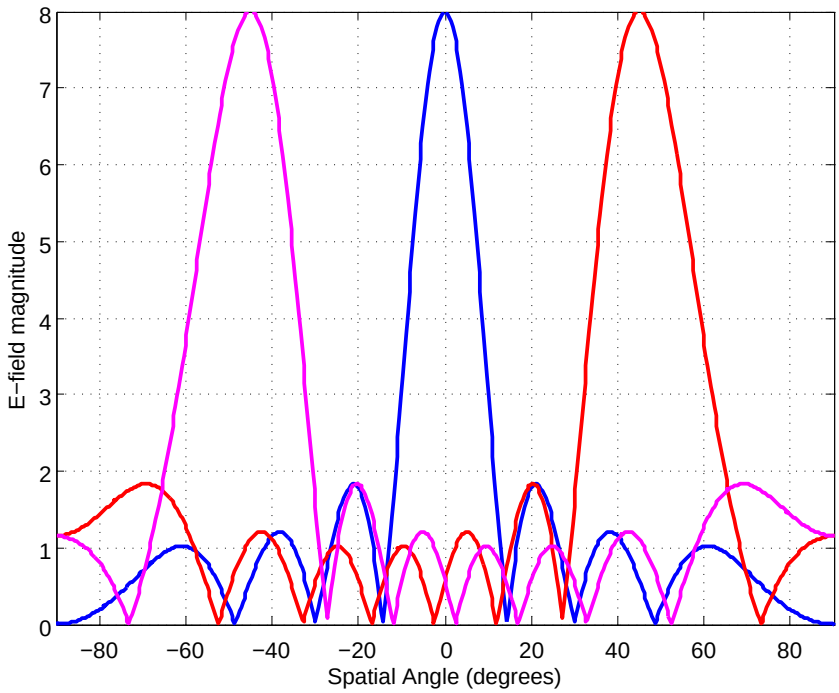
4.3 Phased Array Architectures

There are multiple radio architectures for implementing a phased array, depending on where the phase shift occurs on the signal chain. In a direct conversion transmitter, the phase of the signal can be adjusted at either the RF domain, LO domain, analog baseband domain or digital baseband domain. They correspond to four different phased array architectures, as illustrated in Fig. 4.6. Each architecture has its own advantages and disadvantages, and the optimal architecture is usually application specific, taking into account of various considerations such as frequency, area, power, interference and robustness.

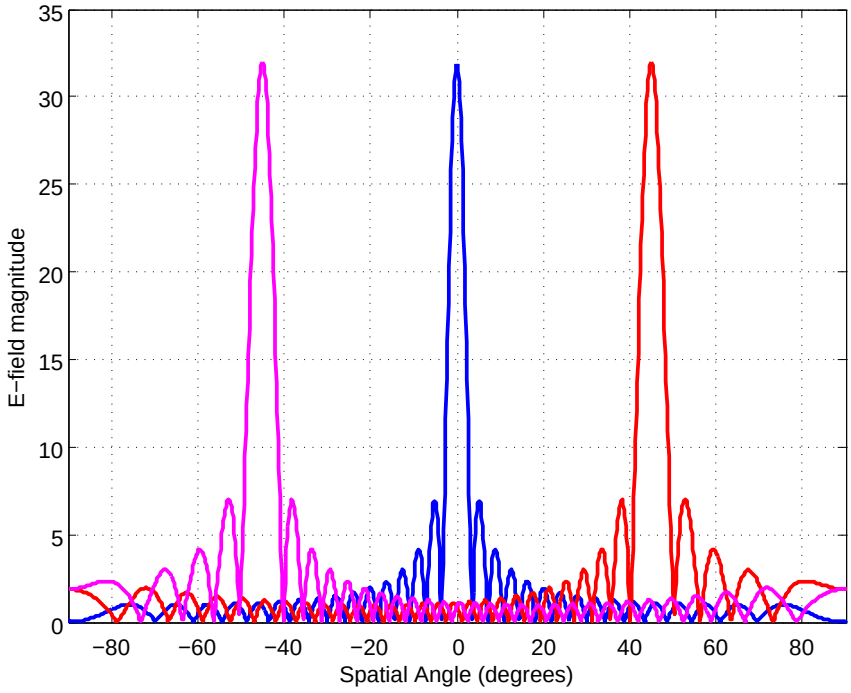
4.3.1 RF Phase Shifting

Shown in Fig. 4.6a, RF phase shifting architecture adjusts the relative phases of different antenna elements in the RF signal domain. Traditionally, the biggest advantage of this architecture is that it requires the least number of circuit components, and therefore potentially the lowest cost and power. In cases of discrete implementation, due to limited routing flexibility and cost of discrete components, RF phase shifting architecture is the most widely used architecture. Since only one mixer is used, there's no need to distribute the LO signal and therefore the noise coupling can be minimized on the LO. In the case of a receiver, another advantage is the relaxed mixer linearity requirement. This is because the signals are combined in the RF domain before entering the mixer, as a result, interference coming from the null direction of the beam will be canceled out during the signal summation, a phenomenon called spatial filtering.

However, phase shifters are usually difficult to implement at high frequencies, especially at mm-wave range. Most mm-wave phase shifters are built from variable passive components and are therefore lossy [20]. In addition, the phase shifter loss is not constant over different phase settings, thus requiring an amplifier to compensate the gain variation [21, 22]. In

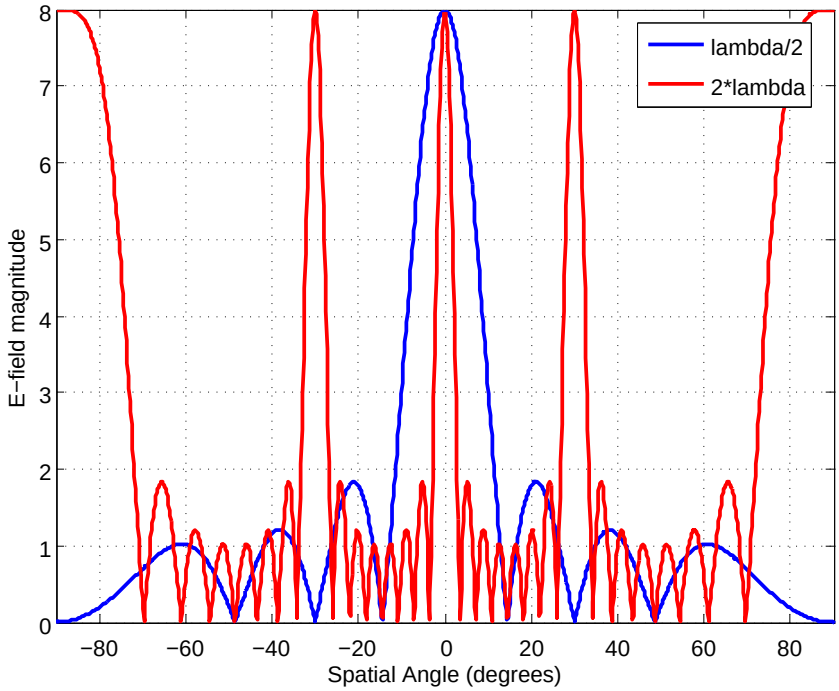


(a) 8-element array

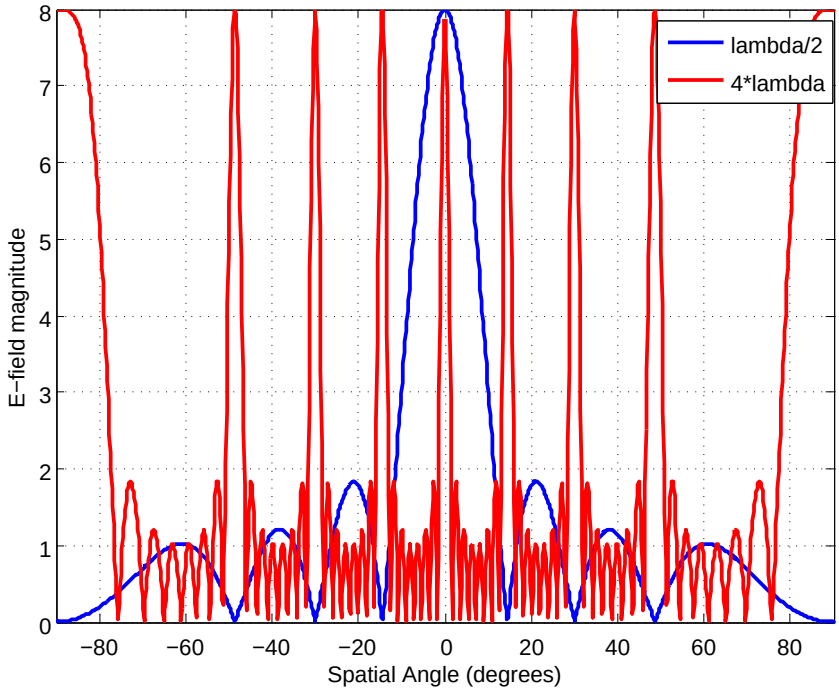


(b) 32-element array

Figure 4.4: Phased array E-field magnitude patterns



(a) Half wavelength spacing and two wavelengths spacing



(b) Half wavelength spacing and four wavelengths spacing

Figure 4.5: Phased array E-field magnitude patterns with various element spacings

addition, the phase shift must maintain the required signal bandwidth as well as linearity since it resides in the signal path.

4.3.2 LO Phase Shifting

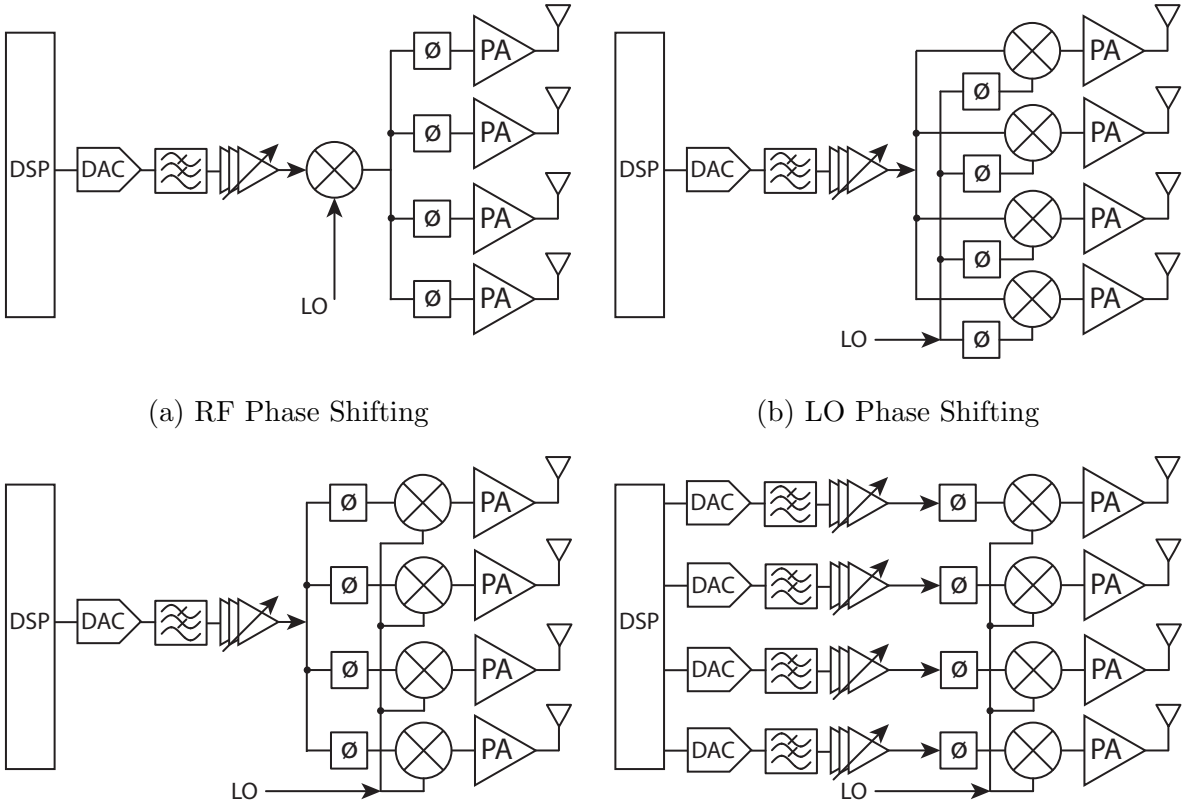
Shown in Fig. 4.6b, LO phase shifting architecture adjusts the relative phases of different antenna element in the LO signal domain. Unlike RF phase shifting architecture, the phase shift function is not implemented in the main signal path, but instead on a constant envelope CW signal. As a result, there's no bandwidth or linearity requirement on the phase shifter. The LO phase shifting architecture requires more circuit blocks, mainly additional mixers, one for each element. As a result, it seems like such architecture will inevitably incur larger power consumption. However, this is not case when the mixer sizing is down scaled accordingly, since each mixer needs to drive only one PA in this architecture whereas the mixer in the RF phase-shifting architecture needs to drive all the PAs unless RF drivers are added.

Similar to RF phase shifting, the main disadvantage of the LO phase shifting architecture is that it requires high frequency phase shifters. In addition, in the case of a receiver, since the signals are combined after down conversion, spatial filtering doesn't occur before the mixer. Therefore, it imposes more stringent requirements on the mixer linearity to handle large signal interferences.

4.3.3 Analog Baseband Phase Shifting

Shown in Fig. 4.6c, analog baseband (BB) phase shifting architecture adjusts the relative phases of different antenna elements in the analog baseband domain. The main advantage of this architecture is that it doesn't need a high frequency phase shifter. Phase shifting at low frequency can be realized with higher resolution, lower power and smaller footprint. Baseband phase shifters are usually implemented using active devices based on I/Q interpolation principles. Unlike passive phase shifters, active phase shifters have very small gain variation across different phase states. Typical analog baseband phase shifters have resolutions higher than 5 bits with gain variation less than 1dB [23, 24]. In the case of a receiver, since the phase shifter loss is eliminated in the RF domain, the noise figure of this architecture is typically smaller than that of a RF phase shifting architecture with the same signal chain power.

Similar to the LO phase shifting architecture, since the LO signal needs to be distributed to the local elements, it's more sensitive to noise. Special care has to be taken to isolate the LO distribution path from noisy circuitry such as the digital signals. Besides, spatial filtering occurs in the analog baseband domain, therefore requiring better mixer linearity.



(a) RF Phase Shifting

(b) LO Phase Shifting

(c) Analog Baseband Phase Shifting

(d) Digital Baseband Phase Shifting

Figure 4.6: Phased array architecutres (Transmitter)

4.3.4 Digital Baseband Phase Shifting

The final architecture is the digital baseband phase shifting architecture in which phase shifting function occurs in the digital domain, as shown in Fig. 4.6d. This architecture provides the greatest flexibility since baseband DSP allows complex beamforming algorithms to be implemented with high accuracy. It also enables spatial multiplexing using Multiple-Input Multiple-Output (MIMO) for higher data rates, which is widely used in 802.11 standards.

However, this architecture suffers from several major disadvantages. First, since the beamforming is implemented in the digital domain, a high resolution DAC/ADC is needed. This is especially unattractive, since most mm-wave radios have data rates in the range of Gb/s and therefore requires GS/s DAC/ADCs. Such high speed data converters are extremely power hungry at high resolution [25]. Even state-of-the-art time interleaved Successive Approximation Register (SAR) ADC has an energy efficiency of 50fJ/conv-step [26], much higher than MS/s range ADCs used for Wi-Fi where energy efficiency is in the order of 10fJ/conv-step. Second, digital baseband phase shifting architecture requires N copies of the entire radio front-end, which increases power and silicon area. Finally the spatial interference signal exists throughout the entire receiver chain, as a result, the linearity requirement extends to the entire analog baseband chain.

4.4 Architecture Power Comparison

Besides the aforementioned performance tradeoffs among different architectures, the most critical factor to consider when choosing the most appropriate architecture is the overall power consumption, which depends on a number of factors such as number of elements, frequency and process. A simplified model is used here to predict the overall power consumption of different architectures.

The analysis starts with a single-element transmitter which consists of a PA, a mixer and a VCO, as shown in Fig. 4.7. The PA delivers an output power of $P_{o,PA}$ and needs a input drive power of $P_{i,PA}$ from the mixer. The VCO drives the mixer with an LO power of P_{LO}. The power consumption of this transmitter is,

$$P_{singleTX} = P_{PA} + P_{mix} + P_{VCO} \tag{4.26}$$

$$= \frac{P_{o,PA}}{\eta_{PA}} + \frac{P_{i,PA}}{\eta_{mix}} + P_{VCO} \tag{4.27}$$

where η_{PA} and η_{mix} are the drain efficiencies of the PA and the mixer respectively. The choice of $P_{i,PA}$ is important for minimizing the overall power consumption. Since mixers generally have lower drain efficiency than amplifiers due to reduced voltage headroom and increased parasitic capacitances, $P_{i,PA}$ should be minimized as much as possible, which means the PA needs to have multiple stages. On the other hand, due to the finite Q of on-chip inductors, the realizable parallel impedance of an inductor is bounded. This means the mixer power consumption cannot be indefinitely down-scaled with $P_{i,PA}$. As a result, an optimal $P_{i,PA}$

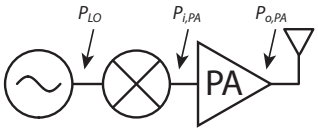
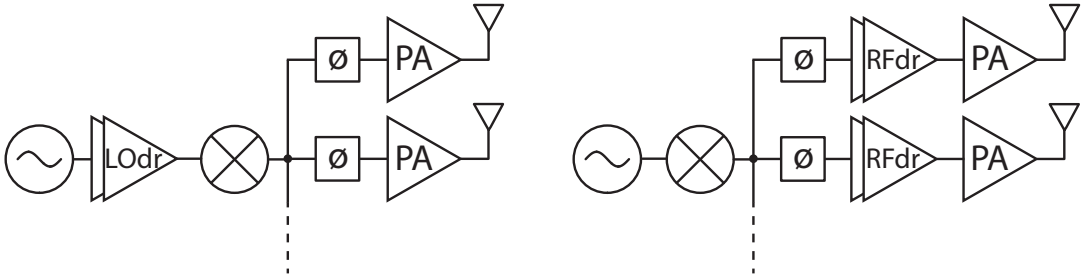


Figure 4.7: Single element transmitter.



(a) Implementation A

(b) Implementation B

Figure 4.8: RF phase shifting.

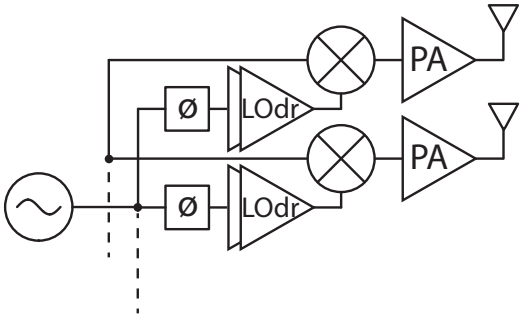


Figure 4.9: LO phase shifting.

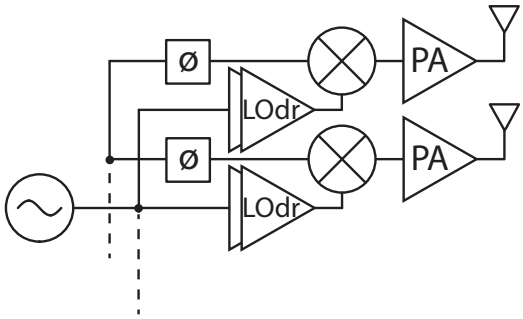


Figure 4.10: Analog baseband phase shifting.

and mixer size exist for a given technology and operating frequency. It's assumed that the $P_{i,PA}$ used in the equation represents such an optimal value.

The following subsections derive the power consumption of the RF blocks for RF, LO and ABB phase shifting architectures. The DBB phase shifting architecture has the same RF power consumption as the ABB phase shifting architecture and therefore is not repeated in the derivation.

4.4.1 RF Phase Shifting

There are two ways of extending a single-element transmitter into a multi-element RF phase shifting transmitter, as shown in Fig. 4.8. In implementation A, the mixer is up-sized to provide larger output power in order to drive multiple elements. In implementation B, the mixer is kept unchanged, instead drivers are added in each element to compensate the power splitting loss and to drive the PA. The RF phase shifters are assumed passive devices and have a power gain of G_{PS} ($G_{PS} < 1$).

In implementation A, the upsized mixer needs to deliver a total output power of $NP_{i,PA}/G_{PS}$. As a result, the mixer power consumption is,

$$P_{mix} = \frac{P_{o,mix}}{\eta_{mix}} \tag{4.28}$$

$$= N\frac{P_{i,PA}}{G_{PS}\eta_{mix}} \tag{4.29}$$

where N is the number of array elements. Since the mixer is upsized, LO drivers are needed after the VCO to provide additional LO power to drive the mixer, the power of the LO drivers can be found according the output power and gain requirement,

$$P_{LOdr} = \frac{P_{LO}\frac{N}{G_{PS}}}{\eta_{amp,non}} \tag{4.30}$$

where $\eta_{amp,non}$ represents the drain efficiency of RF (or LO) amplifiers. Note the subnote *non* indicates that linearity is not required in this amplifier since it's in the LO path, as a result, $\eta_{amp,non}$ can be higher than $\eta_{amp,lin}$. Given the relation between PAE and drain efficiency defined in Eq. 2.17, Eq. 4.30 can be re-arranged to,

$$P_{LOdr} = \frac{P_{LO}N}{G_{PS}PAE_{amp,non}}\frac{G_{LOdr}-1}{G_{LOdr}} \tag{4.31}$$

$$= \frac{P_{LO}}{PAE_{amp,non}}\frac{N-G_{PS}}{G_{PS}} \tag{4.32}$$

$$= \frac{P_{i,PA}}{PAE_{amp,non}}\frac{N-G_{PS}}{G_{PS}G_{mix}} \tag{4.33}$$

where,

$$G_{LOdr} = \frac{N}{G_{PS}} \tag{4.34}$$

and G_{mix} is the power gain of the mixer. Note that PAE is used instead of drain efficiency because it's independent of the number of amplifier stages as shown in Eq. 2.21. Thus, the total power consumption of implementation A is,

$$P_{TX,RFPS-A} = P_{mix} + P_{LOdr} + NP_{PA} \tag{4.35}$$

$$= \frac{P_{i,PA}}{\eta_{mix}}\frac{N}{G_{PS}} + (N - G_{PS})\frac{P_{i,PA}}{PAE_{amp,non}}\frac{1}{G_{mix}G_{PS}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.36}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N - G_{PS})\frac{P_{i,PA}}{\eta_{mix}G_{PS}} + (N - G_{PS})\frac{P_{i,PA}}{PAE_{amp,non}}\frac{1}{G_{mix}G_{PS}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.37}$$

In implementation B, the mixer power consumption is kept unchanged, and the additional power comes from the RF drivers in each element. The power consumption of each driver can be found based on the output power and the gain requirement,

$$P_{RFdr} = \frac{P_{i,PA}}{\eta_{amp,lin}} \tag{4.38}$$

$$= \frac{P_{i,PA}}{PAE_{amp,lin}}\frac{G_{RFdr} - 1}{G_{RFdr}} \tag{4.39}$$

$$= \frac{P_{i,PA}}{PAE_{amp,lin}}\frac{N - G_{PS}}{N} \tag{4.40}$$

where

$$G_{RFdr} = \frac{N}{G_{PS}} \tag{4.41}$$

Therefore the total power consumption of implementation B is,

$$P_{TX,RFPS-B} = P_{mix} + NP_{RFdr} + NP_{PA} \tag{4.42}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N - G_{PS})\frac{P_{i,PA}}{PAE_{amp,lin}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.43}$$

Comparing Eq. 4.37 and Eq. 4.43, it's not difficult to conclude that implementation B is more power efficiency, since $G_{PS}\eta_{mix}$ is most likely to be smaller than $PAE_{amp,lin}$. This is because the efficiency of a mixer is usually smaller than that of an amplifier and the gain of the phase shifter is usually much smaller than one. Besides, Eq. 4.37 has an extra third term which does not exist in Eq. 4.43. The intuition behind the finding is that more efficient blocks such as amplifiers should be used to provide driving capability in order to lower the overall power consumption.

Note that although passive RF phase shifters are assumed for calculation, the results also apply when active phase shifters are used. Active phase shifters are typically made of weighted I/Q summation and can always be separated into the gain stage and the summation stage. The gain stage can be lumped into RF drivers while the summation stage is passive and lossy. The I/Q summation typically introduces 3dB insertion loss.

4.4.2 LO Phase Shifting

In the LO phase shifting architecture, as shown in Fig. 4.9, the mixer and PA in each element remains unchanged while LO drivers are required to compensate for the LO signal distribution and phase shifter loss. The power consumption of each LO driver is,

$$P_{LOdr} = \frac{P_{LO}}{\eta_{amp,non}} \tag{4.44}$$

$$= \frac{P_{LO}}{PAE_{amp,non}} \frac{G_{LOdr} - 1}{G_{LOdr}} \tag{4.45}$$

$$= \frac{P_{LO}}{PAE_{amp,non}} \frac{N - G_{PS}}{N} \tag{4.46}$$

where

$$G_{LOdr} = \frac{N}{G_{PS}} \tag{4.47}$$

The total power consumption of this architecture is,

$$P_{TX,LOPS} = NP_{mix} + NP_{LOdr} + NP_{PA} \tag{4.48}$$

$$= N\frac{P_{i,PA}}{\eta_{mix}} + (N - G_{PS})\frac{P_{LO}}{PAE_{amp,non}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.49}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{i,PA}}{\eta_{mix}} + (N - G_{PS})\frac{P_{i,PA}}{G_{mix}PAE_{amp,non}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.50}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{i,PA}}{PAE_{mix}}\frac{G_{mix} - 1}{G_{mix}} + (N - G_{PS})\frac{P_{i,PA}}{G_{mix}PAE_{amp,non}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.51}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{i,PA}}{PAE_{amp,non}}\frac{1}{G_{mix}}[\frac{N - G_{PS}}{N-1} + \frac{PAE_{amp,non}}{PAE_{mix}}(G_{mix} - 1)] + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.52}$$

where PAE_{mix} is the LO to RF power added efficiency of the mixer.

It's relative hard to conclude whether LO phase shifting is better than RF phase shifting by directly comparing Eq. 4.52 to Eq. 4.43, however, it will be shown that LO phase shifting architecture is almost always inferior to Anlog BB phase shifting architecture.

4.4.3 Analog Baseband Phase Shifting

The transmitter block diagram for analog baseband phase shifting is very similar to that of LO phase shifting, as shown in Fig. 4.10, with the only difference being that the LO phase shifter is now moved to the baseband path. As a result, it's not necessary to repeat the

derivation again. Instead, the total power consumption can be readily obtained by setting the phase shifter gain to unity $G_{PS} = 1$,

$$P_{TX,ABBPS} = NP_{mix} + NP_{LOdr} + NP_{PA} \tag{4.53}$$

$$= N\frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{LO}}{PAE_{amp,non}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.54}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{i,PA}}{G_{mix}PAE_{amp,non}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.55}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{i,PA}}{PAE_{mix}}\frac{G_{mix}-1}{G_{mix}} + (N-1)\frac{P_{i,PA}}{G_{mix}PAE_{amp,non}} + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.56}$$

$$= \frac{P_{i,PA}}{\eta_{mix}} + (N-1)\frac{P_{i,PA}}{PAE_{amp,non}}\frac{1}{G_{mix}}[1 + \frac{PAE_{amp,non}}{PAE_{mix}}(G_{mix}-1)] + N\frac{P_{o,PA}}{\eta_{PA}} \tag{4.57}$$

Comparing Eq. 4.57 to Eq. 4.52, it can be seen that the RF power consumption of analog baseband phase shifting architecture is always smaller than that of the LO phase shifting architecture due to absence of phase shifter loss. Note that the baseband phase shifter power is not included in the comparison and such omission is generally acceptable since baseband circuit blocks typically consume much lower power than RF blocks.

Comparing analog baseband phase shifting architecture and RF phase shifting architecture is less straightforward. Note that the first and third terms in Eq. 4.43 to Eq. 4.57 are identical, therefore only the second term needs to be compared. In fact the second term represents the overhead RF power when scaling the transmitter from single element to N elements,

$$\Delta P_{TX,RFPS-B} = (N - G_{PS})\frac{P_{i,PA}}{PAE_{amp,lin}} \tag{4.58}$$

$$\Delta P_{TX,ABBPS} = (N-1)\frac{P_{i,PA}}{PAE_{amp,non}}\frac{1}{G_{mix}}[1 + \frac{PAE_{amp,non}}{PAE_{mix}}(G_{mix}-1)] \tag{4.59}$$

When the mixer gain G_{mix} is much greater than unity, Eq. 4.59 can be simplified to,

$$\Delta P_{TX,ABBPS} = (N-1)\frac{P_{i,PA}}{PAE_{mix}} \tag{4.60}$$

Since the amplifier PAE is typically larger than the mixer PAE, it's almost certain that for reasonably large arrays, the RF phase shifting is more power efficient. This means for low frequency applications, the RF phase shifting architecture is the best candidate. However, if the ratio of $PAE_{amp,lin}$ to PAE_{mix} is less than two, it's possible that the analog baseband phase shifting architecture is more efficient for small arrays.

4.4.4 Generalized Comparison

To provide more insights on the power comparison, the overhead power of the three architectures is plotted for different array sizes based on different G_{mix} and $\frac{PAE_{amp,lin}}{PAE_{mix}}$ values. With a mixer gain of 2 ($G_{mix} = 2$), if an amplifier is three times more efficient than a mixer ($\frac{PAE_{amp,lin}}{PAE_{mix}} = 3$), the RF phase shifting is always better than baseband phase shifting, no matter how many elements the array has, as shown in Fig. 4.11. If an amplifier is two times more efficient than a mixer, the baseband phase shifting architecture is better for array sizes smaller than 5, as shown in Fig. 4.12. If an amplifier is only 1.75 times more efficient than a mixer, the baseband phase shifting architecture remains better for array sizes up to 8. As a result, the smaller $\frac{PAE_{amp,lin}}{PAE_{mix}}$ is, the more efficient analog baseband phase shifting architecture is. On the other hand, the larger the mixer gain is, the less attractive the analog baseband phase shifting architecture becomes, as evident when comparing Fig. 4.13 and Fig. 4.14. When the mixer gain is doubled, the crossover point for the RF phase shifting and baseband phase shifting drops from 8 to 3. In short, the slope of the overhead power for analog baseband phase shifting architecture increases with mixer gain G_{mix} as well as the efficiency ratio of amplifiers and mixer $\frac{PAE_{amp,lin}}{PAE_{mix}}$. For large array sizes, RF phase shifting is always more power efficient than baseband phase shifting since slope of overhead power for analog baseband phase shifting is always larger than that of RF phase shifting.

4.5 Phase Shifters

4.5.1 Resolution

The minimum phase shifter resolution is determined by the directivity of the array. Fig. 4.15 to Fig. 4.18 show the obtainable array gain as a function of the beamforming angles with different phase shifter resolutions for different array sizes. At each beamforming angle, the phase setting of each element is chosen such that the total array gain is maximized at that angle. It can be seen that 2-bit phase shifter resolution is already sufficient to obtain an array gain within 1dB range of the theoretical maximum. 3-bit resolution closes the gap to less than 0.3dB. Further increasing the resolution has diminishing returns in terms of array gain improvement. As a result, 4-bit might be enough for any array size from the perspective of array gain. On the other hand, phase shifter resolution also affects the sidelobe level and the peak-to-null ratio. Fig. 4.19 and Fig. 4.20 show the maximum sidelobe level and minimum peak-to-null ratio as a function of beamforming angles for a 4-element array. Again, the phase setting for each element is chosen to optimize the array gain at each beamforming angle. Clearly, higher phase shifter resolution reduces the maximum sidelobe and increases the minimum peak-to-null ratio, which is highly desired to minimize interference and multi-path fading. These improvements are more significant than array gain improvement at higher phase shifter resolutions.

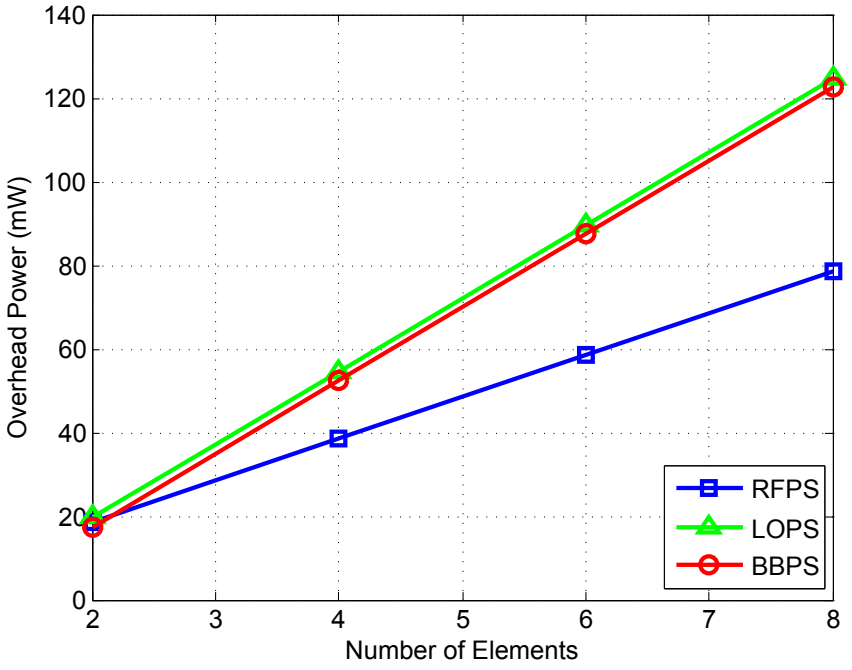


Figure 4.11: Overhead power ($G_{mix} = 2, \frac{PAE_{amp,lin}}{PAE_{mix}} = 3, \frac{PAE_{amp,lin}}{PAE_{amp,non}} = 2, G_{PS} = 1/8$)

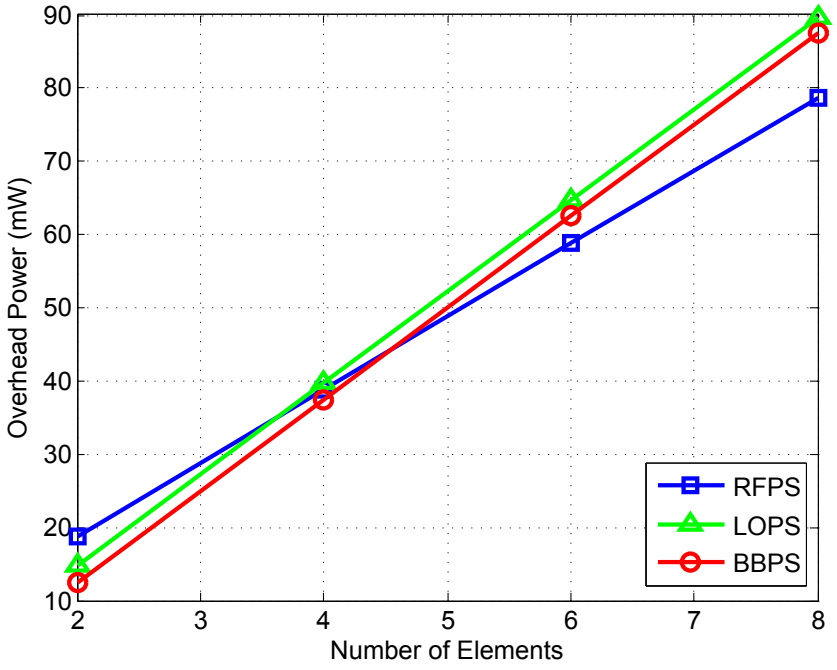


Figure 4.12: Overhead power ($G_{mix} = 2, \frac{PAE_{amp,lin}}{PAE_{mix}} = 2, \frac{PAE_{amp,lin}}{PAE_{amp,non}} = 2, G_{PS} = 1/8$)

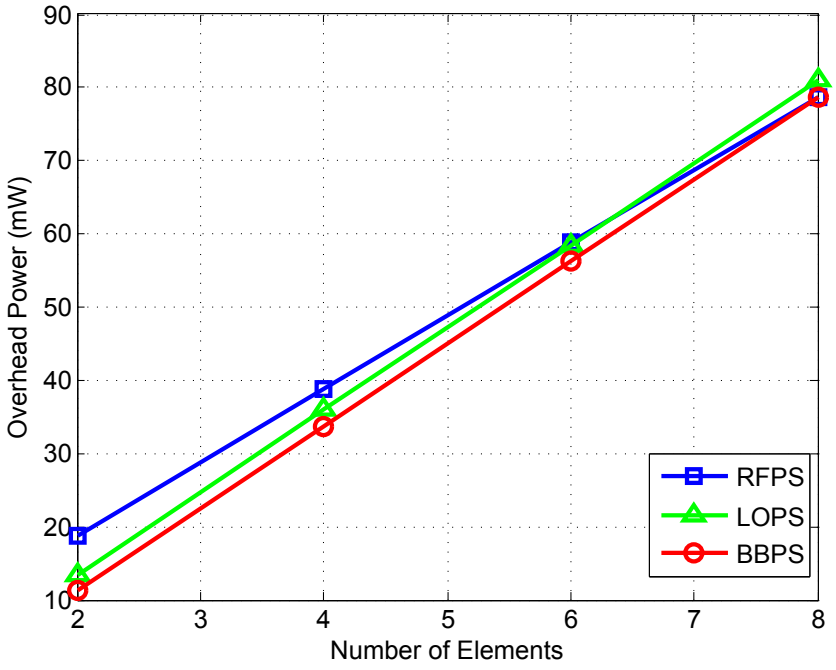


Figure 4.13: Overhead power ($G_{mix} = 2, \frac{PAE_{amp,lin}}{PAE_{mix}} = 1.75, \frac{PAE_{amp,lin}}{PAE_{amp,non}} = 2, G_{PS} = 1/8$)

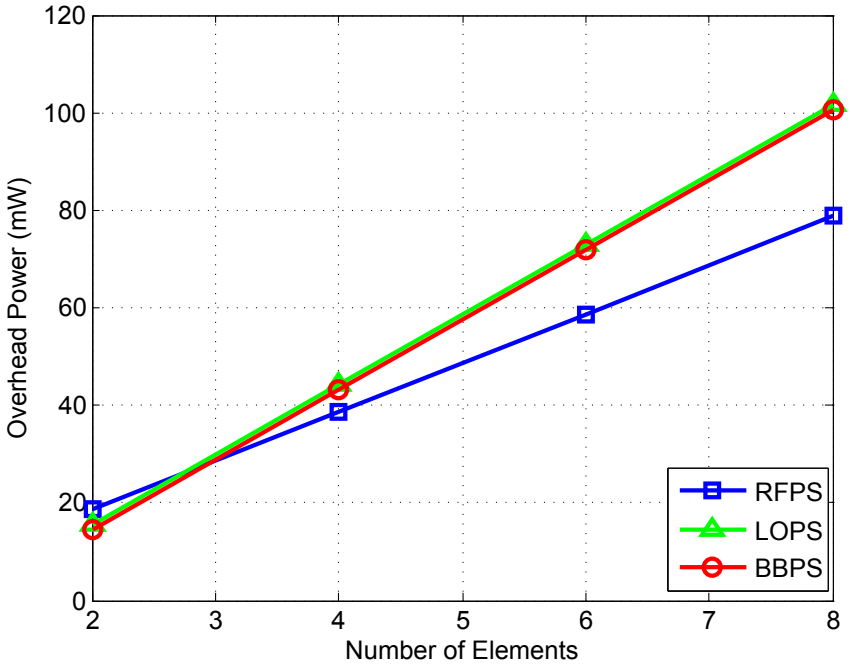


Figure 4.14: Overhead power ($G_{mix} = 4, \frac{PAE_{amp,lin}}{PAE_{mix}} = 1.75, \frac{PAE_{amp,lin}}{PAE_{amp,non}} = 2, G_{PS} = 1/8$)

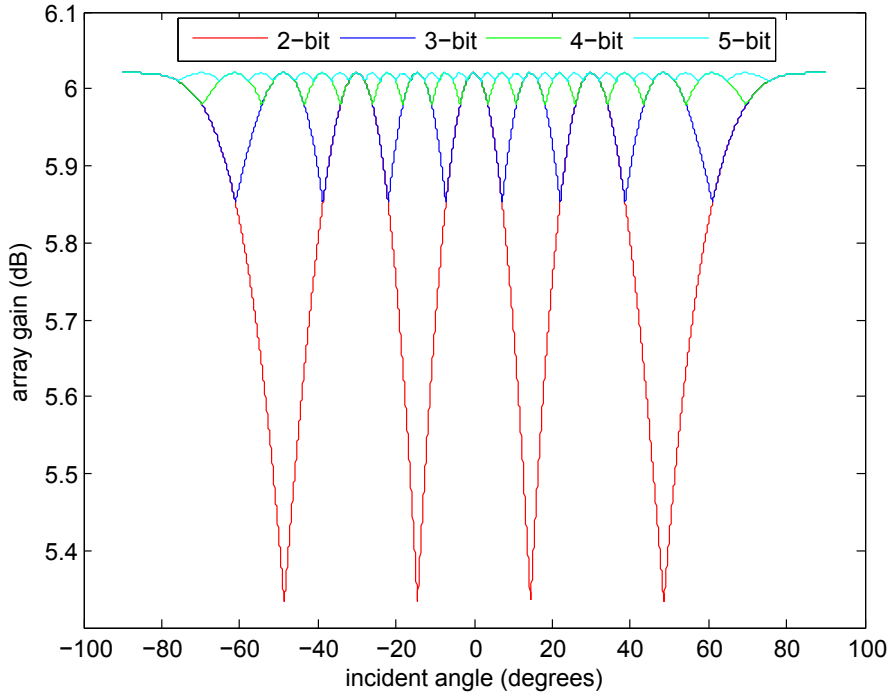


Figure 4.15: Array gain as a function of beamforming angle (2 elements)

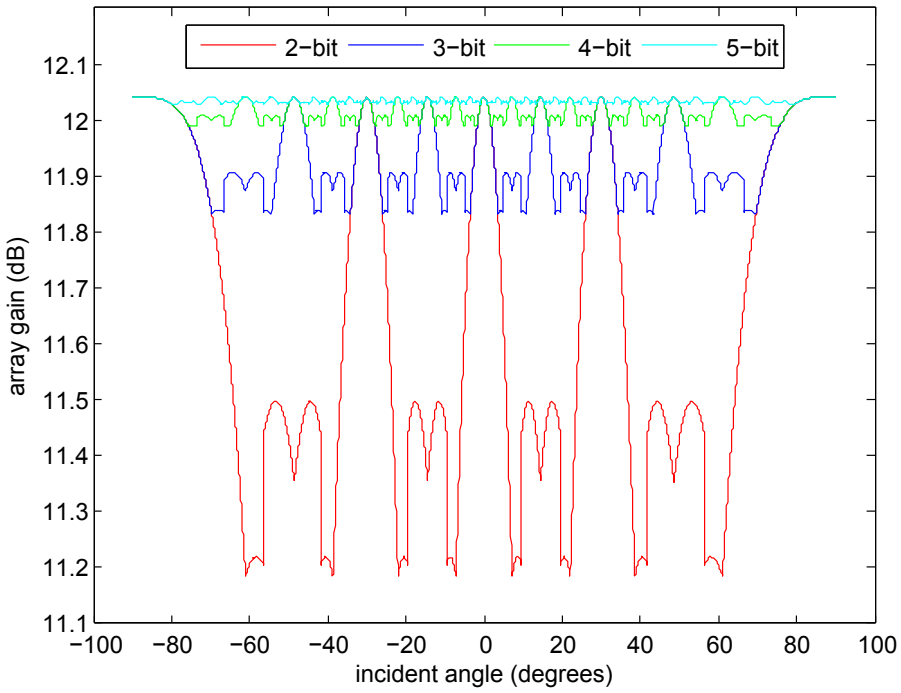


Figure 4.16: Array gain as a function of beamforming angle (4 elements)

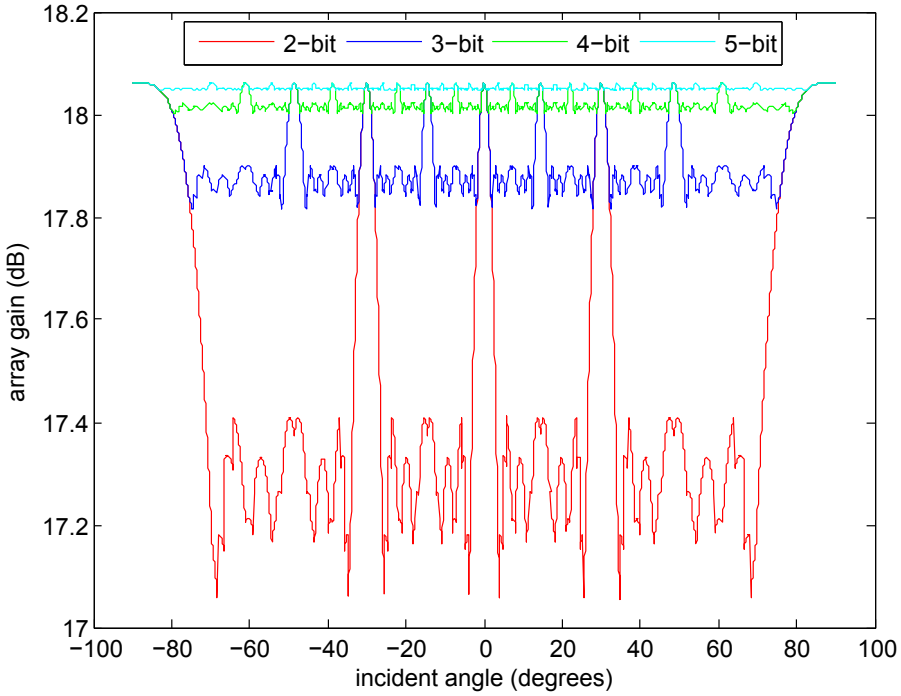


Figure 4.17: Array gain as a function of beamforming angle (8 elements)

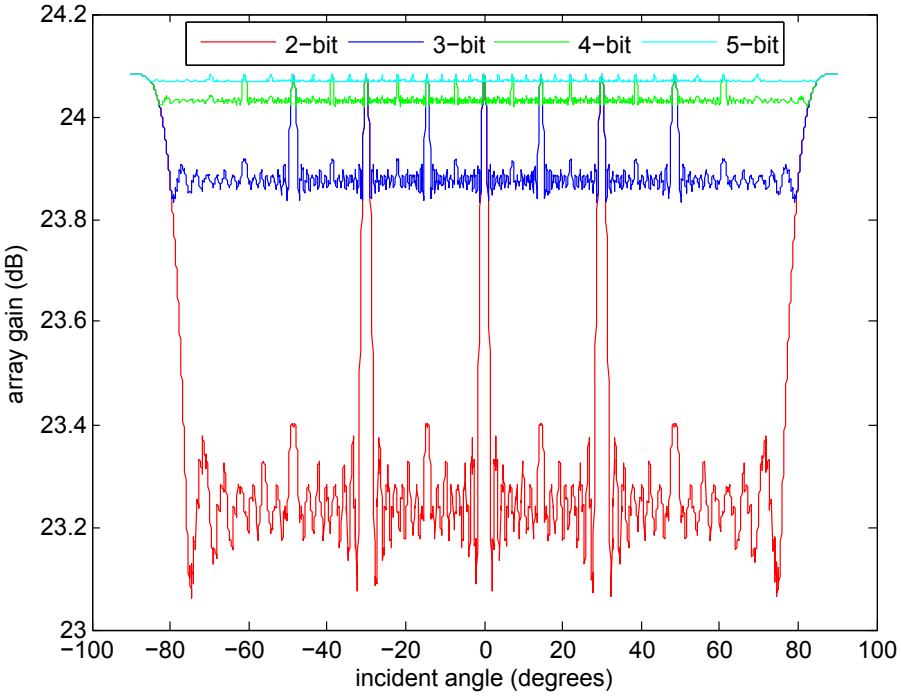


Figure 4.18: Array gain as a function of beamforming angle (16 elements)

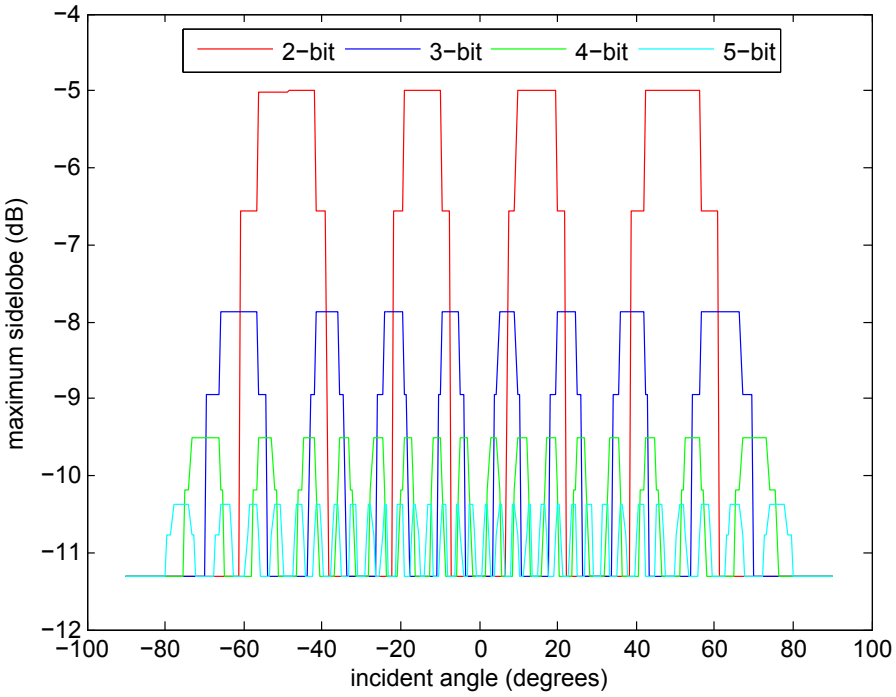


Figure 4.19: Maximum sidelode as a function of beamforming angle (4 elements)

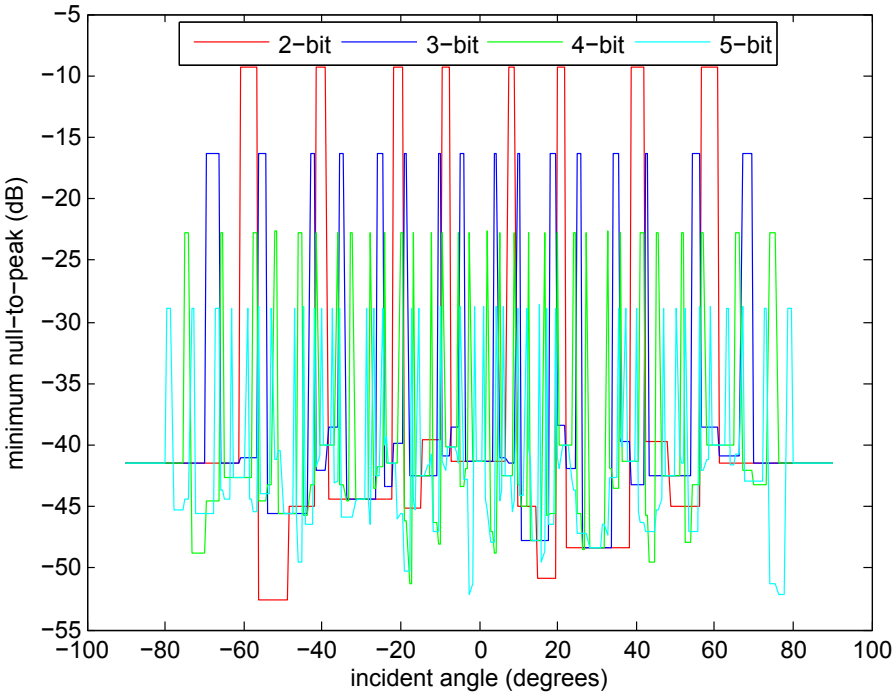


Figure 4.20: Minimum peak-to-null ratio as a function of beamforming angle (4 elements)

4.5.2 Implementation

Depending on the frequency, phase shifters can be either passive or active. Passive phase shifters are mainly used in the RF signal domain. In general, passive phase shifters can be classified into two categories: through type phase shifters and reflective type phase shifters. Through type phase shifters are essentially tunable LC delay structures, as shown in Fig. 4.21. The insertion time delay and phase shift are,

$$t_D = \sqrt{L_{tot}C_{tot}} \tag{4.61}$$

$$\phi = \omega\sqrt{L_{tot}C_{tot}} \tag{4.62}$$

The simplest way to implement such a phase shifter is to have multiple T-lines with different lengths (Fig. 4.21a). The input and output are switched between different T-lines for different phase shift. Obviously such an approach can only have very low resolution or range unless large silicon area is used. A better approach is to synthesize a T-line using variable capacitors and inductors (Fig. 4.21b). Traditionally, only variable capacitors are used since it can be easily implemented by varactors or switched capacitors while variable inductors are much more difficult to implement. However, this changes the characteristic impedance of the T-line and thus reduces the bandwidth. Besides it also limits the obtainable phase shift range. Recently variable inductors were also demonstrated using inductance multiplication techniques [27]. Although only 1-bit variable inductor was implemented, it significantly extended the phase shift range and preserved the wide band characteristic of a T-line.

Reflective type phase shifters exploit the reflection coefficient property of a LC resonance tank. The reflection coefficient of a parallel LC tank has a constant magnitude of unity, but a varying phase depending on the offset from the resonance frequency. By loading a quadrature hybrid with two LC tanks with variable resonance frequency (Fig. 4.22), the signal going through the hybrid will experience variable phase shifts. The phase response of this structure is,

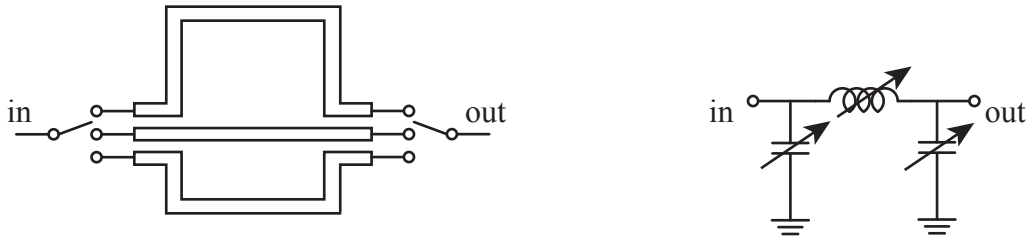
$$\phi \approx -2\arctan(2\frac{Z_0}{Z_{0,tank}}\frac{\delta\omega}{\omega_0}) \tag{4.63}$$

$$\omega_0 = \frac{1}{\sqrt{LC_0}} \tag{4.64}$$

$$Z_{0,tank} = \sqrt{L/C_0} \tag{4.65}$$

where C_0 and $Z_{0,tank}$ are the nominal capacitance and impedance of the tank, and $\delta\omega$ is shift in tank resonance frequency when the capacitance is varied. Although such a structure can provide up to 360° total phase shift in theory, the amount of variable capacitance that can be implemented usually limits the total range. As a result, two or more stages are usually cascaded to cover 360° [28, 27]. Alternatively, an active stage can be used to provide 180° phase invertion [21]. This active stage can also compensate the gain variation of the passive phase shifter at different phase states.

Active phase shifters are based on I/Q interpolation. By summing differently weighted cosine and sine signals, a new phase can be obtained. The relative phase shift is determined



(a) variable T-line lengths (b) Synthesized T-line with variable LC

Figure 4.21: Through type phase shifters

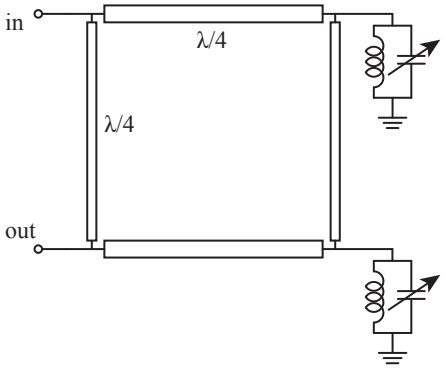


Figure 4.22: Reflection type phase shifter

by the relative weighting assigned to the cos and sin signals,

$$\sqrt{A_I^2 + A_Q^2}\cos(\omega t + \phi) = A_I \cos(\omega t) + A_Q \sin(\omega t) \tag{4.66}$$

$$\phi = -\arctan(\frac{A_Q}{A_I}) \tag{4.67}$$

The programmable weightings are usually implemented by two Variable Gain Amplifiers (VGAs). Fig. 4.23 shows several different implementations. In the first implementation (Fig. 4.23a), the variable gain function is achieved by programming the current source value, and the quadrant selection is performed by directing current to one of the two Gilbert differential pairs. The advantage of this implementation is high current efficiency since all the current is utilized. The downside is the varying output common level and thus varying output voltage swing headroom. In the second implementation (Fig. 4.23b), the variable gain function is achieved by unequally distribute a total constant current to two Gilbert differential pairs to achieve a variable differential output current. Compared to the first implementation, this architecture has constant output common mode but less current efficiency since the differential output current is only a fraction of the total bias current. In both implementations, the original signal (cosine signal or sine signal) has to drive four transistors that form the Gilbert quad to enable quadrant swapping. In contrast, the third and fourth implementation places the quadrant selection at the input of the amplifier using

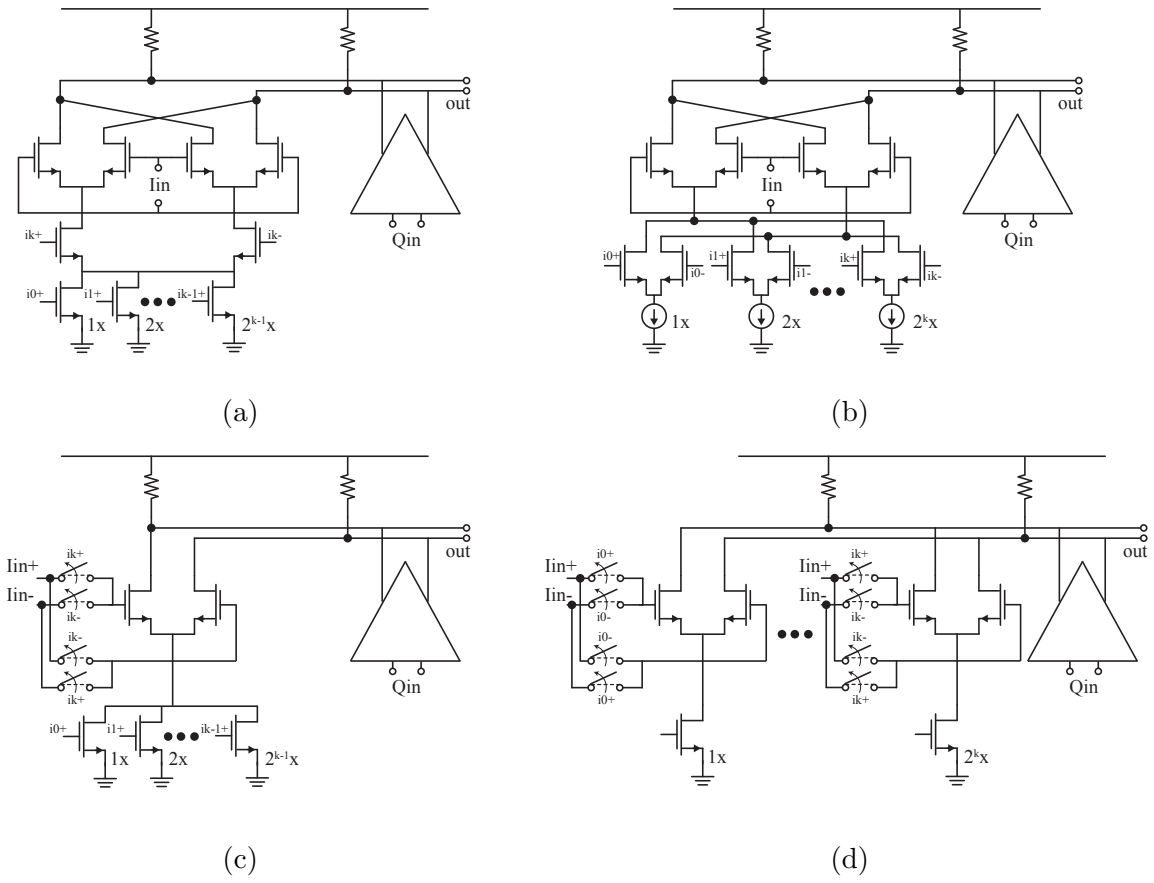


Figure 4.23: I/Q interpolating phase shifters

pass gates (Fig. 4.23c and Fig. 4.23d). As a result, the loading to the signal path can be reduced by half. The disadvantage is the added loss of the pass gate switches and additional noise.

4.6 Low Power 4-Element Phased Array

While the previous sections describe the fundamentals of phased arrays, this section focuses on the design of a mm-wave phased array. Specifically, the phased array operates at 60GHz and the goal is to achieve 10Gb/s data rate over 2 meters Line-of-Sight (LOS) distance while maintaining low power consumption. The number of phased array elements is determined based on the tradeoff of power, efficiency and area. For the same EIRP, less PA power is needed with more transmitter elements. However, this doesn't mean the transmitter power can be reduced indefinitely. In fact, overhead power from mixer and LO distribution increases with the number of elements, as a result, there exists an optimal number of elements for minimal power consumption. This optimal number depends on a

No. of Antennas	4	units
Antenna Gain	-3	dBi
Array Gain	12	dB
PA Power	0	dBm
EIRP	9	dBm
Distance	2	meters
Path Loss	74	dB
RX Input Level	-65	dBm
RX Noise/Hz	-174	dBm/Hz
RX Bandwidth	5	GHz
Noise Level	-69	dBm
RX Noise Figure	8	dB
SNR (one element)	4	dB
Array SNR Gain	6	dB
RX Array Output SNR	10	dB

Table 4.1: Link budget analysis for a 10Gb/s 60GHz QPSK link over 2 meters

number of factors such as circuit architecture, supply voltage and passive Q-factor. In this prototype, four elements are chosen [29]. Table. 4.1 shows the link budget analysis for the desired data rate and distance. A receiver output SNR of 10dB is targeted for obtaining a BER of 10^{-3} with QPSK modulation.

Based on the preceeding comparison of four different phased array architectures, the analog baseband phase shifting architecture is chosen for better phase shifter performance and lower power consumption due to small array size and limited mixer gain ($G_{mix} < 2$). The block diagram of the entire transceiver is shown in Fig. 4.24. The transceiver consists of four TX, four RX, an integrated integer-N frequency synthesizer and a LO distribution network.

Each transmitter element consists of an analog baseband phase shifter, a double-balanced quadrature Gilbert mixer and a ZVS power amplifier (Fig. 4.25). In a relatively low output power transmitter, the overall efficiency is no longer determined by the PA since other blocks in the transmitter chain also contribute a significant portion of the total power consumption. Therefore, efficiency optimization is a key design component for every block in the transmitter chain and this is achieved by various techniques addressed in the following subsections.

4.6.1 Phase Rotating Quadrature Mixer

Maximizing the current efficiency of the mixer is critical not only in reducing the mixer power, but also in maximizing the impedance at the LO ports which reduces the required

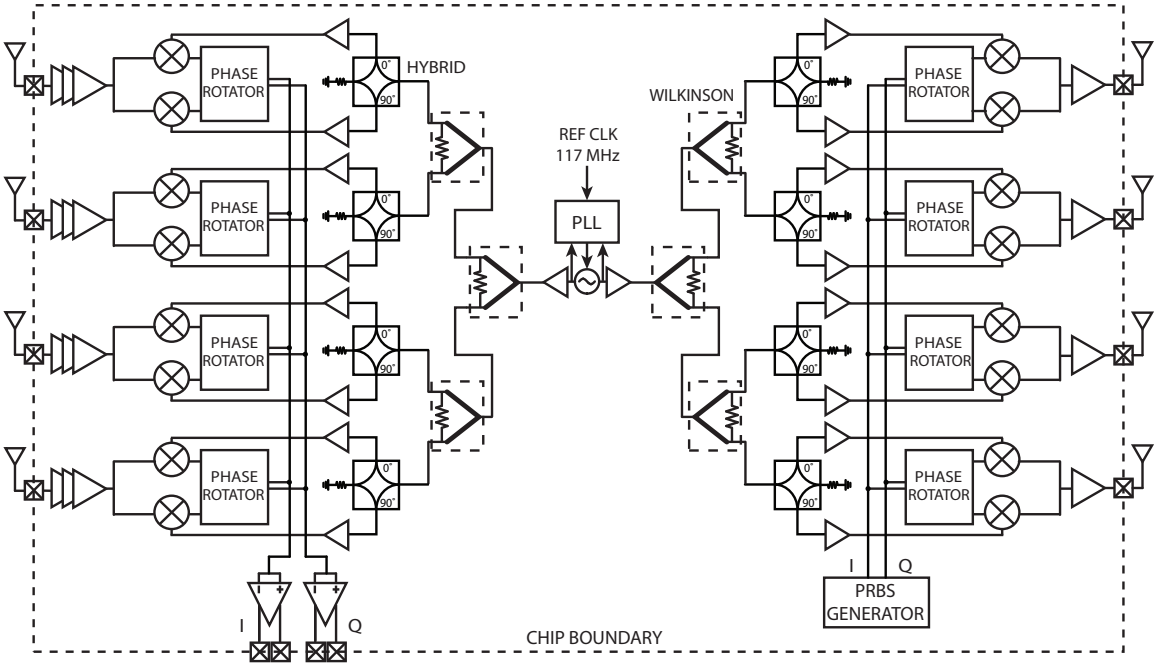


Figure 4.24: Block diagram of the 60GHz four-element phased array transceiver

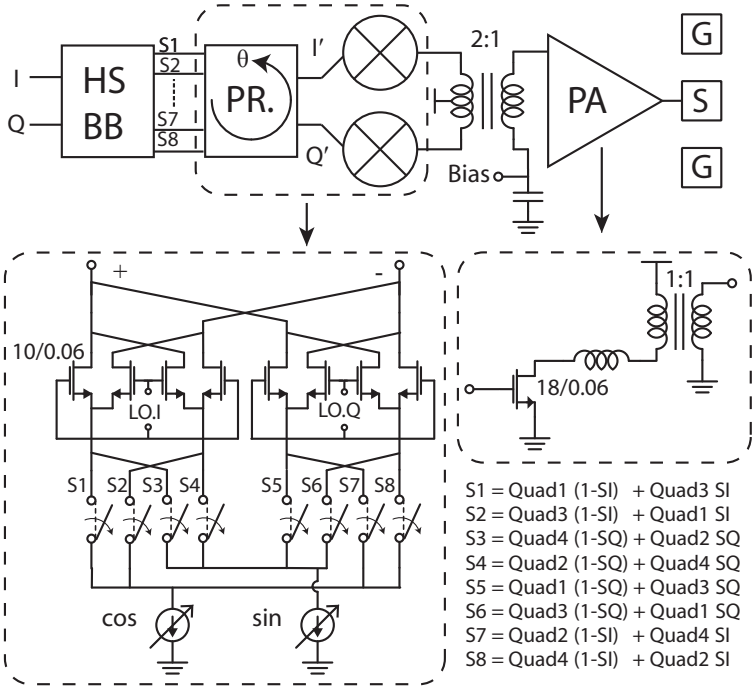


Figure 4.25: Schematic of the transmitter element

LO buffer current swing. Combining the baseband phase shifter and the quadrature mixer into a single structure is attractive since the bias current can be reused. In such a combined structure, the conventional means to achieve phase shift functionality is to current sum the weighted original I and Q signals, as illustrated in Fig. 4.26a. However, this architecture suffers from relatively low efficiency. To gain insight into the root cause of this poor efficiency, consider the current efficiency of the phase rotator, which is defined as the ratio of the effective differential output current magnitude to the total DC current,

$$\eta_{traditional} = \frac{\sqrt{I'^2 + Q'^2}}{2(\cos\theta + \sin\theta)} = \frac{\sqrt{I^2 + Q^2}}{2(\cos\theta + \sin\theta)} \tag{4.68}$$

$$= \frac{1}{\sqrt{2}(\cos\theta + \sin\theta)} \tag{4.69}$$

where I' and Q' are the phase shifted baseband signals and the final equality is based on the fact that the original baseband I/Q values are either 1 or -1 for QPSK modulation. Note that the efficiency falls between 50% to 71% depending on phase shift angle θ. The low efficiency is further explained by an example shown in Fig. 4.26b. At 45° phase shift angle with both initial I/Q inputs of 1, one pair of the cos and sin current appears at the output as common-mode signals to achieve an effective I magnitude of zero. In other words, half of the total current is wasted due to the current-mode subtraction inherent in this structure. To improve the current efficiency, an improved baseband phase shifter architecture is proposed in Fig. 4.27. In the proposed architecture, only two current sources are used and instead of carrying $\cos\theta$ and $\sin\theta$ weightings, they carry $\cos(\theta + 45°)$ and $\sin(\theta + 45°)$ weightings, representing the magnitude of the phase-rotated I/Q signals. Depending on in which quadrant the final phase-shifted data lands, the $\cos(\theta + 45°)$ weighting should represent the I' signal or the Q' signal. Since the final data quadrant depends on both the initial data quadrant and the phase shift angle, eight current distribution switches are used to dynamically distribute the two current sources to the appropriate I/Q output ports. Since the DC current is reduced, the current efficiency can be improved. The new current efficiency is,

$$\eta_{proposed} = \frac{\sqrt{I'^2 + Q'^2}}{2(\cos(\theta + 45) + \sin(\theta + 45))} \tag{4.70}$$

$$= \frac{\sqrt{\cos^2(\theta + 45) + \sin^2(\theta + 45)}}{2(\cos(\theta + 45) + \sin(\theta + 45))} \tag{4.71}$$

$$= \frac{1}{(\cos(\theta + 45) + \sin(\theta + 45))} \tag{4.72}$$

Note that the new efficiency falls between 71% and 100%, a significant improvement compared to the conventional approach. Fig. 4.28 plots the efficiency comparison for different phase shift angles and the new architecture is superior to the conventional one at any angle with an average improvement of around 40%. The current distribution switches S1-S8 are controlled by the baseband input quadrant information (Quad1 to Quad4), which are computed from the original I/Q signals, as well as the sign bits of the 7-bit phase shift control, as illustrated in Fig. 4.25. The truth table of the quadrant signals and the correspondence

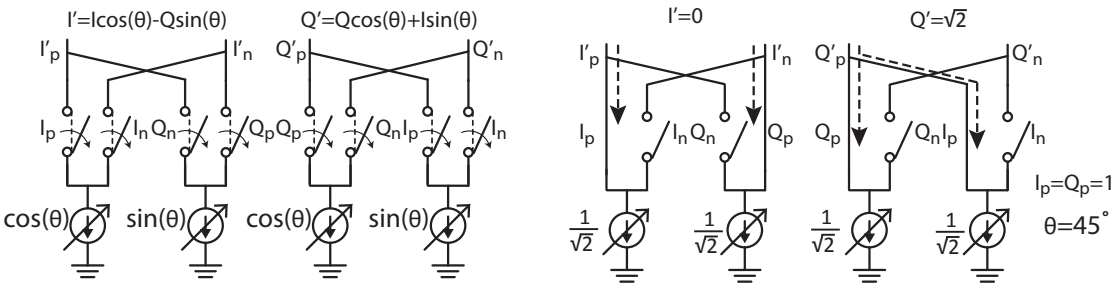


Figure 4.26: Conventional BB phase shifter architecture

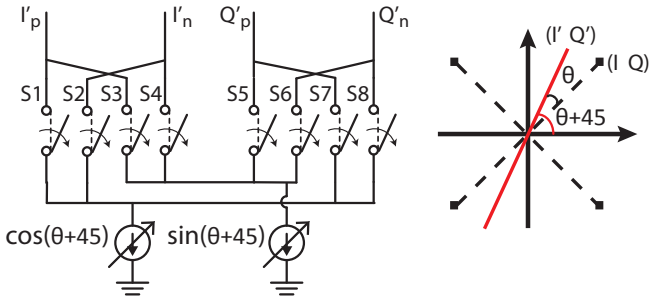


Figure 4.27: Proposed BB phase shifter architecture

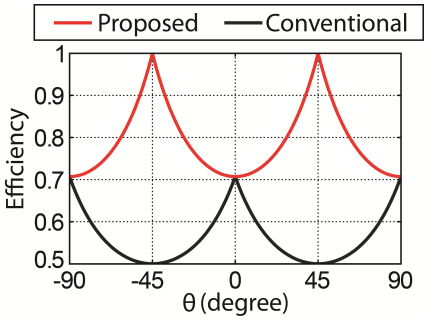


Figure 4.28: Efficiency comparison of the conventional and proposed phase shifters

of the two sign bits and the phase shift region are shown in Table. 4.2 and Table. 4.3. At each clock sampling point, only one of the four switches (S1 to S4) connecting the I mixer is turned on while others remain off. The same principle applies to the four switches (S5 to S8) connecting the Q mixer. The current sources are implemented by 5-bit DACs with upper 4-bits thermometer encoded. The phase shifted baseband signals are fed into two double-balanced mixers for quadrature modulation. Since the sizes of the Gilbert quad switches set the driving requirement for the LO buffers, smaller size transistors are preferred to increase the input impedance and reduce the buffer power. Due to reduced phase shifter current, $10\mu m$ transistors are used, which present a differential impedance of $1.2k\Omega$. With a differen-

Input (I,Q)	Quad1	Quad2	Quad3	Quad4
(1,1)	1	0	0	0
(-1,1)	0	1	0	0
(-1,-1)	0	0	1	0
(1,-1)	0	0	0	1

Table 4.2: Truth table of the quadrant signals

SI,SQ (Sign bits of the phase shift setting)	Phase shift region in radians
(0,0)	0 to $\pi/2$
(1,0)	$\pi/2$ to π
(1,1)	π to $3\pi/2$
(0,1)	$3\pi/2$ to 2π

Table 4.3: Phase shift region and its corresponding control sign bits

tial input swing of 600mV, the required LO input power is kept below -8dBm. The outputs of the two double-balanced mixers are current combined before feeding to the transformer used for coupling the PA and the mixer. The power consumption of the phase rotating mixer varies between 6-8.75mA depending on the phase shift setting.

4.6.2 ZVS PA

Since QPSK modulation has relatively small PAPR ratio before filtering, this enables the use of a zero-voltage-switching (ZVS) amplifier for improved efficiency[2]. Recall in Chapter. 2, it's shown in Fig. 2.15c that traditional Class-E amplifiers have non-optimal PAE at mm-wave frequency. With harmonic impedance tuning, the extended Class-E/F$_X$ amplifiers can potentially improve the PAE by more than 20%. On the other hand, adding higher order harmonics beyond the second has diminishing returns in efficiency improvement but results in more complexity (and hence additional passive loss) in implementing the on-chip tuning network, and therefore this design limits the harmonic tuning to the second order. In fact, it will be shown that involving second harmonic tuning is already challenging at this frequency range.

Similar to Fig. 2.15c, the drain efficiency, power gain, and PAE of the core PA can be evaluated as a function of the normalized second harmonic reactance (X_2) presented to the transistor (Fig. 4.29). Unlike Fig. 2.15c which is based a generic 65nm process, Fig. 4.29 is based on ST 65nm GP process with transistor parameters extracted from layout. In addition, a Q-factor of 10 is assumed for an inductor at the fundamental frequency while a Q-factor of

[2]In cases where the signal is filtered, a linear PA is required to preserve the output spectrum shape. In other words, filtered QPSK signals have non-negligible PAPR ratio.

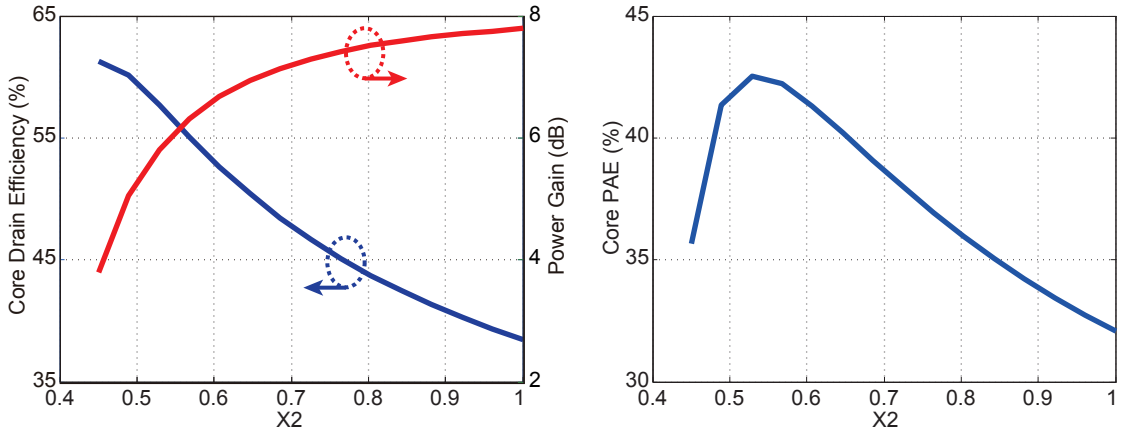


Figure 4.29: Predicted PA drain efficiency, power gain and PAE

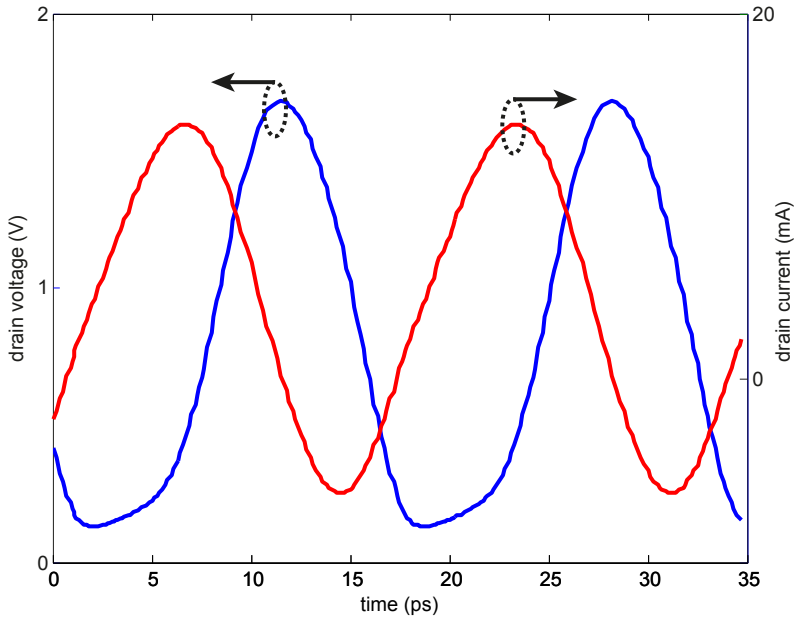


Figure 4.30: Schematic of the transmitter element

4 is assumed for the second harmonic impedance. Due to decreased F_C, decreasing second harmonic reactance improves the capacitance tolerance of the ZVS topology and allows a larger switch to be used. As a result, the drain efficiency can be significantly improved. Meanwhile, larger switches lead to larger input power and therefore lower power gain. Finally, the optimal X_2 for PAE is around 0.52 at 60GHz, quite close to what's predicted based on a generic 65nm process.

The tuning network for this design is implemented by a 1-to-1 wide trace broadside

coupled transformer with a diameter of 42 μm and a series inductor of 100 pH. This passive network up-converts 50Ω to 150Ω at the fundamental frequency for optimal power delivery while providing roughly 40 pH of inductance at the second harmonic. The reduced second harmonic inductance is mainly achieved by operating the transformer beyond SRF. This means the transformer acts as a capacitor at the second harmonic and therefore reduces the effective series inductance. The output transformer has a trace width of $12\mu m$ and provides inherent electrostatic discharge (ESD) protection. The ZVS amplifier operates from a 0.8 V (4.4mA) supply for reliability and has a simulated power gain of 5dB. The simulated core drain efficiency is 54% while the overall drain efficiency including the passive loss and the pad loss is 34%.

Although this tuning network is able to provide second harmonic inductance, the actual realized Q-factor is very low at the second harmonic. This is because the transformer is loaded with 50Ω on the secondary side, and therefore the effective capacitance that the transformer presents has fairly low Q. As a result, the second harmonic content is the current waveform is not rich, evident in the V-I waveforms (Fig. 4.30).

As the target output power is 0dBm for each element, a single-ended PA is used instead of a differential one since the differential PA requires an additional 4X impedance upconversion and would hence suffer from excessive loss. A 2-to-1 transformer is used to achieve the differential to single-ended conversion as well as to perform the required impedance matching between the mixer and the PA. To improve the common-mode rejection of the transformer balun, primary and secondary windings are implemented in two different top metal layers such that the capacitive coupling can be minimized. In addition, a 600fF de-coupling capacitor is added to the center-tap of the primary winding in order to resonate with the common-mode winding inductance. In order to capture the lead inductance of the current summing structure at the output of the quadrature mixer, an 8-port input structure is attached with the transformer primary in the EM simulation. The combined structure has an insertion loss of 3dB.

4.6.3 Experimental Results

The detailed description of the rest of the transceiver blocks can be found in [30]. The transceiver was realized in a 65nm bulk CMOS process without any special RF options. Fig. 4.31 shows the die micrograph. The chip measures 2.5mm by 3.5mm. The measurement was performed by direct on wafer probing of mm-wave signals in a chip-on-board configuration which allowed DC and IF signals to be wire bonded to the PCB. Fig. 4.32 shows the measured output power of each transmitter element. The measured peak output power and PA drain efficiency is -1.5dBm and 20% respectively with 3-dB bandwidth of more than 8GHz. The output power difference among four elements is less than 0.5dB. The measured power is about 1.5dB lower than the simulation results. The main suspect for this difference is the modeling of the 2-to-1 transformer Balun connecting the mixer and the PA. In simulation, the transformer operates very close to SRF, and in addition the insertion loss depends on the common-mode termination, both lead to increased sensitivity.

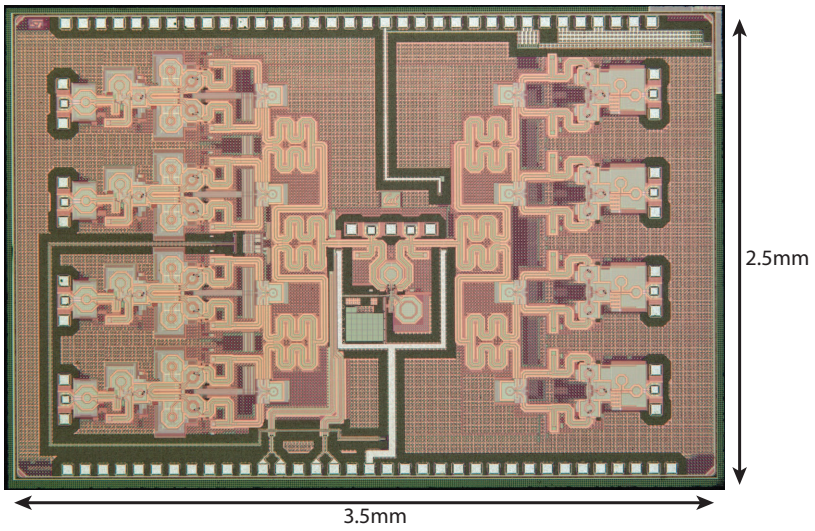


Figure 4.31: Chip micrograph of the 60GHz 4-element phased array transceiver.

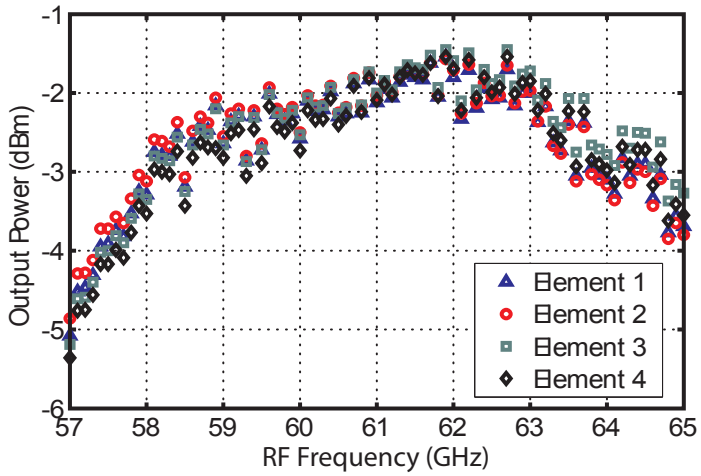


Figure 4.32: Measured transmitter output power.

Each transmitter element was tested using QPSK modulated signals mapped from PRBS sequences generated on-chip. The 60 GHz output was down-converted using an external harmonic mixer. Fig. 4.33 shows the measured eye-diagram when transmitting 5Gb/s data on the I-channel. The transmitter consumes a total worst-case (45° phase shift) power of 27mW/element including LO distribution.

The measured TX phase constellations of the four elements are shown in Fig. 4.34. In the TX, the VNA is used to measure the relative phase shift. An external LO signal is fed from Port 1 of the VNA to the distribution tree while taking the output from the PA to Port 2. The relative phase shift is observed. The TX achieves 360° of phase shifting range with a worst-case phase steps of 5°. The gain variation across all phase settings and all elements is less than ±0.5dB.

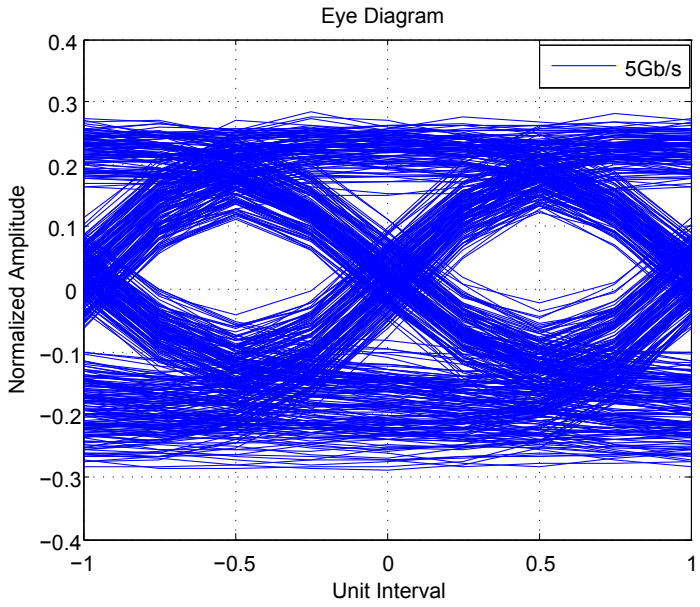


Figure 4.33: Measured eye-diagram of the I-channel while transmitting 5Gb/s QPSK data.

The two-element phased array normalized gain is shown in Fig. 4.34a as a function of relative phase shift angle. This two-element measurement was performed by using dual GSG probes and off-chip power combiners which allowed simultaneous probing of both elements. The phase setting of one element was held constant while the phase shift on the other element was swept over its entire range. Due to the high phase resolution and low gain mismatch, the measured peak-to-null ratio on the TX is 40dB. This also confirms that the on-chip isolation between elements is sufficient to obtain a good peak-to-null ratio.

Although 4-element beamforming pattern is unavailable due to lack of antenna array integration, a synthesized pattern was constructed using the measured 4 TX element power and phase characteristics (Fig. 4.34b). The pattern is based on a $\lambda/2$ uniformly spaced array. In such an array pattern, the peak gain is not as sensitive as the nulls to the mismatch in gain and phase between elements. Thus, the close matching between ideal and synthesized array patterns shown in Fig. 4.34b further confirms the high resolution and accuracy of the proposed transceiver design.

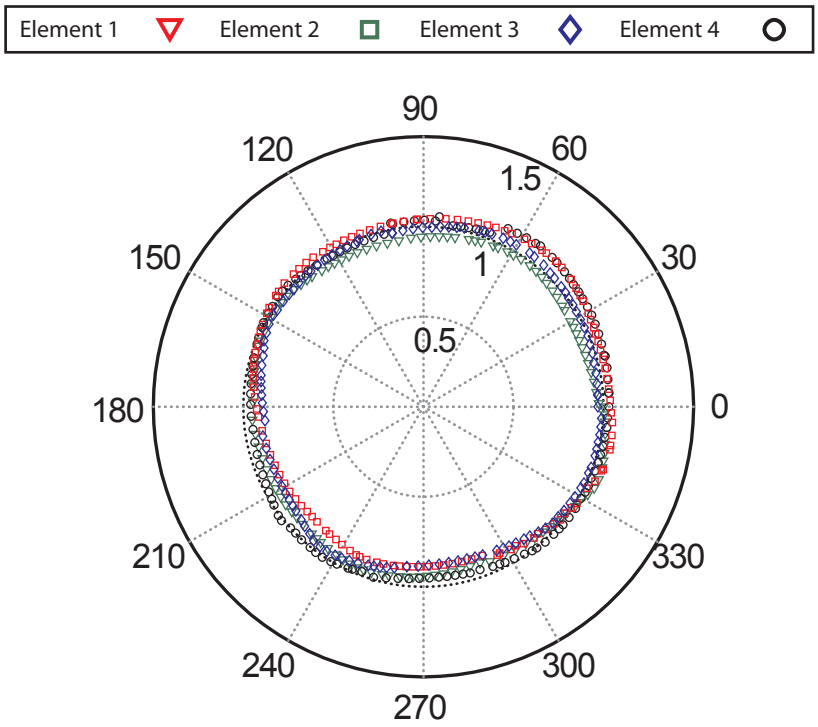
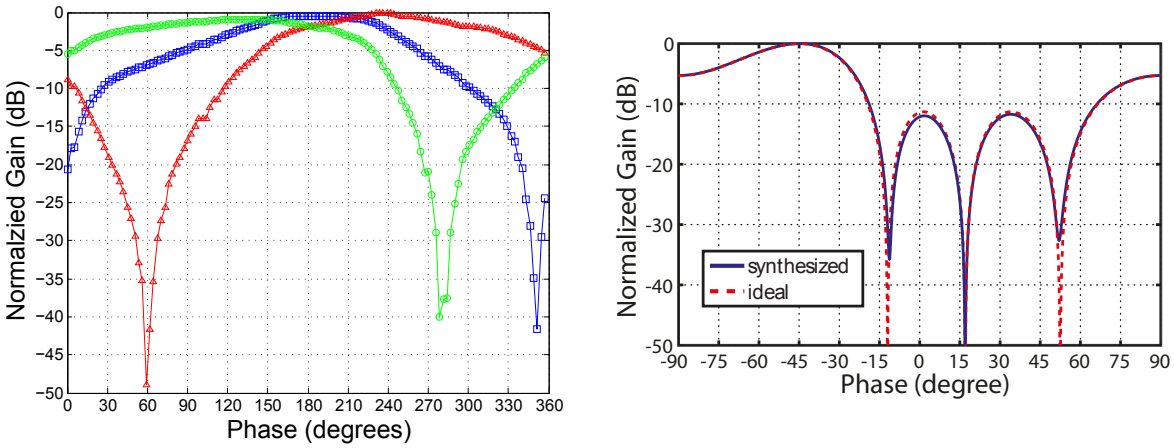


Figure 4.34: Measured phase constellations for four TX elements.



(a) Two element measured

(b) Four element synthesized

Figure 4.35: Transmitter beamforming pattern.

Chapter 5

Direct Digital-to-RF Transmitter

The previous chapters present the architectures and circuit techniques mainly for improving the mm-wave transmitter peak output power/EIRP. However, mm-wave transmitters suffer from not only low power handling capability, but also low efficiency, as evident in Fig. 2.16. This is mainly caused by reduced available power gain and increased device parasitic loss. Unlike lower frequency counterparts, mm-wave PAs typically include multiple driver stages to obtain the desired power gain, which leads to a larger difference between the 1dB gain compressed output power (P_{1dB}) and the saturated output power (P_{sat}) due to simultaneous gain compression from multiple stages. Such discrepancy can be as large as 3-4dB [31, 32]. Since the extra power beyond P_{1dB} cannot be utilized in a linear transmitter due to amplitude distortion, the actual peak efficiency is already significantly reduced[1]. To make matters worse, most digital modulation schemes widely used today have a fairly high PAPR, which demands the transmitter to back-off from its peak linear output power level P_{1dB}. Since the efficiency of Class-A PAs drops linearly with output power, a 6dB power back-off reduces the PA efficiency by 75%. As a result, the average efficiency during data transmission is only a small fraction of the peak efficiency. Various techniques have been introduced to enhance the average efficiency, however, challenges remain to adopt them for mm-wave transmitters. This chapter presents a direct digital-to-RF conversion architecture that enables dynamic DC power scaling at very high data rates and thus significantly improves the mm-wave transmitter average efficiency. By utilizing a concept called Quadrature Spatial Combining, both I/Q load-pull and insertion loss present in a traditional Cartesian transmitter can be minimized simultaneously. The rest of this chapter is organized as follows. It first reviews traditional efficiency enhancing architectures that are widely used at lower frequencies and discusses challenges for high frequency adoption. Section 2 introduces the direct digital-to-RF architecture and discusses the practical implementation at mm-wave frequency domain. Section 3 presents the Quadrature Spatial Combining concept and its implications on the PA performance and beamforming. The circuit implementation of an eight-element

[1]The efficiency numbers in Fig. 2.16 are mostly those at P_{sat} for mm-wave PAs. Efficiencies at P_{1dB} are usually half.

(4I+4Q) 60GHz digital-to-RF conversion beamforming transmitter is presented in section 4, followed by experimental results in section 5. Finally, digital calibration and pre-distortion techniques are discussed in section 6.

5.1 Traditional Efficiency Enhancing Architectures

The main reason behind the low average efficiency of Class-A PAs is that the DC power consumption remains constant and independent of the instantaneous signal amplitude. In other words, the larger the PAPR is, the lower the average efficiency becomes. A number of transmitter architectures have been proposed to improve the average efficiency by enabling dynamic DC power scaling through various ways. Three most popular architectures are envelope tracking, envelope elimination and restoration and outphasing[2].

In an envelope tracking transmitter (Fig. 5.1), dynamic DC power scaling is implemented by modulating the supply voltage according to the signal envelope through an amplitude amplifier. The envelope signal can be generated either directly from the baseband DSP through a DAC, or by passing the analog signal through an envelope detector. Since the supply voltage scales with the signal envelope while the DC current remains constant to the first order, the efficiency backs off linearly with the output amplitude, resembling a Class-B amplifier behavior. In deep sub-micron process nodes with small channel resistance r_o, the DC current of a transistor also drops with reduced drain voltage. As a result, the actual efficiency back-off is slightly better than Class-B. The downside of this architecture is that a linear PA is still needed to amplify the non-constant envelope input signal, as shown in Fig. 5.1. Therefore, the peak efficiency of this system is still similar to a traditional transmitter.

To further enhance the efficiency, a concept called Envelope Elimination and Restoration (EER) was invented. The most practical EER implementation is the Polar architecture, as shown in Fig. 5.2. In such a transmitter, the baseband signal is first converted to amplitude (AM) and phase (PM) signals. Then a highly efficient but nonlinear PA can be used to amplify the constant envelope PM signal while the amplitude information is restored at the PA output through a supply modulator similar to that used in an envelope tracker. The benefits of this architecture are two folds. First, much higher peak efficiency can be obtained with non-linear switching PAs. Second, since the efficiency of switching PAs is independent of the supply voltage to the first order[3], the peak efficiency can be maintained at different back-off power levels. Although EER has higher average efficiency than envelope tracking, it requires much more stringent alignment between AM and PM signals [33]. Typically a feedback loop is used to dynamically correct the timing error for optimal EVM.

Both envelope tracking and EER rely on a supply modulator, which usually has limited bandwidth. Increasing the bandwidth of a supply modulator leads to significant overhead

[2]Sometimes referred to as LInear amplification using Non-linear Components (LINC)

[3]With strong short channel effect, the efficiency does degrade with decreasing supply voltage

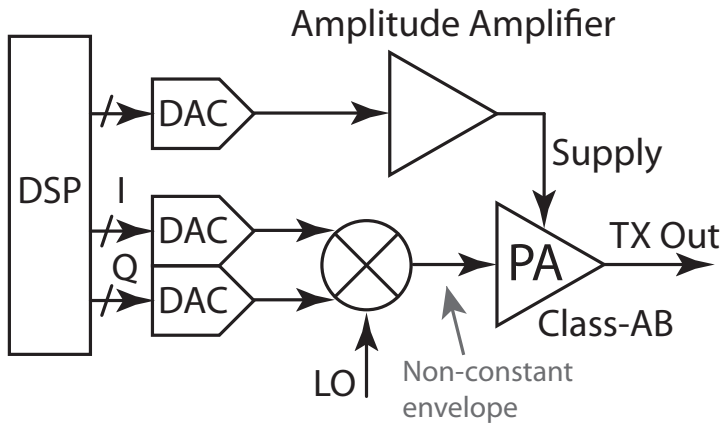


Figure 5.1: Envelope tracking through supply path.

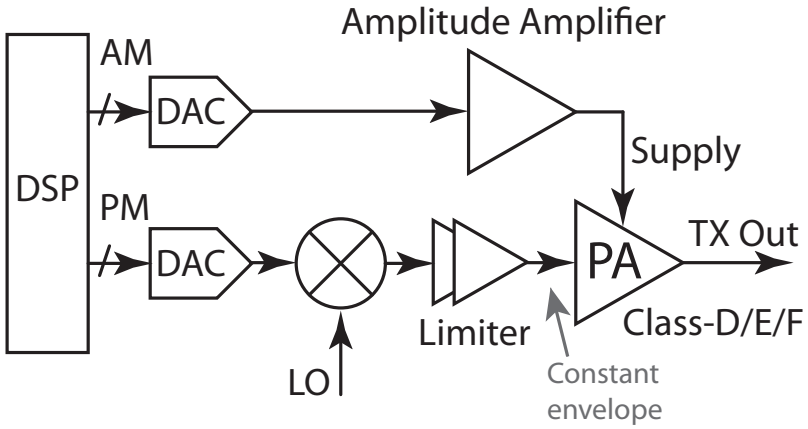


Figure 5.2: Envelope elimination and resotoration (Polar)

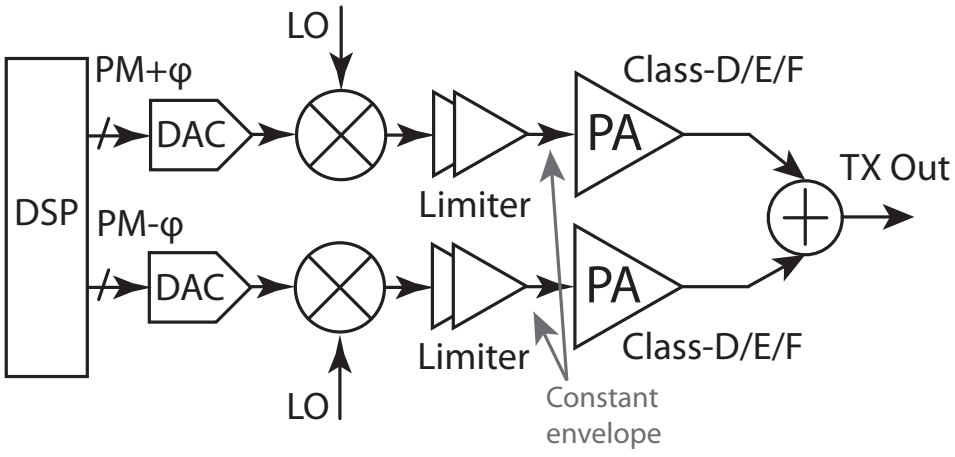


Figure 5.3: Outphasing

power. Therefore these two architectures can hardly satisfy the data rate requirement for mm-wave transmitters.

Another architecture using non-linear amplifiers is the outphasing architecture[4] [34]. In such a transmitter (Fig. 5.3), the amplitude information is created by summing signals from two non-linear PAs with the same amplitude but different phases. The phase difference between the two signals is called the outphasing angle. The summed signal amplitude depends on the outphasing angle: the larger the outphasing angle is, the smaller the summed amplitude becomes. Since each PA is transmitting a constant envelope signal, non-linear switching amplifier can be used to maximize the peak efficiency. In addition, since no supply modulator is needed in the outphasing architecture, high data rates can be achieved. The back-off characteristic of an outphasing amplifier depends on the type of signal combiner used. If isolating combiners such as the Wilkinson combiner are used, the transmitter exhibits a Class-A type of back-off since the DC power consumption is fixed and independent of the outphasing angle. On the other hand, if non-isolating combiners are used, DC power consumption may decrease with increasing outhphasing angles due to increased load impedance seen by the amplifier as a result of a load-pull. However, there are two problems with non-isolating combiners. First, the change of outphasing angle also introduces reactive impedance seen by the amplifier and hence reduces the efficiency at back-off power levels. Second, it causes amplitude distortion, especially if the amplifier cannot be modeled as a voltage source, which might be the case for most high frequency applications. The problem can be mitigated by using a Chireix combiner which compensates the change of load reactance by a fixed susceptance [35]. However, since it can only cancel the load reactance at one outphasing angle, it doesn't eliminate the distortion. A combination of dynamically variable PA size and switched tuning capacitor bank were introduced in [36] to compensate the load modulation at multiple outphasing angles and eliminate the distortion. Recently, mm-Wave outphasing transmitters have been demonstrated with both transformer combining and spatial combining [37, 38]. However due to lack of load compensation, the power stays relatively constant and the back-off curve is close to Class-A. The major efficiency enhancement comes from the utilization of the power beyond P_{1dB}.

5.2 Direct Digital-to-RF Conversion

Unlike a traditional linear transmitter with a single analog PA, the direct digital-to-RF conversion transmitter consists of an array of sub-PAs [39, 40]. Each PA amplifies a constant envelope signal while the amplitude information is created by digitally switching on and off a certain number of sub-PAs. Since each PA does not need to convey envelope information, a nonlinear amplifier class can be used to maximize the peak efficiency. On the other hand, the total DC current dynamically scales with the number of sub-PAs that are switched on, and therefore it's proportional to the output amplitude. This means the PA efficiency backs off linearly with the output amplitude, or square root of output power, similar to a Class-B

[4]Sometimes referred to as LIear amplification using Non-linear Components (LINC).

amplifier. The efficiency back-off can be described as follows,

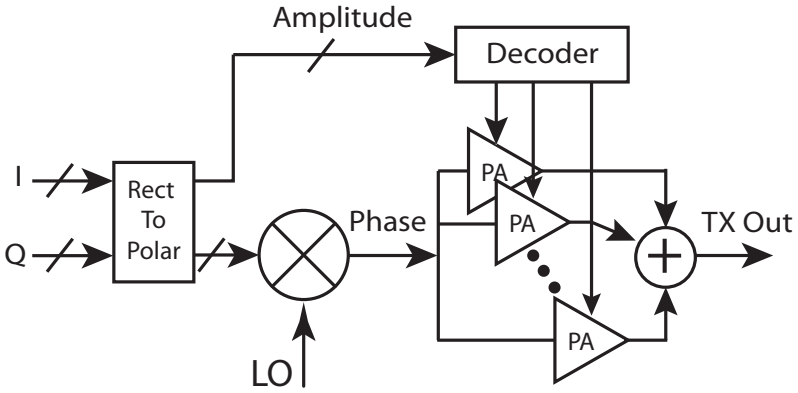
$$\eta = \frac{V_{out}}{V_{out,peak}}\eta_{peak} = \sqrt{\frac{P_{out}}{P_{out,peak}}}\eta_{peak} \tag{5.1}$$

Therefore the direct digital-to-RF conversion not only enhances the peak efficiency but also improves the back-off characteristic. There are two popular architectures for implementing the direct digital-to-RF conversion for QAM signals: the Polar architecture and the Cartesian architecture (Fig. 5.4). In the Polar architecture, the baseband digital I/Q signals are first converted to AM bits and PM bits. A phase modulator is used to upconvert the PM bits into a phase modulated carrier signal to drive the PA while the AM bits are used for sub-PA on/off switching. In the Cartesian architecture, quadrature carrier signals are used to directly drive two identical PAs which are switched on/off by the digital baseband I/Q signals respectively. Then a RF signal combiner is used to sum the I/Q PA outputs. The Polar architecture has been demonstrated at 2.4GHz for WLAN and has shown remarkable efficiency enhancement [40]. However, the Polar architecture suffers from several drawbacks. First, the bandwidth of the PM signal is significantly larger than that of the original I/Q baseband signal due to nonlinear Cartesian to Polar conversion [41]. Maintaining extra RF signal bandwidth for high data rate mm-wave systems is power costly. Second, due to different types of circuit blocks employed in the AM and PM paths, it's difficult to maintain the delay matching between the two unless a feedback loop is used. In contrast, the Cartesian architecture is free of these issues and appears to be a better candidate for mm-wave transmitters.

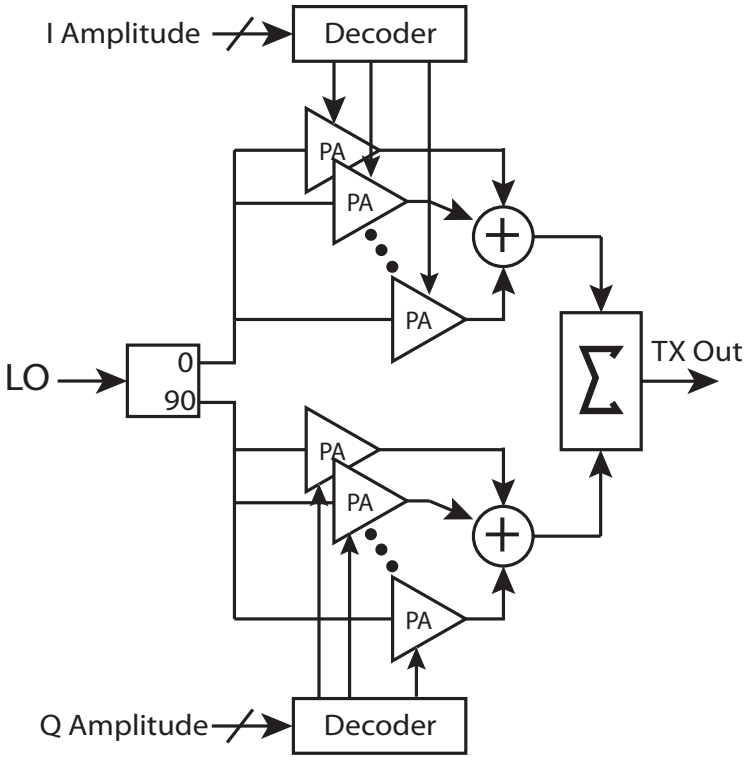
Although the Cartesian architecture has many advantages over the Polar counterpart, the biggest challenge is how to implement the final stage signal combiner. Traditionally, there are two types of combiners: non-isolating combiners and isolating combiners. Non-isolating combiners such as current summing or transformers typically have very small insertion loss, but suffer from I/Q mutual load-pull due to lack of isolation [42]. Since I/Q signals are uncorrelated, a two dimensional digital pre-distortion lookup table is required to linearize the transmitter, which adds significant complexity and overhead power. In contrast, isolating combiners such as Wilkinson combiners eliminate the I/Q mutual load-pull and preserve each PA amplitude response. However, such combiners usually have much larger insertion loss and thus degrade the PA efficiency. In other words, there's no on-chip passive combiner that can accomplish low insertion loss and high isolation simultaneously.

5.3 Quadrature Spatial Combining

To eliminate the aforementioned tradeoff between insertion loss and isolation, we propose to combine the quadrature signals in space, a concept called Quadrature Spatial Combining [43, 44] (Fig. 5.5). In a quadrature spatial combined transmitter, at least two antennas are used, one for transmitting the I path signal while the other for transmitting the Q path signal. Since the E-field vectors can be losslessly summed in the far field, no insertion loss



(a) Polar architecture



(b) Cartesian architecture

Figure 5.4: Direct digital-to-RF conversion transmitters.

is incurred. Besides, since there's no physical connection between the I/Q PAs, the mutual load-pull is eliminated. In this way, two goals can be accomplished simultaneously.

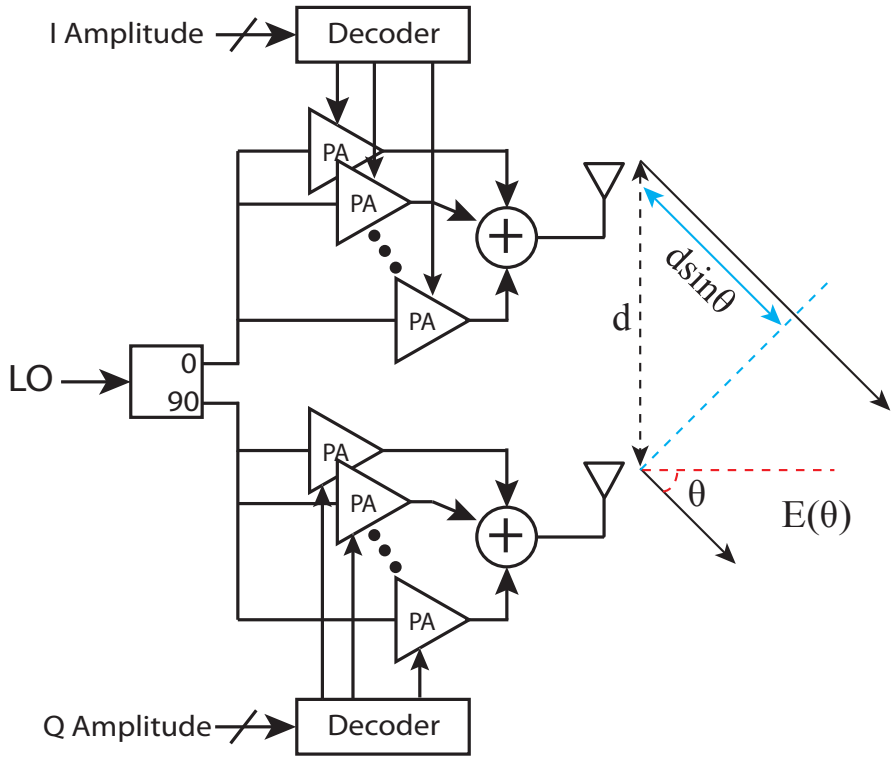


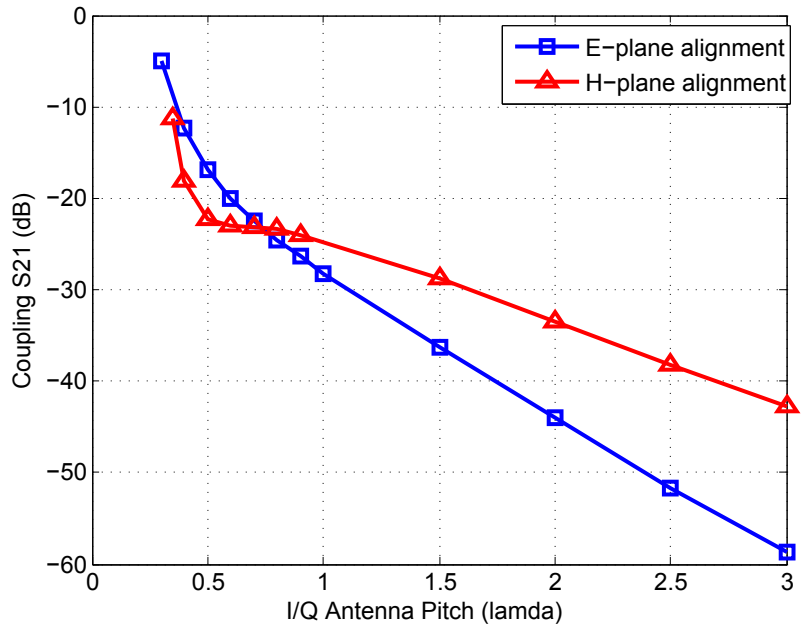
Figure 5.5: Cartesian transmitter with quadrature spatial combining.

Although the on-chip coupling between I/Q PAs is eliminated, coupling between I/Q antennas may also results in mutual load-pull and therefore should be minimized as well. Fig. 5.6a shows the simulated coupling factor S21 between two patch antennas for E-plane and H-plane alignment. The S21 drops to below -20dB once the antenna spacing is greater than half wavelength. Fig. 5.6b shows that the theoretical Error-Vector-Magnitude (EVM) of a (I=1,Q=1) symbol as a function of I/Q antenna spacing. Depending on the desired modulation scheme and actual antenna implementation, minimal spacing can be determined. For 16QAM modulation which requires roughly -19dB EVM, half wavelength spacing is sufficient.

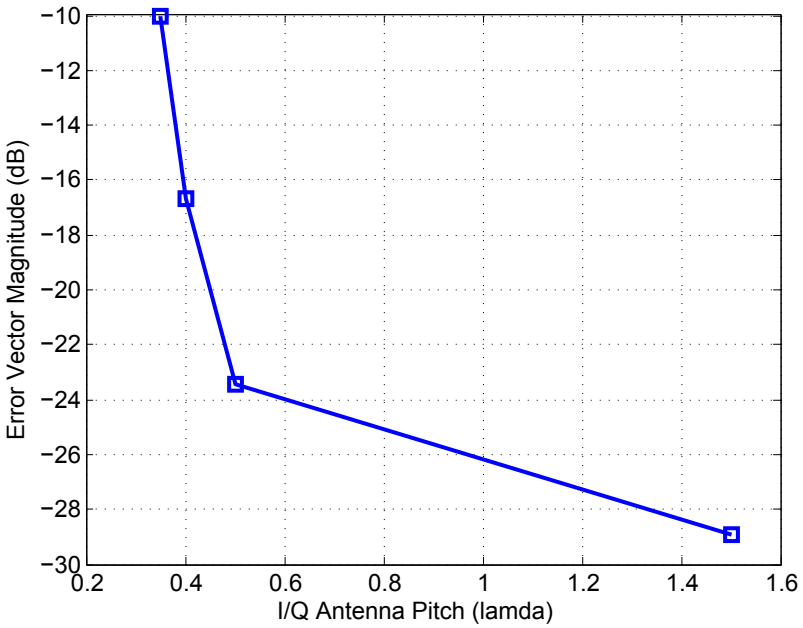
Unlike traditional beamforming, quadrature spatial combining has several unique properties, which will be explained in the following subsections.

5.3.1 Transmitter EVM

In a phased array, signals transmitted from different antenna elements experience different time delays which vary with the spatial angle θ (Fig. 5.5). As a result, quadrature phase summation can be obtained only at one particular spatial angle. Assuming the Q-PA LO is



(a)



(b)

Figure 5.6: Antenna mutual coupling (a) Coupling factor S21. (b) EVM.

phase shifted 90° relative to the I-PA LO, the E-fields radiated at θ degree angle are,

$$E_i = i(t - t_0)\cos[\omega(t - t_0)] \tag{5.2}$$

$$E_q = q(t)\cos[\omega t - \pi/2] \tag{5.3}$$

$$E = E_i + E_q = i(t - t_0)\cos(\omega t - \omega t_0) + q(t)\sin(\omega t) \tag{5.4}$$

where t_0 is the free space delay difference between the two radiated signals,

$$t_0 = \frac{d\sin\theta}{c} \tag{5.5}$$

For narrow band systems in which $T_{bb} \gg T_{rf}/2 \geq \frac{d\sin\theta}{c}$, $i(t - t_0)$ can be approximated by $i(t)$, therefore the summed E-field can be simplified to,

$$E(\theta) \approx i(t)\cos(\omega t - \frac{\omega d\sin\theta}{c}) + q(t)\sin(\omega t) \tag{5.6}$$

This means the desired modulation can be only obtained at 0° spatial angle, while signals at other transmission angles are distorted due to I/Q phase imbalance. The I/Q phase imbalance at θ angle is,

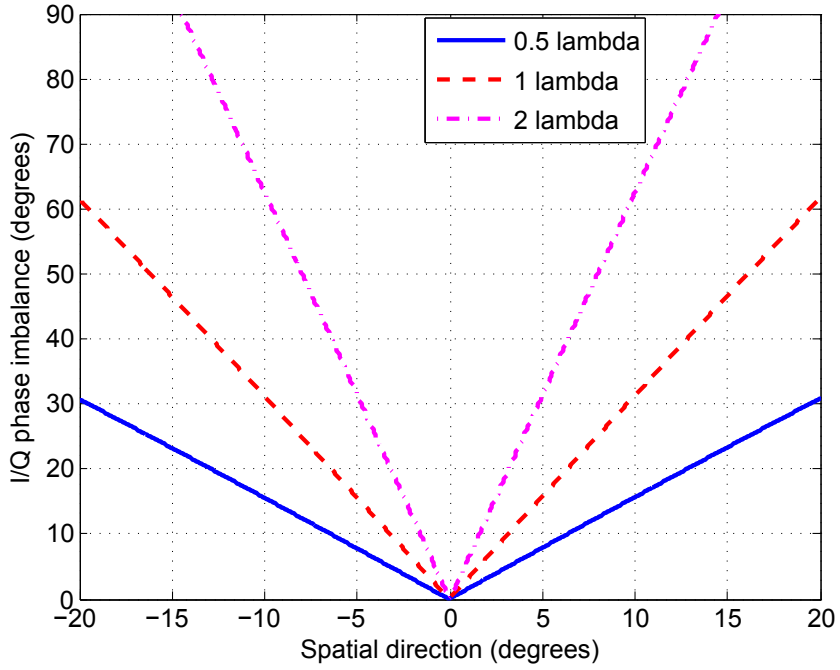
$$\Delta\phi_{IQ} = \frac{\omega d\sin\theta}{c} \tag{5.7}$$

As a result, the transmitter output EVM gradually degrades when the receiver is moved away from the center transmission direction. This provides a secure information link through a narrow spatial window and prevents eavesdropping from other directions. Fig. 5.7 plots the I/Q phase imbalance and the theoretical transmitter EVM of a QPSK signal as a function of the spatial angle. Here we define the information beamwidth as the range of spatial angles within which the transmitted data can be received correctly [45]. In other words, the transmitter EVM is below the decoding limit within the information beamwidth. Apparently, the information beamwidth depends on the modulation scheme used. Higher order modulation schemes require lower EVM for the same symbol error rate and thus result in narrower information beamwidth. Note that the patterns in Fig. 5.7 also depend on the I/Q antenna spacing. The larger the spacing, the narrower the information beamwidth becomes. Similar to a phased array, the desired information direction can be steered by setting the relative phase offset between the I/Q elements.

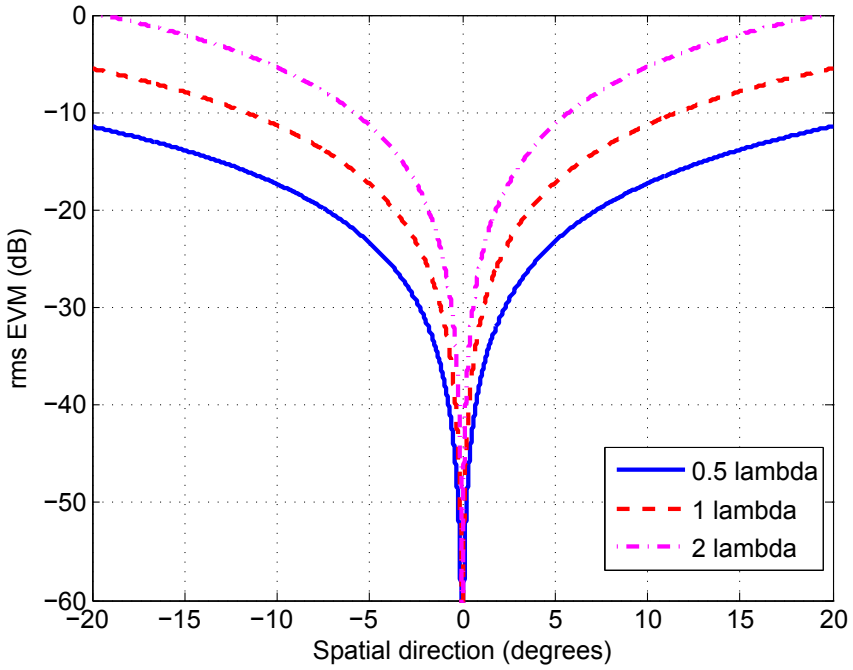
In the case of a point-to-point wireless link such as cellular backhauls, highly directional antennas are used, which means the antenna half power beamwidth is very small. Therefore quadrature spatial combing will not pose additional beam constraints as long as the information beamwidth is larger than the power beamwidth. Fig. 5.8 plots the directional gain of an antenna as a function of the half power beamwidth. Typical backhaul antennas have gain larger than 30dBi, as a result, the power beamwidth is less than 10°(±5°), whereas the information beamwidth (-19dB EVM) for 16QAM signal is around 16°.

5.3.2 Radiation Pattern

Another unique property of quadrature spatial combining is regarding the radiation beam patterns. Since the antenna array radiation pattern depends on the relative signal amplitude



(a)



(b)

Figure 5.7: Quadrature spatial combining output (a) I/Q phase imbalance. (b) EVM.

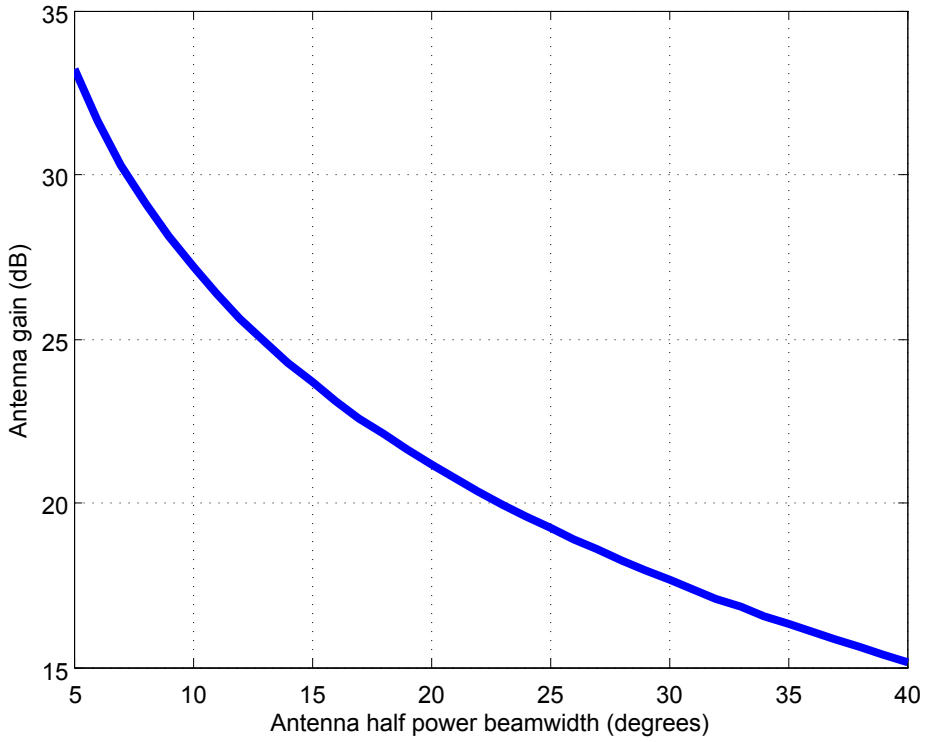


Figure 5.8: Antenna half power beamwidth as a function of antenna gain.

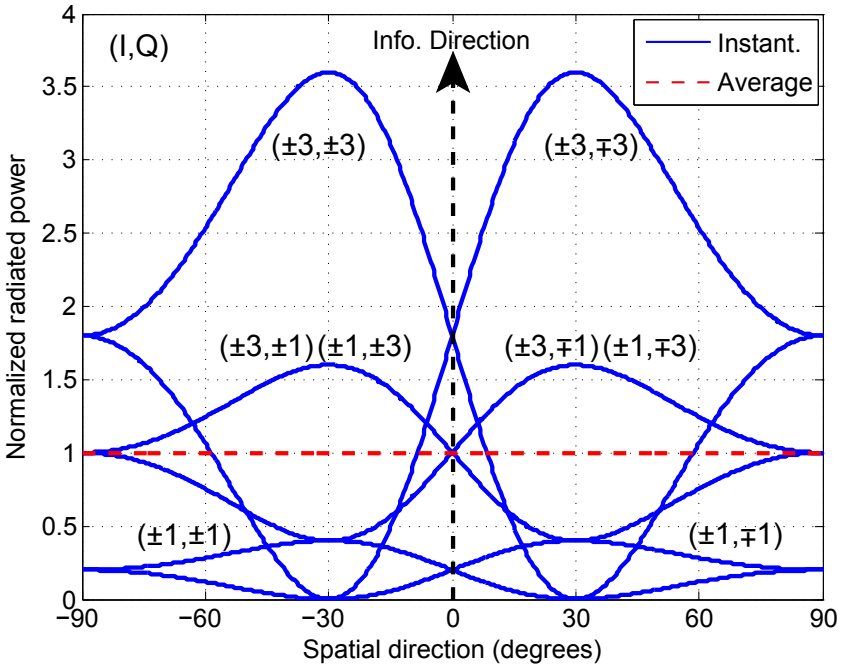
and phase of individual elements, the quadrature spatial combining pattern varies as the baseband I/Q signals change. The instantaneous power pattern is given by,

$$P(\theta) = i^2(t) + q^2(t) + 2i(t)q(t)\sin(\frac{\omega d \sin\theta}{c}) \tag{5.8}$$

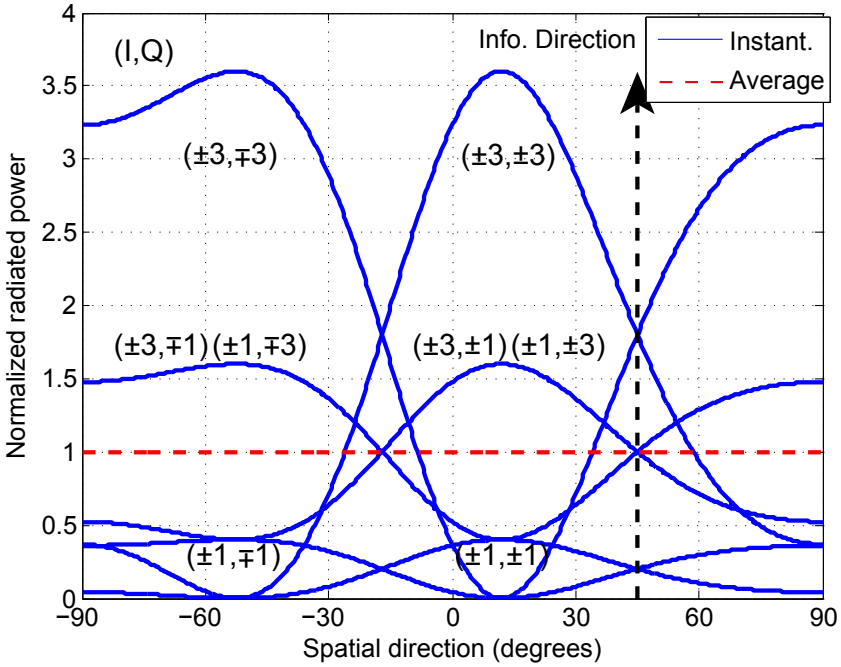
Fig. 5.9 shows the power pattern of a quadrature spatial combined transmitter with two antennas for 16QAM signals. Clearly the instantaneous power pattern depends on the transmitted symbol. Due to quadrature signal summation, the power radiated at the information direction is always lower than the peak power, and the difference lies between 0 to 3dB depending on the I/Q amplitude ratio. The peak instantaneous power happens at,

$$\theta_{pk} = \arcsin(\frac{\pi}{2}\frac{c}{\omega d}) \tag{5.9}$$

Despite the data dependent instantaneous power pattern, the time averaged power pattern is uniform, as shown in the dashed horizontal line in Fig. 5.9. This is because the pattern for symbol (1,1) is complementary to the pattern for symbol (1,-1), which means the peak of the former lands on the null of the latter, and vice versa. As a result, the average of the two patterns is flat across the entire space. This is also independent of the information direction, as shown in Fig. 5.9. This leads to another important property: although employing two antenna elements, quadrature spatial combining does not create averaged directivity for balanced I/Q modulation signals.



(a) 0° information direction



(b) 45° information direction

Figure 5.9: Instantaneous and time averaged radiation power pattern of a quadrature spatial combined transmitter with two antennas for 16QAM signal.

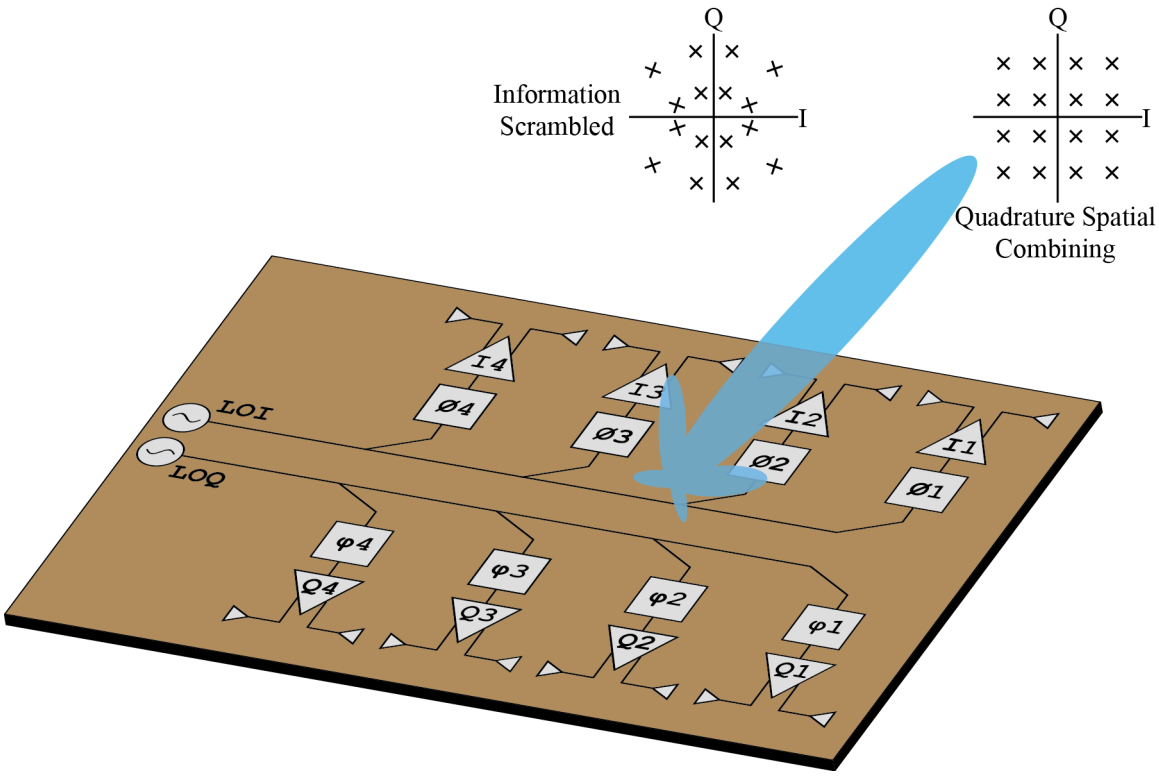


Figure 5.10: Multi-element beamforming transmitter with quadrature spatial combining.

The whole concept can be further extended to a beamforming transmitter where each I/Q transmitter consists of multiple elements, as shown in Fig. 5.10. The information beam can be steered along with the power beam of each I/Q antenna array. In a beamforming transmitter with multiple PAs, there are several ways of implementing direct digital-to-RF conversion:

i). Digital PA only. In this scheme, each PA consists of a sub-PA array driven by the baseband digital bits. The PAs in different elements are driven in the same manner by the same digital bits. Hence the resolution of each PA is equal to the resolution of the entire transmitter.

ii). Digital Antenna only. In this scheme, each PA is a 1-bit DAC, either on or off. The digital-to-RF conversion is achieved by turning on and off individual elements. Hence the number of antenna elements determines the amplitude resolution.

iii). Digital PA/Antenna. In this segmented approach, the LSB digital bits drive the first element sub-PA arrays while the MSB digital bits drive the PAs in other elements. The resolution of the entire transmitter is the sum of the PA resolution and the number of the antenna elements.

The advantage of digitally switching antenna elements is higher amplitude resolution and therefore lower output spectral noise floor.

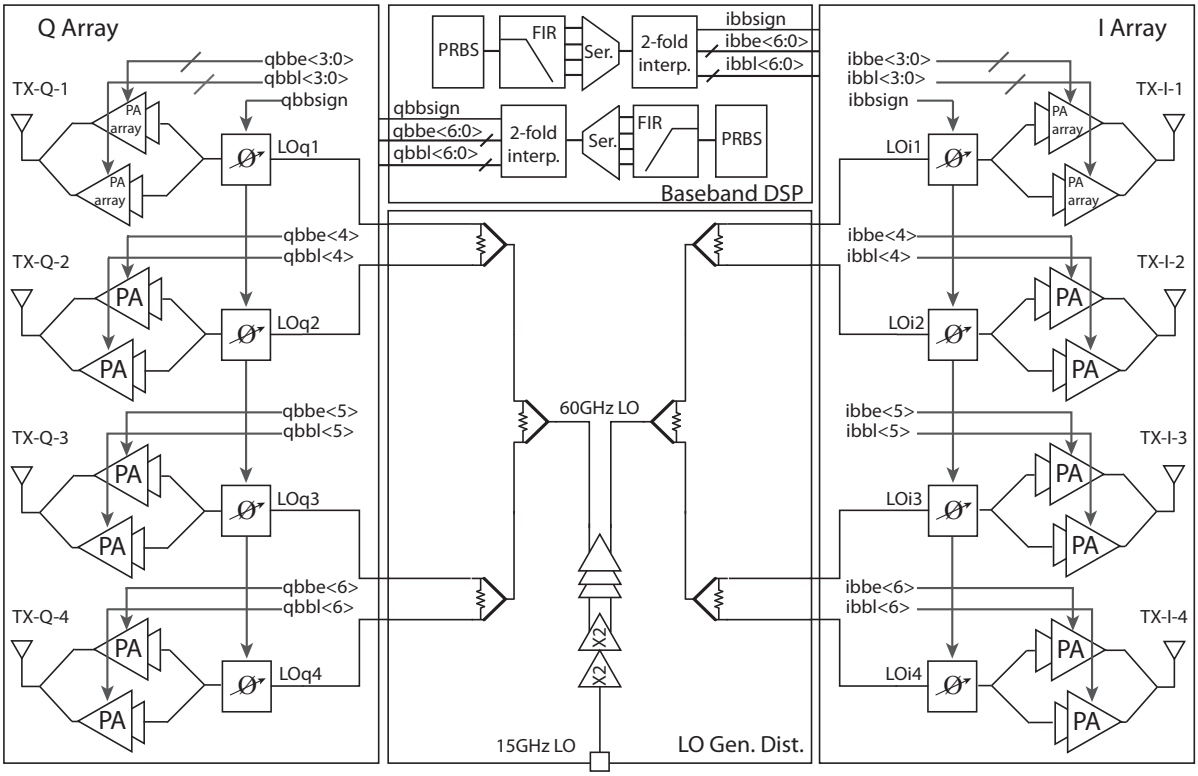


Figure 5.11: System block diagram of the 60GHz beamforming transmitter with quadrature spatial combining.

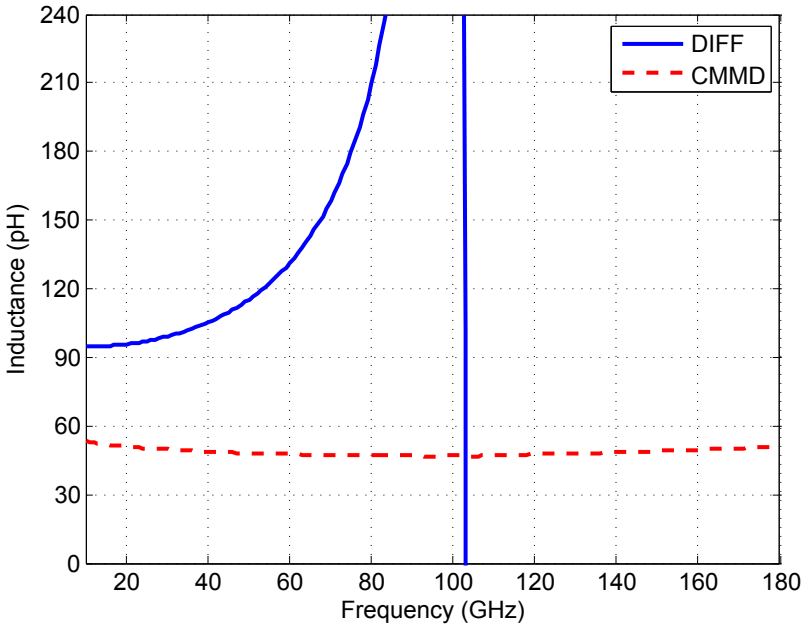
5.4 Circuit Implementation

To demonstrate the concept, a 60GHz beamforming transmitter was implemented targeting WiGig applications. A segmented digital PA/antenna approach is chosen to achieve higher amplitude resolution. In this transmitter (Fig. 5.11), each I/Q array consists of 4 elements, hence providing 2 bits of coarse amplitude resolution (bit 4-6). The PA in the first element in further segmented into 4 thermometer cells to provide 2 bits of fine amplitude resolution (bit 0-3). As a result, the total amplitude resolution is 4 bits for each I/Q array. Both coarse and fine digital bits are thermometer coded. The LO signal is distributed through a chain of lumped Wilkinson power dividers and LO phase shifters are used to locally adjust the phase for beam steering as well as quadrature signal summation. Finally a mixed-signal approach combining baseband DSP and two-fold linear interpolation is implemented for output spectral filtering. The details of each block are described in the following subsections.

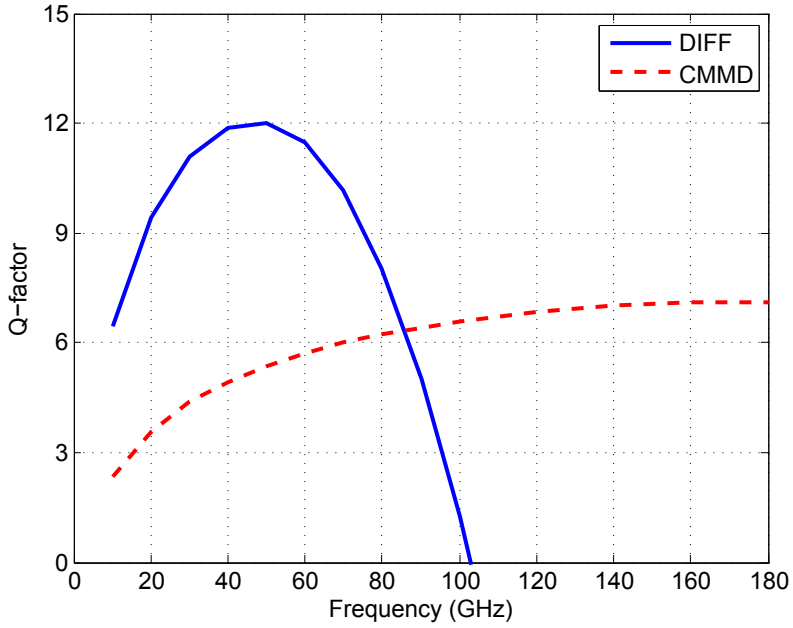
5.4.1 mm-Wave Switching PA-DAC

The key advantage of the direct digital-to-RF conversion architecture is the ability to use switching amplifier for higher PA efficiency. The most attractive switching amplifier class for high frequency applications is the Class-E/F family due to transistor capacitance tolerance [8]. It's shown in Chapter. 2 that second harmonic tuning can significantly improve the PA efficiency, and such tuning class is known as Class-E/F$_{X2}$ with optimal $X2$ being around 0.5. Such tuning requires a passive matching network that can resonate with the transistor drain parasitic capacitance at both the fundamental and the second harmonic frequency. Implementing such a matching network usually involves several series or parallel branches, which not only increases the insertion loss at the fundamental frequency, but also reduces the second harmonic content, as discussed in Chapter. 4 and [30]. Here we propose a simple tuning network that can achieve the desired impedance with minimum complexity. The key component is a two turn inductor. When a two turn inductor is excited by differential signals, the magnetic flux generated by the two windings add constructively, as shown in Fig. 5.13a, therefore the total inductance is 4 times larger than that of a single turn inductor. On the other hand, if the two turn inductor is excited by common-mode signals, the magnetic flux generated by the two windings cancel out each other, as shown in Fig. 5.13b, and the effective inductive area is the enclosed gap between the two windings. This means the common-mode inductance is much smaller than the differential-mode inductance. Since the second harmonic signals in a differential PA are common-mode signals and the required second harmonic load inductance is one quarter of the fundamental load inductance, a two turn inductor can satisfy the tuning requirement by properly adjusting the winding diameter/spacing and bypassing the center tap. Fig. 5.12 shows the differential mode and common-mode inductance for a two turn inductor with 25μm inner diameter and 2.5μm spacing. This inductor is used as the primary side of a 2:1 transformer which performs differential to single-ended signal conversion. The V-I waveforms of the Class-E/F$_2$ PA using the 2:1 transformer is shown in Fig. 5.14 for both square wave input and sine wave input.

Fig. 5.15 shows the schematic of the 2-bit PA-DAC in the first transmitter element. The PA is segmented into four thermometer sub-PA units for direct digital modulation, each unit being a differential pair. The digital switching function is implemented by a bottom switch connected to the source of a differential pair. Each switch is sized roughly twice as large as the combined differential pair size so that the power degradation is reduced to less than 0.2dB. To compensate for the compressive AM-AM distortion present in a typical PA, non-uniform device sizing is applied to the thermometer units. The sizes are chosen to minimize the DNL. Similar to an analog PA, this PA-DAC also exhibits AM-PM distortion due to capacitance variation when the sub-PA units are being switched on and off. To compensate for this effect, a switched capacitor array is added at the differential input of the PA-DAC. One pair of switched capacitor is turned on when one sub-PA unit is turned off. The switched capacitor array has a unit size of 3.4fF and is binary coded. A bottom switch is also inserted in the drive stage differential pair to completely turn off the driver in order to minimize the signal leakage at code zero. This also reduces the power consumption at back-off power levels and improves the efficiency. Finally, an identical copy of the 2-bit PA-DAC is connected in parallel, forming an early and late PA array pair. The late PA

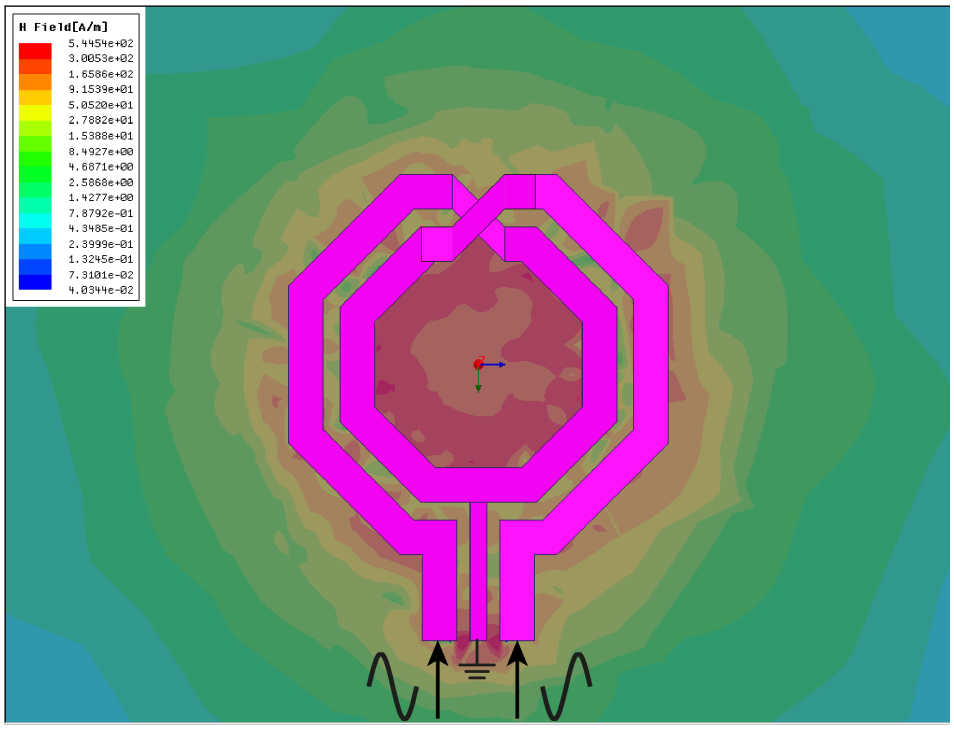


(a) Inductance

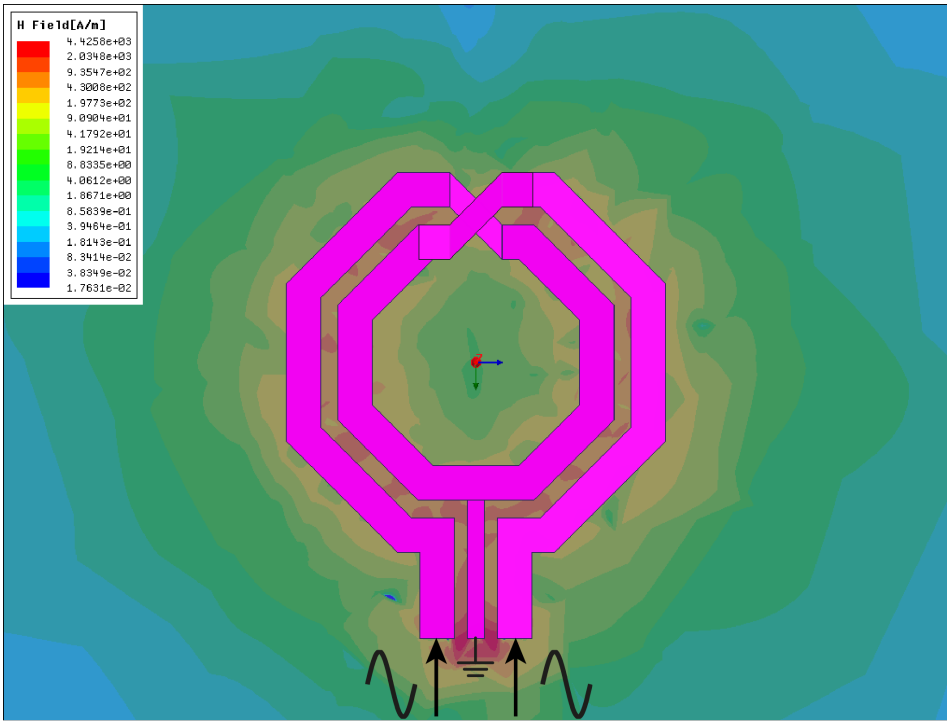


(b) Q-factor

Figure 5.12: Inductance and Q-factor of a two turn inductor ($25\mu m$ inner diameter and $2.5\mu m$ spacing).

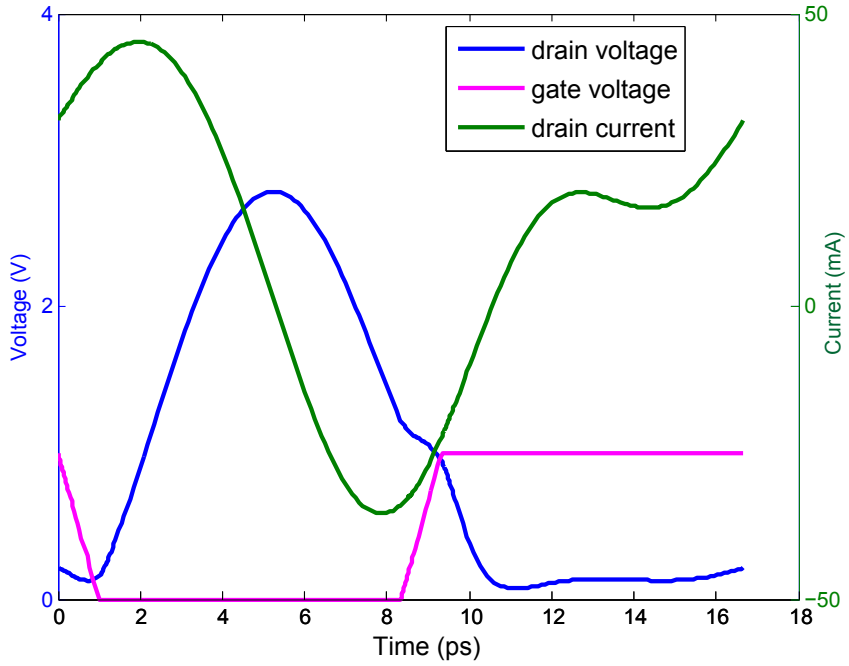


(a) Differential mode

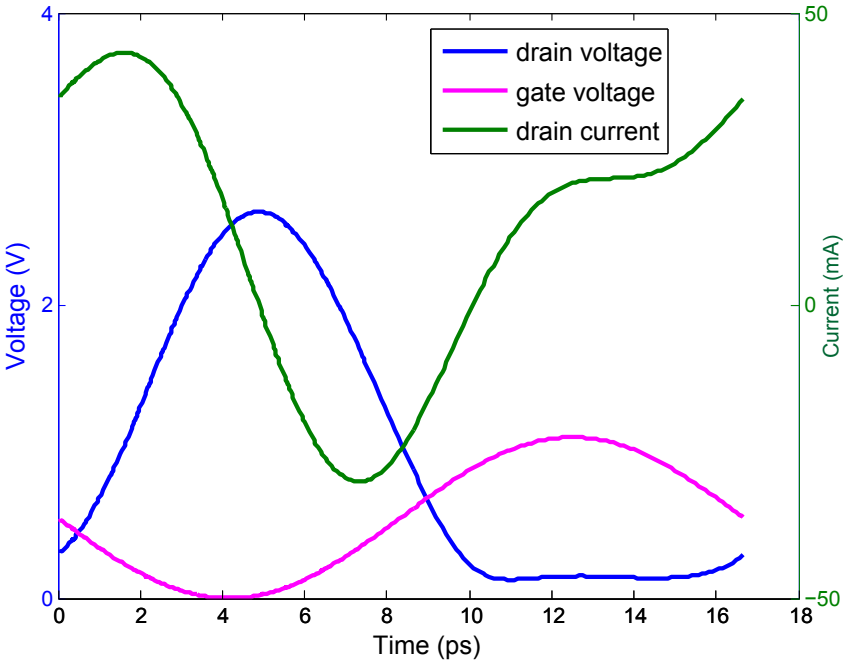


(b) Common mode

Figure 5.13: Magnetic field of a two-turn inductor when exited by differential signals and common-mode signals



(a) Square wave drive



(b) Sine wave drive

Figure 5.14: V-I waveforms of the Class-E/F_2 PA using 2:1 transformer.

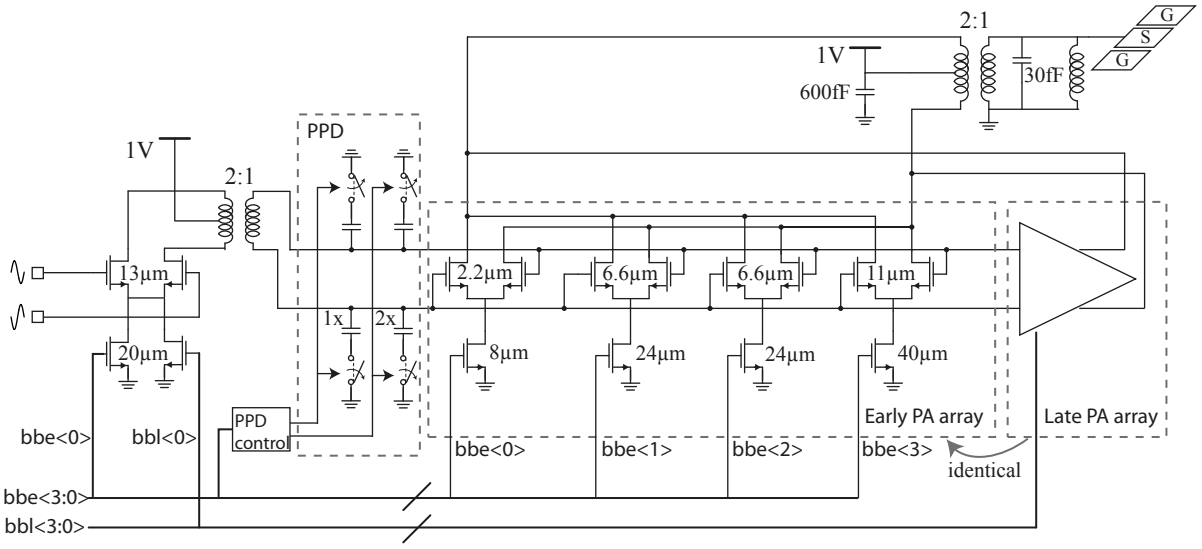


Figure 5.15: Schematic of the PA-DAC in the first transmitter element.

array is always digitally activated and de-activated half a baseband clock cycle later than the early array. This two-fold operation approximates a first-order-hold (FOH) digital-to-analog conversion for lower spectral images. This particular part will be further discussed in the mixed-signal filtering subsection. The MSB three transmitter elements utilize the same design as the first element but with all the differential pairs tied together and controlled by a single coarse amplitude bit.

5.4.2 Phase Shifters

Unlike a traditional phased array where the phase shifter resolution is set to achieve the desired directivity and therefore can be relatively coarse, the phase shifter resolution in a quadrature spatial combined transmitter directly determines the worst case I/Q phase imbalance when the information beam is being steered. As a result, a high resolution phase shifter is needed to minimize the I/Q phase imbalance. Fig. 5.16 shows the transmitter EVM contour as a function of I/Q amplitude and phase imbalance. To obtain an EVM better than -20dB for 16QAM signal while allowing amplitude mismatch of 1dB, the phase mismatch needs to be smaller than 9°. Note that this EVM calculation does not include any non-idealities such as non-linearity and noise. Therefore, even better phase resolution is required. In this design, an 8-bit I/Q interpolating type active phase shifter is implemented with an effective phase resolution better than 2° (Fig. 5.17). The required quadrature LO signals for the phase shifter are generated by a quadrature hybrid [30]. Two cascaded 1:2 transformers are used to interface the hybrid and the LO phase shifter. The cascaded transformer performs the matching between the high impedance LO port of the phase shifter and the 50Ω output of the hybrid. Besides, the cascaded transformer also acts as a balun with improved common-mode rejection [46].

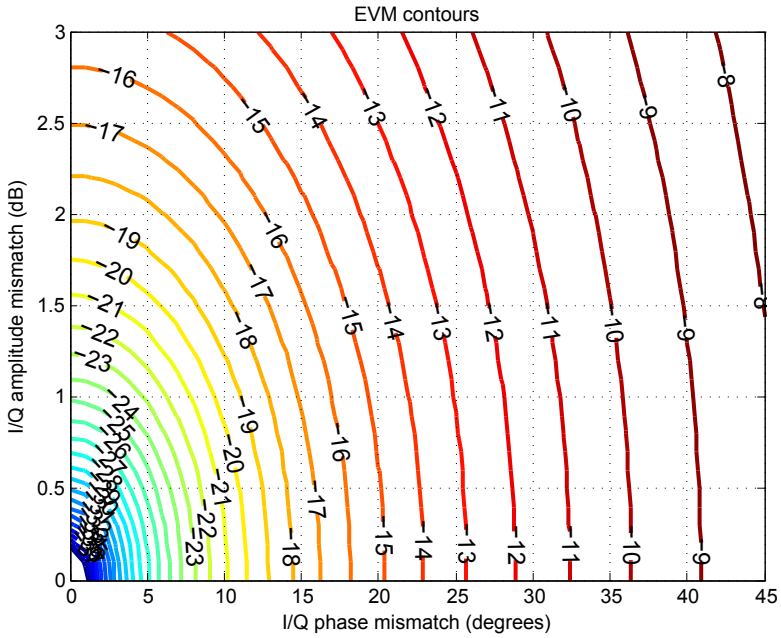


Figure 5.16: EVM contour as a function of I/Q amplitude and phase imbalance.

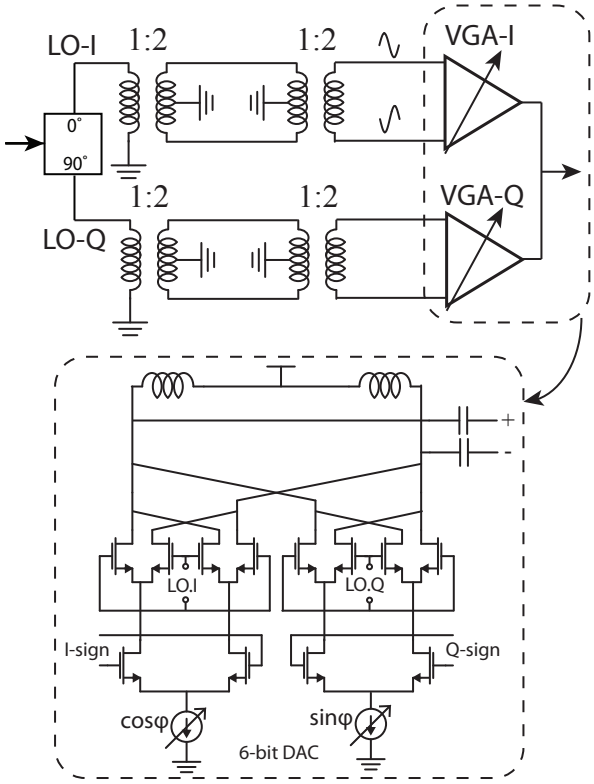


Figure 5.17: Schematic of the LO phase shifter.

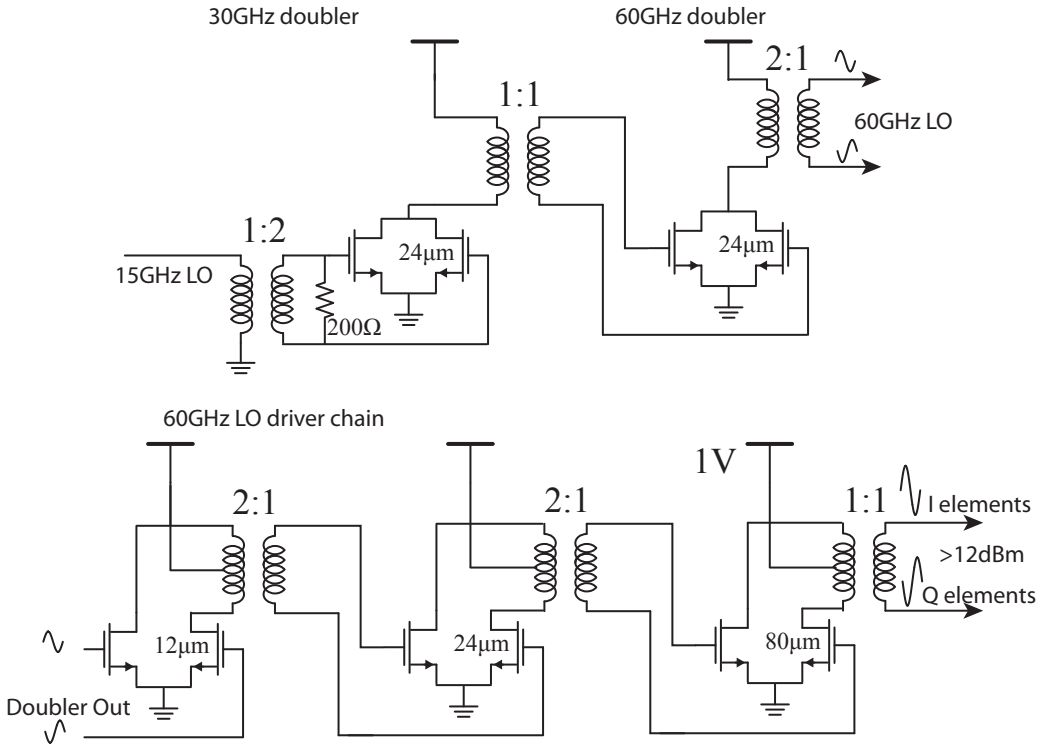


Figure 5.18: Schematic of the LO frequency doublers and the driver chain.

5.4.3 LO Generation and Distribution

The 60GHz signal is derived from an external 15GHz signal through two stages of frequency doublers. Each frequency doubler uses common-source push-push architecture, as shown in Fig. 5.18. The single-ended signal is then converted to differential signal to feed the next stage through a balun. The loss of the two stages is around 10dB with 0dBm input power level. To drive eight transmitter elements, three driver stages are used to boost the output power level to 12dBm (Fig. 5.18). The differential output signals of the LO driver chain are distributed to the I/Q elements by a chain of Wilkinson power dividers. The downside of Wilkinson dividers is the large footprint. To reduce the area, lumped versions are implemented. Fig. 5.19 shows the schematic of a lumped Wilkinson divider. To approximate a distributed Wilkinson, the required inductance and capacitance values are,

$$L = \frac{Z_T}{\omega_0} \tag{5.10}$$

$$C = \frac{1}{Z_T \omega_0} \tag{5.11}$$

For standard Wilkinson dividers, both input and outputs are matched to $Z_0 = 50\Omega$, therefore the characteristic impedance of the transmission Z_T needs to be $\sqrt{2}Z_0$. On the other hand, the final stage LO driver delivers large output power and therefore needs to see a low impedance. If standard Wilkinson dividers were used, the total load impedance would be

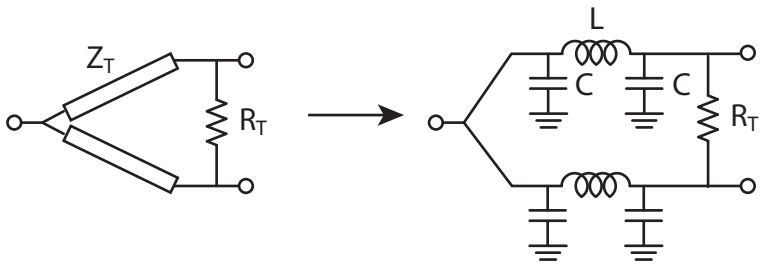


Figure 5.19: Lumped Wilkinson power dividers.

100Ω, which is too large. As a result, asymmetrical IO impedance is chosen for the first two Wilkinson dividers. The input impedance is set to 25Ω while the output impedance is kept 50Ω. This leads to a total load impedance of 50Ω for the LO driver. Such an asymmetrical divider can be realized based on the following IO impedance equations,

$$Z_{in} = \frac{1}{2}\frac{Z_T^2}{Z_L} \tag{5.12}$$

$$Z_{out}^{odd} = \frac{1}{2}R_T \tag{5.13}$$

$$Z_{out}^{even} = \frac{Z_T^2}{2Z_S} \tag{5.14}$$

where Z_L and Z_S are the load and source impedance respectively. By setting transmission line impedance Z_T to $Z_0 = 50\Omega$ and termination resistance R_T to $2Z_0 = 100\Omega$, the desired input and output impedance can be achieved.

$$Z_{in} = \frac{1}{2}\frac{Z_0^2}{Z_0} = \frac{1}{2}Z_0 \tag{5.15}$$

$$Z_{out}^{odd} = \frac{1}{2}(2Z_0) = Z_0 \tag{5.16}$$

$$Z_{out}^{even} = \frac{Z_0^2}{2\frac{1}{2}Z_0} = Z_0 \tag{5.17}$$

5.4.4 Mixed-Signal Baseband Signal Processing

Digital to analog conversions result in spectral images at multiples of the sampling frequency, therefore filters are required to suppress the images in order to meet the transmitter mask specified by the standard. Due to the lack of analog baseband signal in a direct digital-to-RF conversion transmitter, traditional analog low-pass filters cannot be used. This demands that the filtering be performed in the discrete time domain. As a result, an oversampling FIR filter is required. On the other hand, even after digital filtering, spectral images

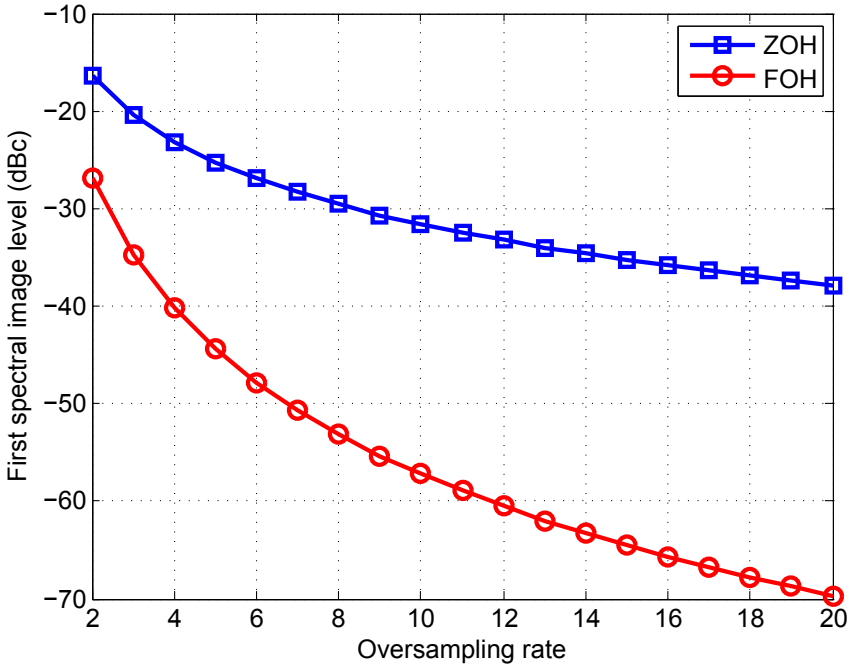


Figure 5.20: First spectral image level as a function of the oversampling rate.

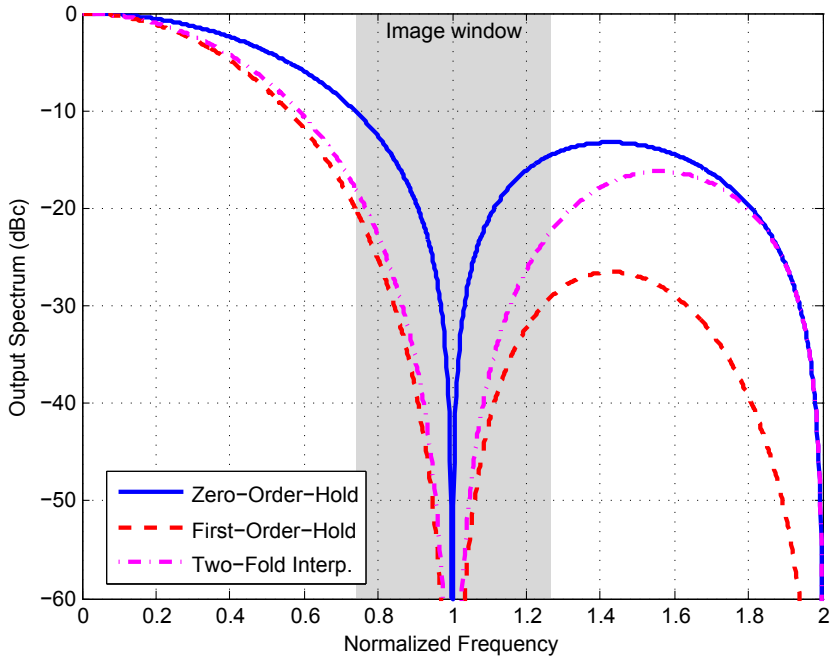


Figure 5.21: Output spectrum of different types of digital to analog conversion.

still exist at multiples of the oversampled clock frequency. The largest spectral image level of a Zero-Order-Hold (ZOH) digital to analog converter is a function of the oversampling rate (OSR),

$$P_{image} = 20\log(\text{sinc}(1 - \frac{1}{2OSR})) - 6 \tag{5.18}$$

As the WiGig standard requires out-of-band emission lower than -30dBc, an oversampling rate of 8 is required if traditional ZOH is used. However, running the digital filter at 14GHz ($1.76G \times 8$) is impractical. In addition, distributing such high speed digital signals costs significant power consumption. To remedy this problem, a mixed signal approach is used. Compared to traditional ZOH digital to analog conversion whose output spectrum exhibits a sinc rolloff, FOH conversion exhibits a sinc^2 rolloff and therefore provides additional suppression on the spectral image, as shown in Fig. 5.20. This means lower oversampling rate is required for the same image level. In this design, an oversampling rate of 4 is chosen. To simplify the implementation of a FOH conversion, multi-fold linear interpolation can be used as a close approximation [39]. Since the first spectral image is usually much larger than other images, two-fold interpolation is sufficient to suppress this dominant image. Fig. 5.21 compares the output spectrum of ZOH, FOH and two-fold interpolation. Clearly, both FOH and two-fold interpolation exhibit lower spectral amplitude than ZOH near the sampling frequency, and the two-fold interpolation well matches the FOH within the spectral image window. The implementation of the entire mixed-signal baseband signal processing unit is shown in Fig. 5.22. The FIR consists of four sub-FIRs, each running at the Nyquist sampling frequency to reduce power consumption. A 4:1 serializer is used at the output to produce the oversampled data stream. The output of the serializer is divided into two parallel bit streams, an early signal and a late signal. Two-fold linear interpolation requires half baseband clock cycle delay between the two bit streams, which is achieved by adding one additional latch in the late signal path. The early and late digital signals are distributed to eight transmitter elements to activate/de-activate the early and late PAs. A MUX is inserted to enable FIR and two-fold interpolation bypassing for comparison.

The detailed serializer schematic is shown in Fig. 5.23. It consists of two stages of 2:1 serializers. Latches are inserted in front of the MUX in order to ensure the serialization sequence. Due to additional one latch delay, bits going through "1" input of the MUX always get serialized before bits going through "0" input of the MUX. The 1x and 2x clocks are derived from two frequency dividers. Note that the frequency dividers need to be shared between the I and Q baseband units in order to align the serialized I and Q data.

5.4.5 Supply Bypass Network

Due to dynamic current scaling, the supply voltage experiences significant ripples. The supply ripples are mainly caused by the inductance introduced by bondwires or micro-bumps. Since the supply ripple increases with the loop inductance, low inductance packaging such as flip-chip is preferred. On the other hand, to minimize the impact of supply noise on the transmitter EVM, the supply path needs to have a fast settling time. The supply bypass

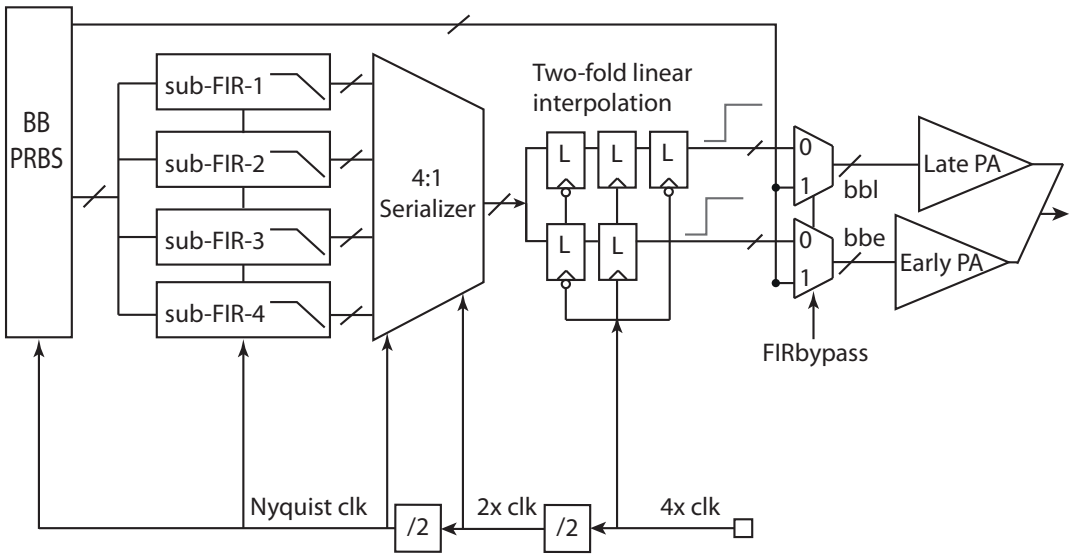


Figure 5.22: Mixed-signal baseband signal processing for spectrum filtering.

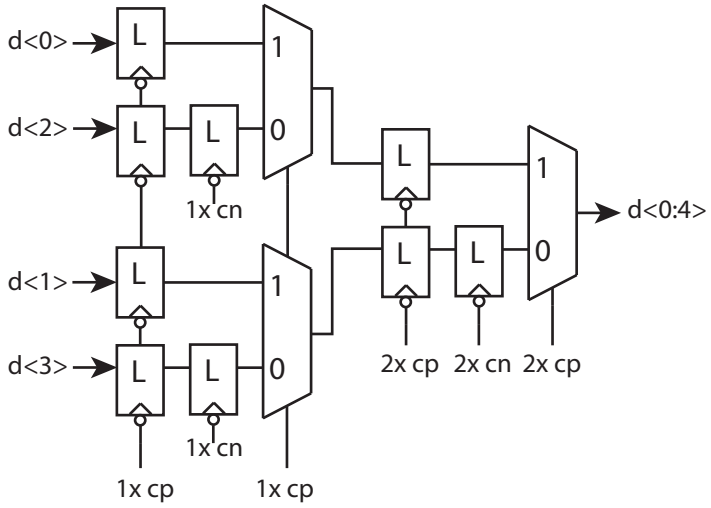


Figure 5.23: Schematic of the 4-to-1 serializer.

network can be modeled by a second order RLC circuit with a Q-factor of $\sqrt{\frac{L}{C}}\frac{1}{R}$. For a given LC, the settling time can be minimized by setting the second order system slightly overdamped (Q≈1). Under that condition, the settling time can be approximated by,

$$t_s \approx \frac{\ln(\epsilon)Q}{\omega_n} = \frac{\ln(\epsilon)L}{R} \tag{5.19}$$

where ϵ is the relative error tolerance. Apparently the settling time can be minimized by increasing R. However, digital amplitude code switching results in delta pulses on the supply voltage and the pulse amplitude is proportional to R. As a result, R cannot be arbitrarily

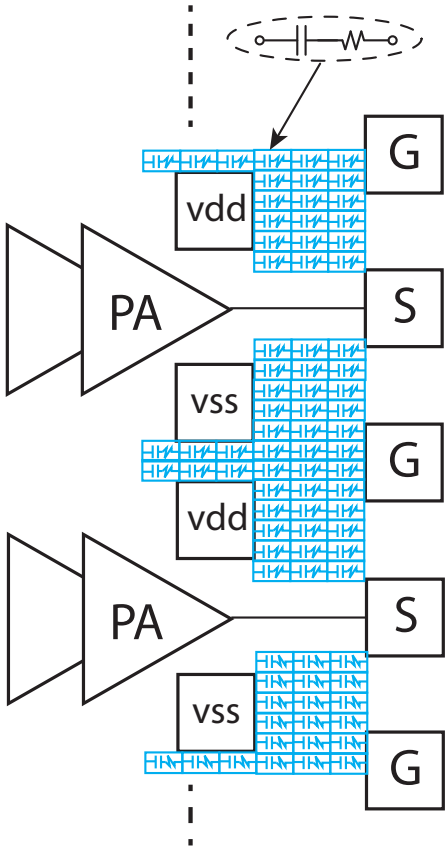


Figure 5.24: Supply bypass network floorplan.

increased. In this design, a series combination of 200pF bypass capacitor and 1.5Ω resistor is used for each transmitter element so that the worst-case voltage pulse is less than 250mV and the pulse can be settled within 350ps. Fig. 5.24 shows the bypass network floorplan. Each transmitter has a vdd/vss bump pad pair and the de-Qued bypass capacitor is placed between them.

5.5 Experimental Results

The beamforming transmitter is fabricated in a bulk 65nm CMOS technology with no special RF options and the die measures 3×3 mm^2. Fig. 5.25 shows the chip micrograph. This section is divided into two parts. The first part describes the Continuous-Wave (CW) mode measurement results from on-die probing and the second part describes the modulation test results from the mm-wave package.

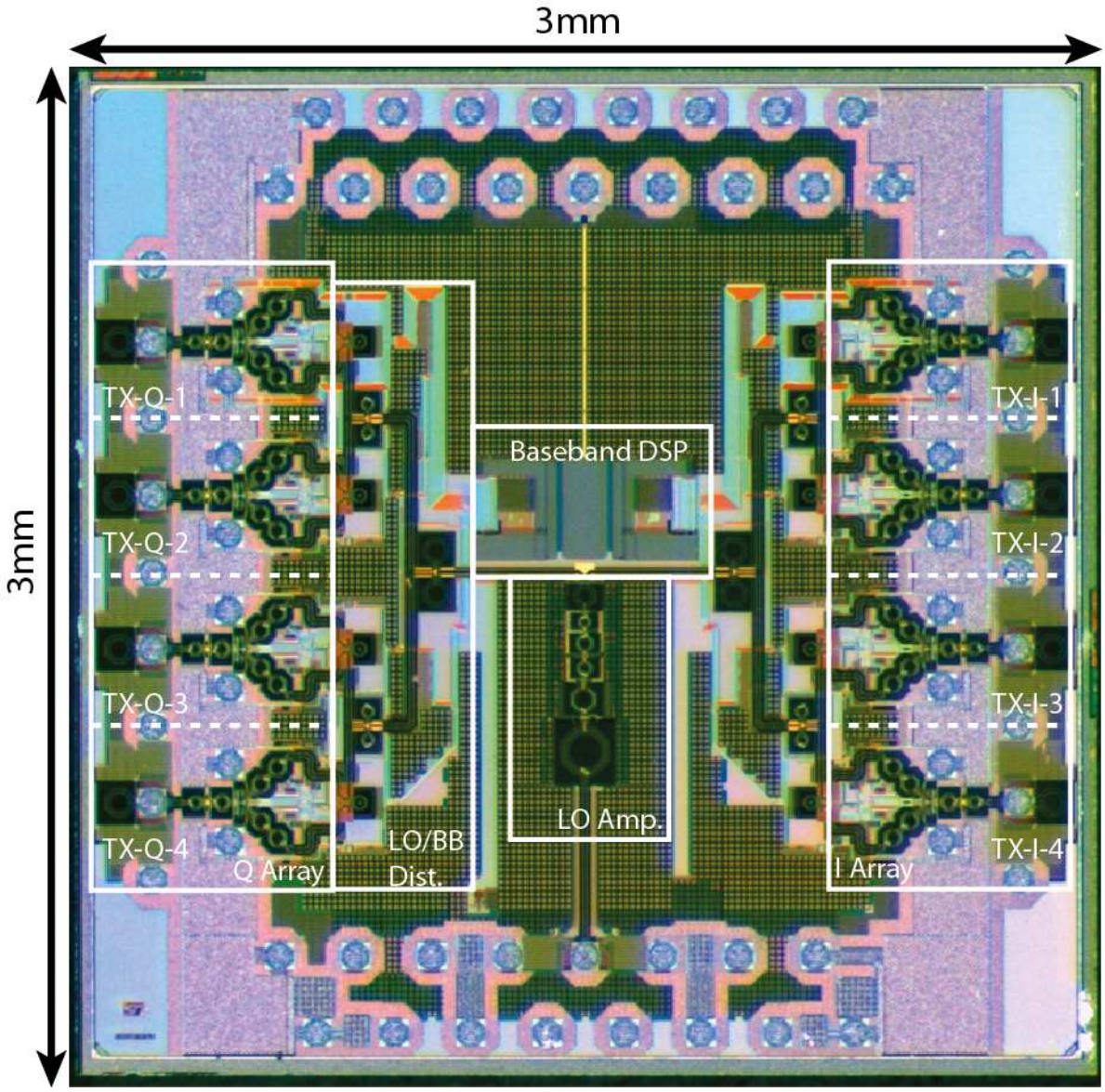


Figure 5.25: Chip Micrograph.

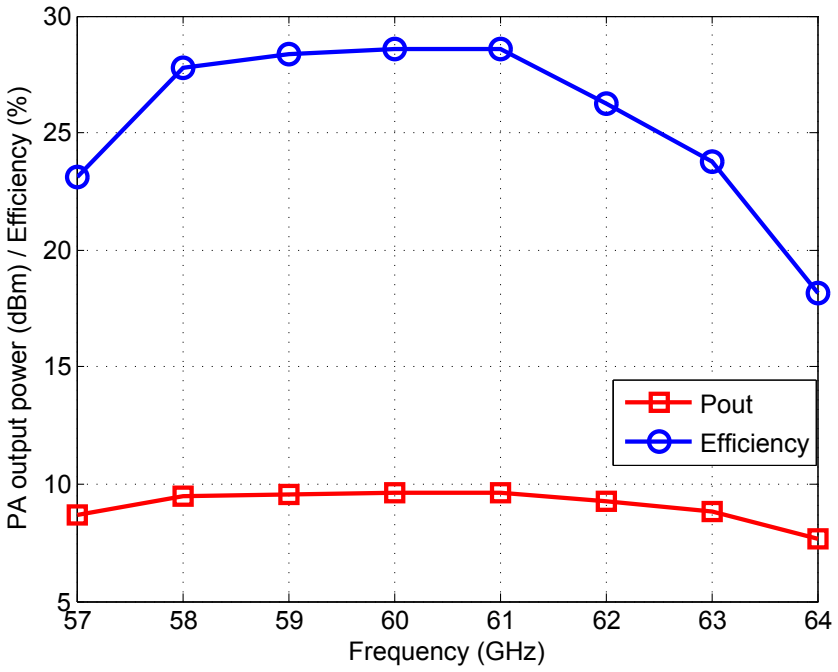
5.5.1 CW Mode Measurement Results

The CW mode performance is measured by probing the output of each transmitter element. Fig. 5.26 shows the PA peak output power and peak drain efficiency across the WiGig frequency band. The peak output power of each element is around 9.6dBm and the peak drain efficiency is 28.5%. The back-off characteristic of this transmitter can be observed in the PA efficiency versus the output power plot in Fig. 5.26. The measured back-off curve closely matches a Class-B PA behavior, confirming the theoretical prediction. Note that at 6dB CW power back-off, the PA still has an efficiency of 15%, which is twice of what the efficiency would be if this were a Class-A PA. The digital AM-AM and digital AM-PM curve of the first transmitter element is shown in Fig. 5.27. Due to non-uniform device sizing on the sub-PA thermometer cells, the PA output voltage amplitude is very linear with respect to the digital amplitude codeword, and the difference between I and Q PAs is negligible. When the phase-pre-distortion is off, the worst case AM-PM distortion is more than 30°. By enabling the phase pre-distortion, the AM-PM distortion can be reduced to 10°. Fig. 5.28 shows the output power of the 8 transmitter elements. Since the supply and ground pads for inner two elements are not wire bonded, the output power of the measured inner elements are slightly lower than the outer elements. However, this difference is eliminated in the flip-chip package in which IR drop for all the elements are comparable. The LO phase shifter performance was measured by comparing the relative delay of two transmitter element output signals. Fig. 5.29 shows the measured codeword to phase transfer curve and the corresponding phase step. The phase shifter has a RMS phase resolution of 2.2° and worst case phase step of 5°. The fluctuation in the phase step is mainly due to the timing accuracy of the sub-sampling oscilloscope.

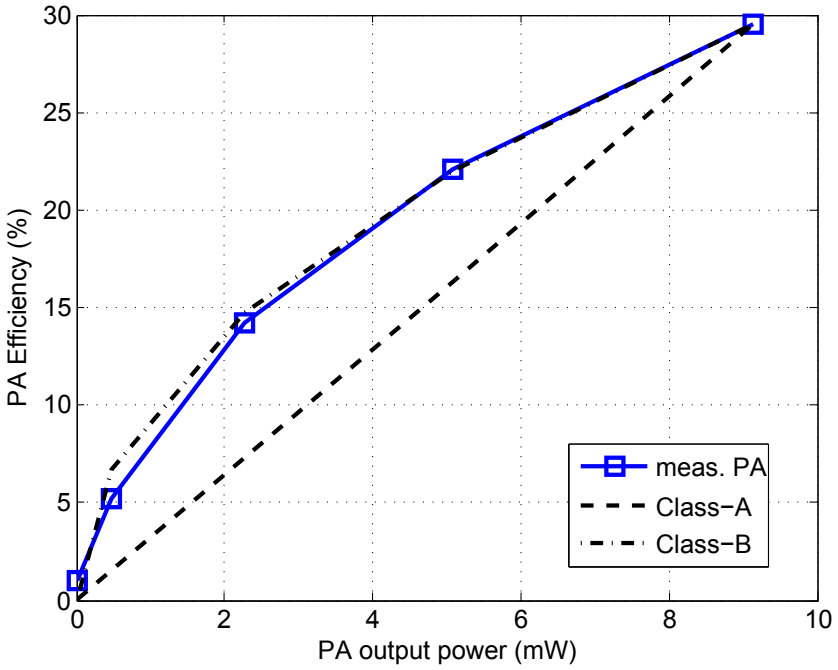
5.5.2 Package Measurement Results

To verify the concept of quadrature spatial combining, a complete mm-wave module was also designed and fabricated, as shown in Fig. 5.30. The package consists of a 3-layer PCB made of Rogers 4350 material. The silicon die is attached to the front-side of the package in a flip-chip configuration and the patch antenna arrays are printed on the back-side. The patches are aperture coupled from the feed lines on the front-side. Unassembled PCBs are used for antenna characterization. The measured antenna radiation pattern and gain is shown in Fig. 5.31. The measured H plane half power beamwidth is 70°. The peak antenna frequency response is shifted to lower frequency, with a peak gain of 4.4dBi at 54GHz. The gain at 60GHz is around 2.5dBi with a 3dB bandwidth of 3GHz. The total radiation efficiency is 43%, including the loss of the feed lines. Due to a modeling error, the inner two patches have lower gain (around 6dB) than the outer two patches.

The measurement setup for characterizing the mm-wave package is shown in Fig. 5.32. A horn antenna is placed in the far field to receive the spatially reconstructed signal radiated from the mm-wave module. A subharmonic mixer is used to down-convert the V-band signal to an IF frequency of 2GHz. The IF signal is then sent to the oscilloscope and spectrum

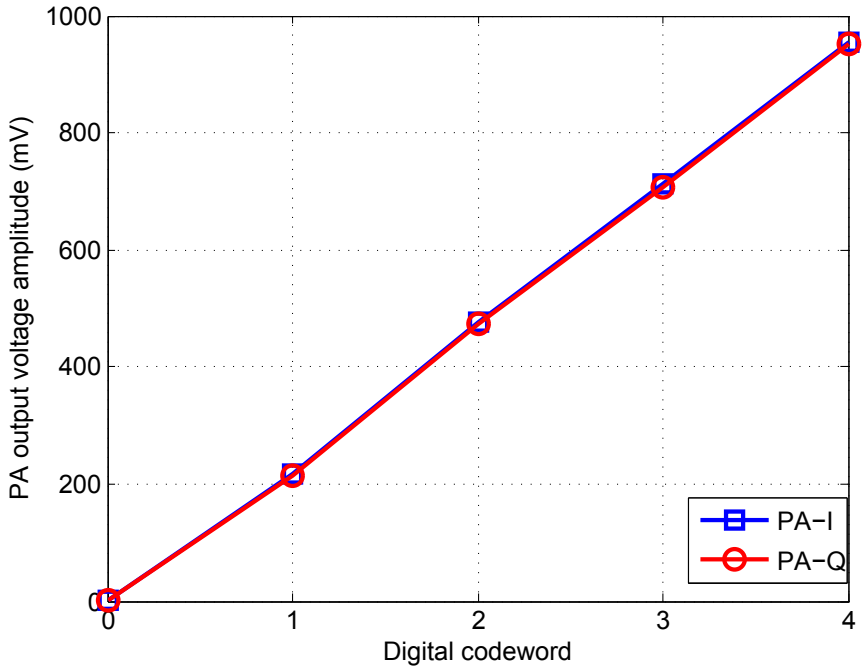


(a)

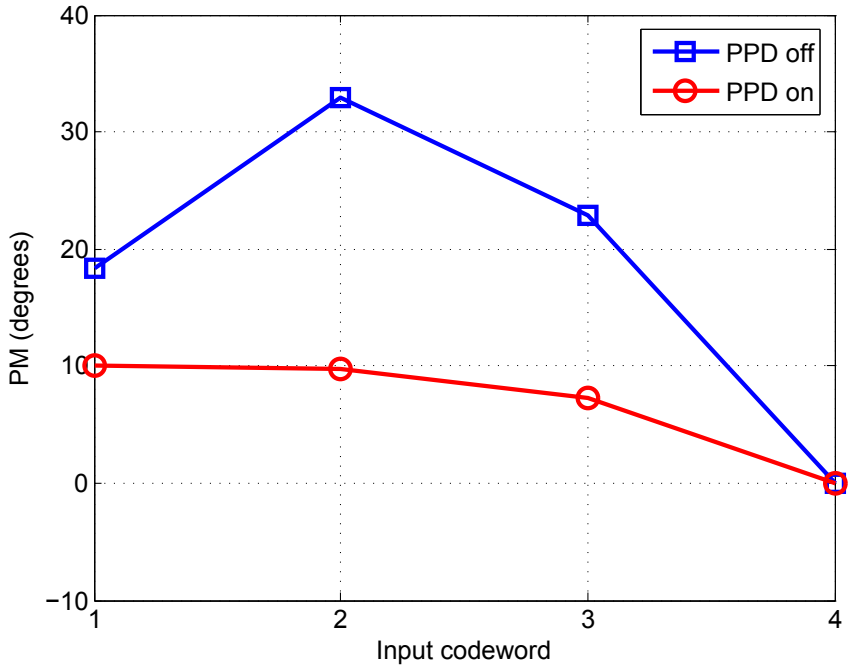


(b)

Figure 5.26: Measured transmitter output power and drain efficiency.



(a)



(b)

Figure 5.27: Measured first transmitter digital AM-AM and digital AM-PM behaviors.

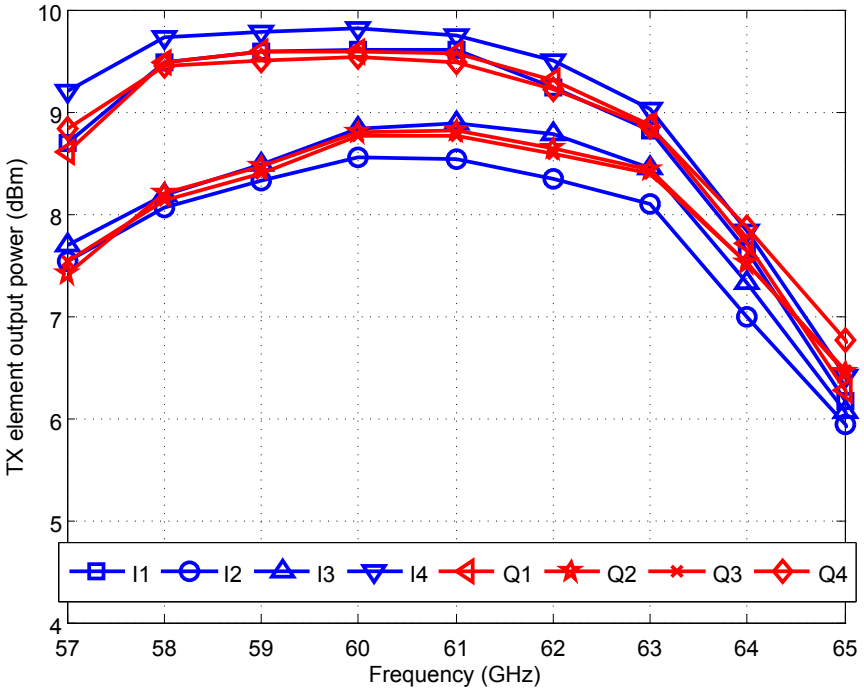


Figure 5.28: Measured output power of 8 transmitter elements.

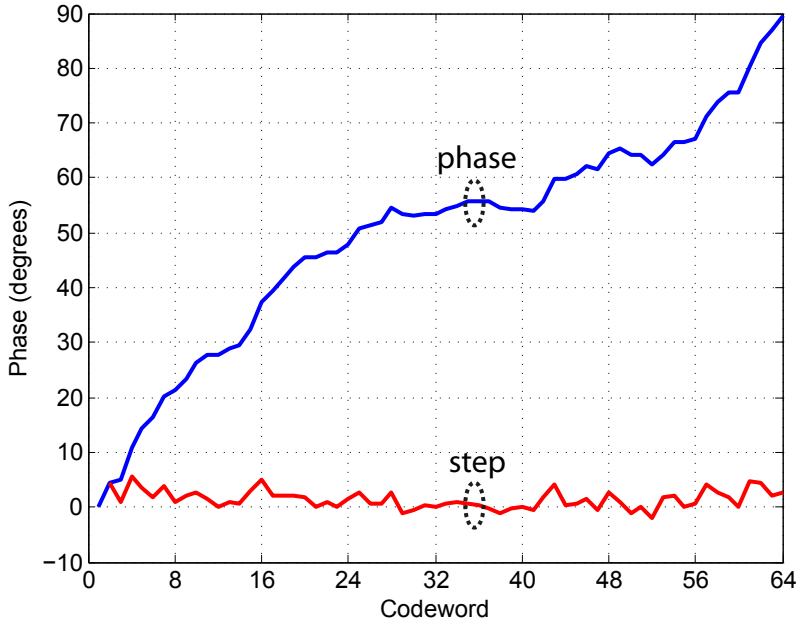


Figure 5.29: Measured LO phase shifter codeword to phase transfer curve and the corresponding phase step.

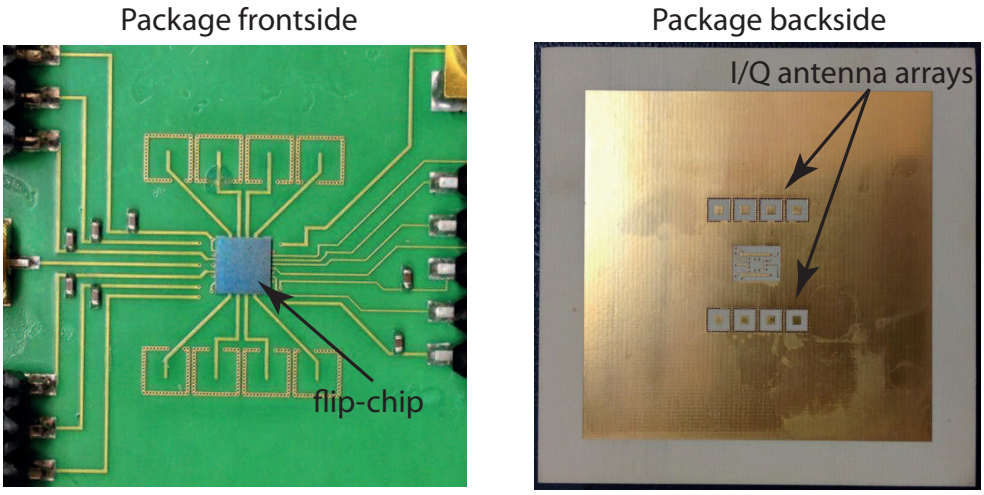


Figure 5.30: Flip-chip packaged module with antenna arrays.

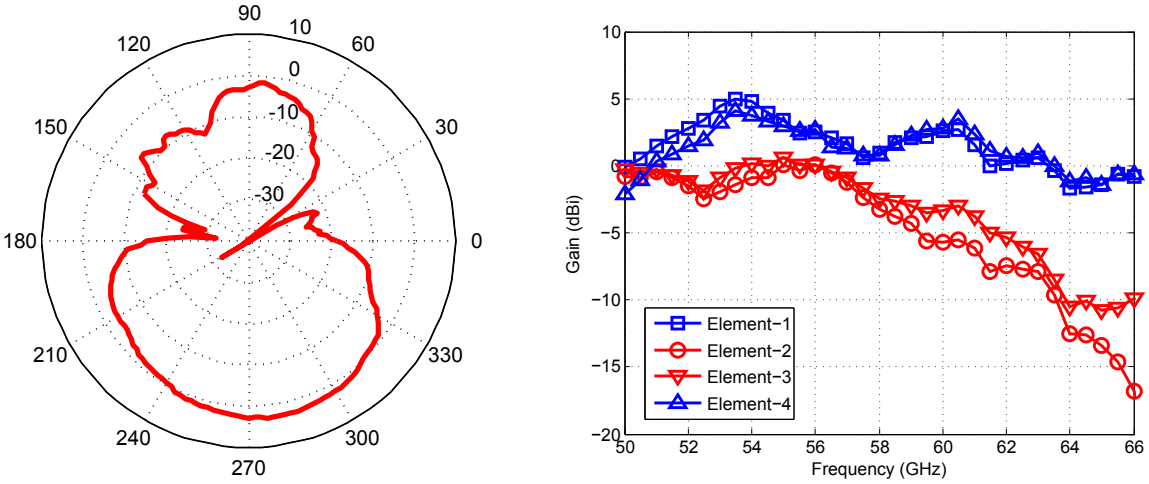


Figure 5.31: Measured antenna radiation pattern and freqeuncy response.

analyzer for EVM and spectrum analysis. A V-band power meter is also used to measure the EIRP and the digital AM- RF AM curve.

The measured peak EIRP is 22dBm at 60GHz in the broadside direction. Fig. 5.33 shows the digital AM - RF AM curve of the entire transmitter array, which is normalized to the peak value (22dBm). It can be observed that the curve is piece-wise linear within 4 codes, re-confirming the probing measurement results. However, due to reduced antenna gain of the inner two patches, the AM-AM curve exhibits two non-monotonicity at code 5 and code 9. Although these codewords are avoided at Nyquist samples by digital pre-distortion to prevent EVM degradation, there are residual effects such as spectral re-growth and increased out-of-band emission. This error can be eliminated once the inner patches are properly redesigned.

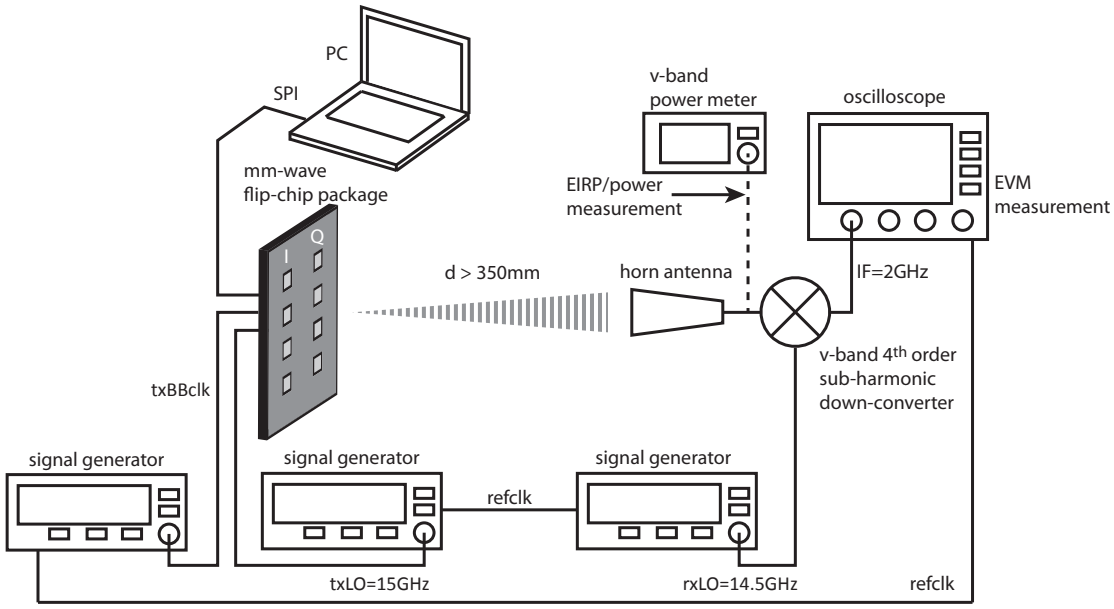


Figure 5.32: Measurement setup for wireless transmission using the mm-wave module.

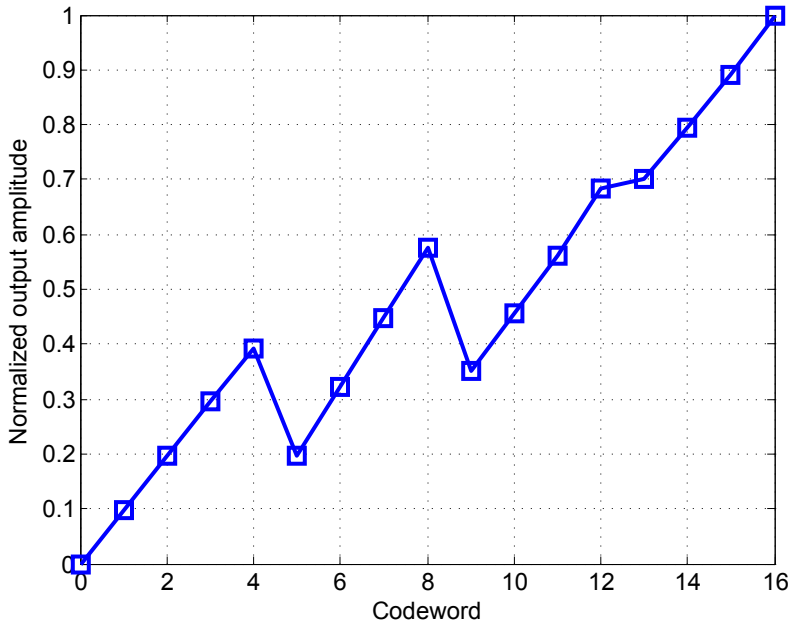


Figure 5.33: Measured radiated transmitter amplitude as a function of amplitude codeword.

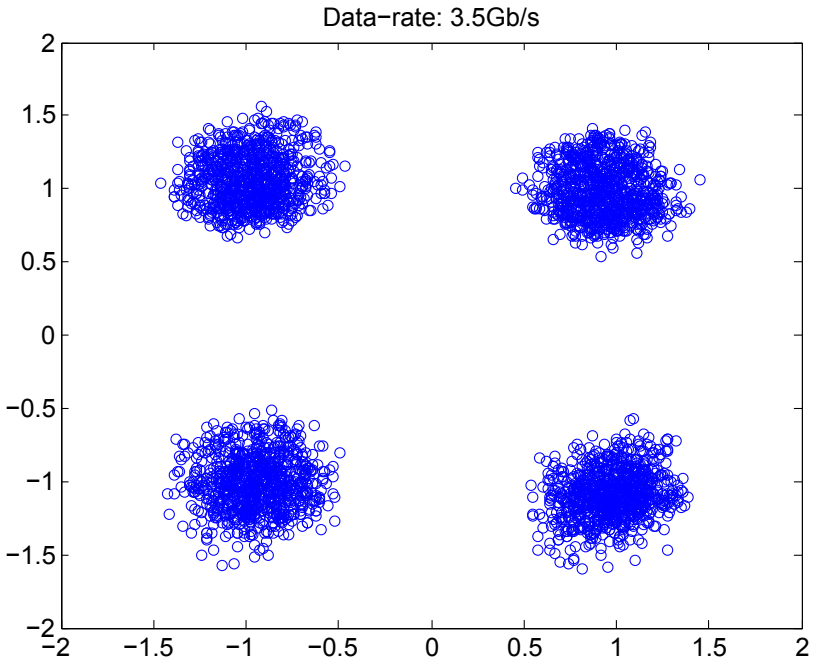


Figure 5.34: Received signal constellation for QPSK modulation at 3.5Gb/s.

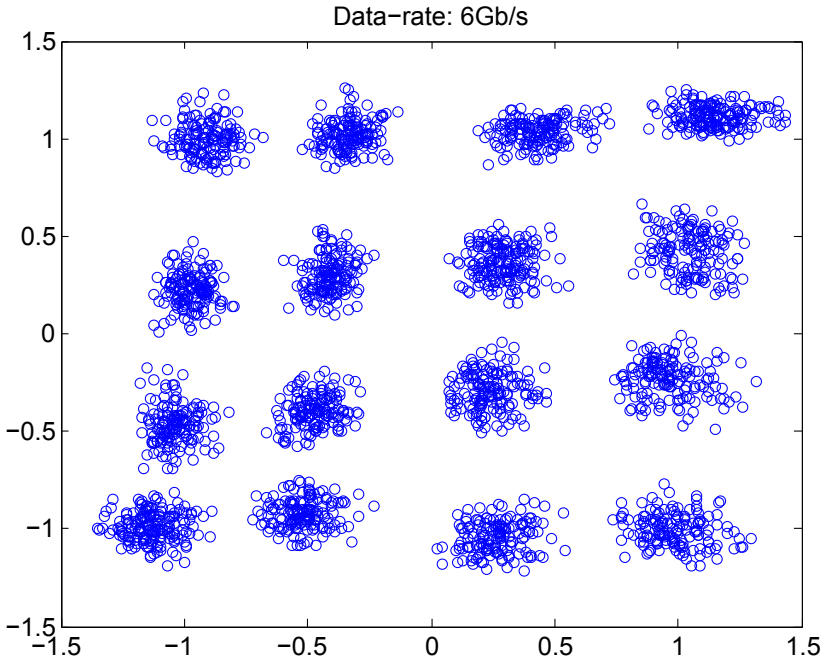


Figure 5.35: Received signal constellation for 16QAM modulation at 6Gb/s.

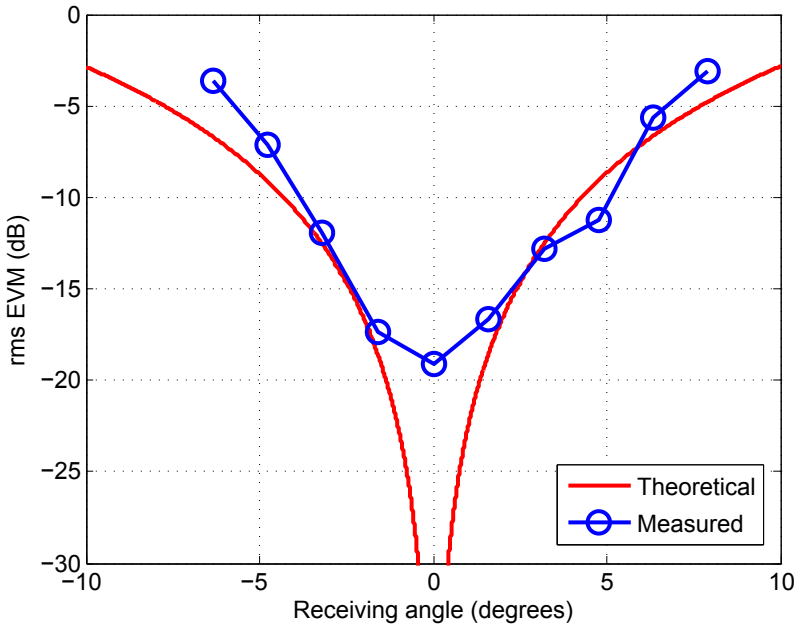
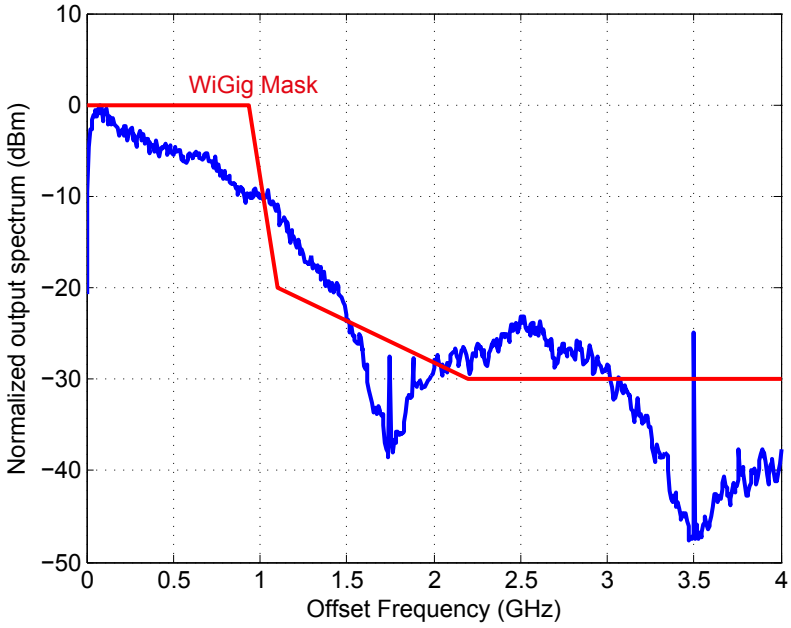


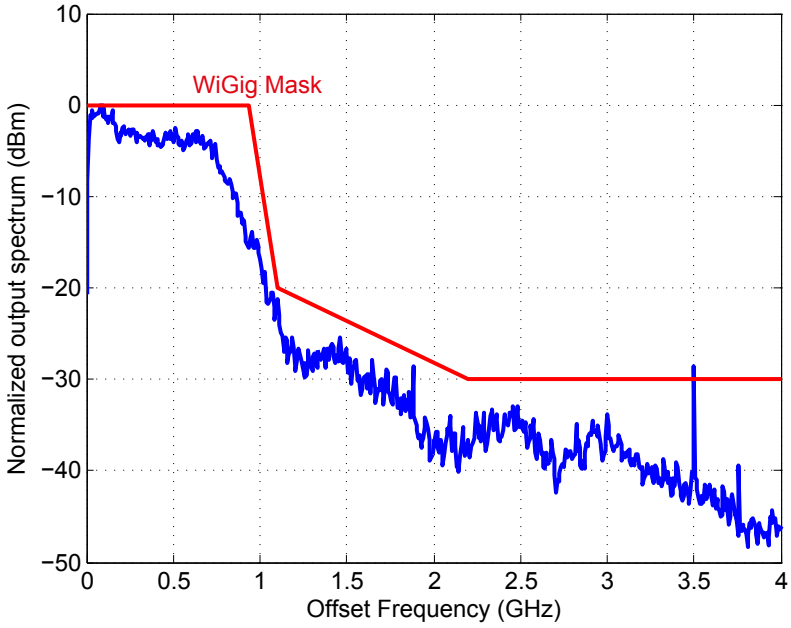
Figure 5.36: Transmitter output EVM as a function of spatial angle.

Modulated signal transmission was conducted by sending an I/Q pseudo random bit sequence (PRBS) from the transmitter and recovering the data by the aforementioned external receiver. Fig. 5.34 plots the received signal constellation for QPSK modulation at 3.5Gb/s data rate with a carrier frequency of 60GHz. The corresponding EVM is -15dB. Similarly, the constellation for 16QAM modulation is shown in Fig. 5.35. The data rate for 16QAM modulation is 6Gb/s and the EVM is -16.2dB. The average EIRP is around 16.4dBm. The transmitter EVM was also tested as a function of the receiving angle. Fig. 5.36 shows the measured EVM pattern when the information direction is centered around 0°. QPSK modulation is used in this experiment with a carrier frequency of 60GHz and a data rate of 2Gb/s. The measured EVM spatial behavior matches closely with the theoretical prediction. Finally, the mixed-signal baseband signal processing function is verified by monitoring the transmitter output spectrum. Fig. 5.37 and Fig. 5.38 show the downconverted output spectrum for QPSK and 16QAM modulation schemes respectively. Comparing the spectrums before and after the FIR and interpolation are turned on, it's clear that the mixed-signal baseband signal processing can effectively suppress the image and lower the noise floor. With mixed-signal filtering on, the output spectrum is compliant with the WiGig mask.

The entire circuit is powered from 1V supplies. The I/Q PA arrays consume a peak DC power of 229mW and the entire transmitter consumes a peak power of 382mW. The power consumption of other blocks are summarized in Table. 5.1. The measured peak PA efficiency is 28.5% and the peak transmitter efficiency is 17.4%. The average efficiency of

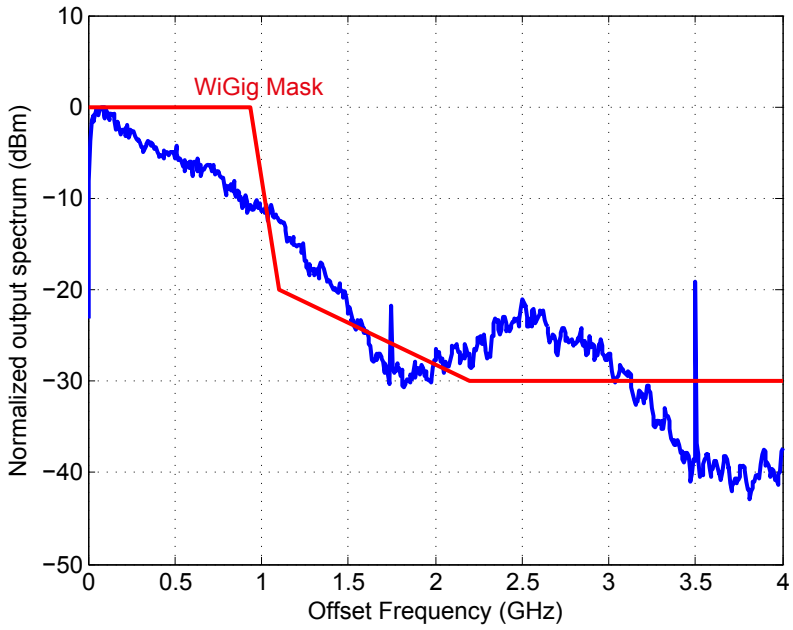


(a) FIR and Interp OFF

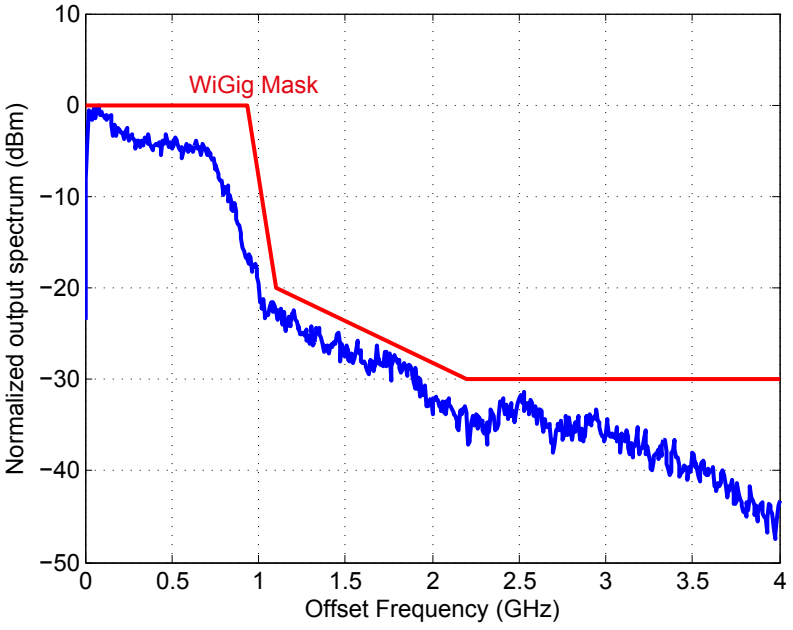


(b) FIR and Interp ON

Figure 5.37: Measured QPSK transmitter output spectrum.



(a) FIR and Interp OFF



(b) FIR and Interp ON

Figure 5.38: Measured 16QAM transmitter output spectrum.

Block	Power (mW)
PA	229
Phase Shifter	48
BB Signal Processing	40
BB Signal Distribution	30
LO Drivers	35
Total	382

Table 5.1: Chip power breakdown.

CW Mode	
Peak EIRP (dBm)	22
TX Element Pout (dBm)	9.6
PA peak efficiency (%)	28.5
TX peak efficiency (%)	17.4
Total peak DC power (mW)	382
Data transmission at 6dB backoff	
PA average efficiency (%)	16.5
TX average efficiency (%)	7
Peak data rate (Gb/s)	6
QPSK EVM	-15
16QAM EVM	-16.2
TX average DC power (mW)	260

Table 5.2: Chip performance summary.

the transmitter is defined by the following equation,

$$\eta_{average} = \eta_{peak} \frac{EIRP_{average}}{EIRP_{peak}} \frac{P_{DC,peak}}{P_{DC,average}} \tag{5.20}$$

where η_{peak} is the peak efficiency of the transmitter, which can be measured by probing the CW output power of each transmitter. Note that EIRP is used instead of total output power to account for the antenna array gain change in such a digital antenna system during the transmission of a modulated signal. In a single antenna transmitter, the equation falls back to the original form,

$$\eta_{average} = \eta_{peak} \frac{P_{o,average}}{P_{o,peak}} \frac{P_{DC,peak}}{P_{DC,average}} = \frac{P_{o,average}}{P_{DC,average}} \tag{5.21}$$

At 6dB back-off when transmitting QPSK or 16QAM data, the measured average PA efficiency is 16.5% and the average transmitter efficiency is 7%, based on Eq. 5.20. Table. 5.2

	[37]	[38]	[47]	**This Work**
TX architecture	Outphasing	Outphasing	Class-AB	Cartesian digital-to-RF
Signal reconstruction	Spatial combining	Transformer combining	No need	Spatial combining
Single PA P_{sat} (dBm)	9.7	15.6	10	9.6
η_{PA} at P_{sat} (%)	11	25	10.8	28.5
η_{PA} at 6dB BO[a] (%)	N.A.	9	5.7[b]	15
TX beamforming?	Yes (2 PAs)	No	No	Yes (8 PAs)
Antenna package?	No	No	No	Yes (EIRP=22dBm)
Modulation	16QAM	16QAM	16QAM	16QAM
Peak data rate (Gb/s)	0.2	0.5	7	6
Technology	65nm	40nm	40nm	65nm

Table 5.3: Comparison to efficiency enhancing 60GHz transmitters.

[a] CW mode
[b] At 5dB back-off

summarizes the chip performance and Table. 5.3 compares this work to other state-of-the-art 60GHz transmitters with various efficiency enhancing techniques. Clearly, substantial improvement has been achieved in terms of average efficiency.

5.6 Digital Calibration and Pre-distortion

Digital pre-distortion and calibration are required to linearize and compensate the direct digital-to-RF transmitter, as a result it's critical to analyze all the potential non-idealities and figure out the corresponding algorithms. Some non-idealities in this architecture are common to those in traditional phased array transmitters or single-element transmitters and therefore can be corrected by existing algorithms while others require additional training or calibration.

As evident in the measurement results, the largest mismatch in the systems comes from antenna mismatch, due to PCB manufacturing tolerances. This affects the transmitter performance in a number of ways.

Gain mismatch between I and Q antennas results in I/Q amplitude imbalance, which is a common issue in any transmitter. This can be calibrated either in the digital domain

by re-adjusting the full-scale of each I/Q DAC or in the analog domain by re-adjusting the LO amplitude. Gain mismatch within each antenna array (between four patches in the prototype) results in AM-AM non-linearity. This can be corrected by digital pre-distortion through a LUT. Note that since I/Q arrays are de-coupled, the pre-distortion is only one dimensional. Since the antenna elements are unlikely to see identical environment, gain mismatch can be angle dependent. As a result, both I/Q amplitude imbalance and DAC AM-AM non-linearity can be angle dependent. To address this issue, both calibration and 1-D digital pre-distortion can be performed in different angles. Beam steering is usually sectorized in a typical phased array, and a LUT may be loaded with different values for different sectors.

The antenna mismatch can be also frequency dependent. The frequency dependent gain mismatch between I and Q antennas results in non-flat channel frequency response, and is no different from frequency dependent gain mismatch between I and Q circuitry on the chip. Such mismatch can be either compensated by a pre-emphasis filter [48] or lumped into the wireless channel response and corrected by the receiver equalizer.

The correction for frequency dependent gain mismatch within each antenna array is more subtle. Note that three different schemes of digital-to-RF conversion implementation in a beamforming array are proposed. For Digital PA only scheme, different antenna elements inside the same array is transmitting the same signal (either I or Q modulated signal), with only a relative delay, therefore frequency dependent gain mismatch only results in non-flat channel frequency response. This is the same problem that every phased array system faces, and it can be corrected by pre-emphasis or receiver equalizer. For the Digital Antenna only scheme and the mixed Digital PA/Antenna scheme, frequency dependent gain mismatch results in both non-flat channel frequency response and non-linearity. In this case, a MMSE algorithm may be used to minimize the EVM for all the possible symbols, which is a potentially more involving task.

Another important design choice is regarding where the digital pre-distortion should be implemented. In this implementation, digital pre-distortion was only applied to Nyquist rate samples preceding the mixed-signal filtering unit. As a result, linearization is only effective at Nyquist rate samples [5]. This is effective in improving EVM, but much less effective in reducing spectral regrowth and out-of-band emissions, especially when the non-linearity of the digital AM - RF AM transfer function is severe. A better approach is to place the digital pre-distortion after filtering. This ensures that the oversampled digital codewords are completely linear, and both the EVM and spectral regrowth can be minimized. The price of such an approach is increased power consumption and slightly reduced DAC resolution.

[5] Assuming the transmitter has sufficient bandwidth to preserve the Nyquist rate samples

Chapter 6

Conclusions

This dissertation focuses on the architecture and the implementation of efficient mm-wave wireless transmitters in standard CMOS technologies for both regional and personal area networks. Low breakdown voltage and scarce power gain of the CMOS process have been identified as the main contributors to inferior transmitter performance at mm-wave.

To overcome the low supply voltage imposed by the breakdown voltage, transformer power combining techniques are proposed and analyzed. A 65nm 60GHz PA utilizing a quad-input DAT was designed and fabricated. Measurement results prove that by doubling the number of input ports in the combiner, the PA output power can be quadruped without sacrificing the efficiency. Given a particular CMOS process, maximum obtainable output power based on such type of combiners can be predicted. To further extend the range of a mm-wave wireless link, beamforming using spatial power combining can be utilized. Four types of beamforming radio architectures have been analyzed and compared. Analog baseband beamforming was chosen as an optimal architecture for an 4-element transceiver array design. The 65nm prototype achieves low power consumption.

Besides obtaining higher output power and peak efficiency, improving average efficiency is equally important as most modulation schemes used today present non-constant envelope signals. A key concept of digital quadrature spatial combing was introduced as an enabler for highly efficient mm-wave transmitters. The direct digital-to-RF conversion architecture can improve the transmitter peak efficiency by using non-linear class PAs and enhance the back-off efficiency by dynamically scaling the total DC power consumption according to the signal envelope. At the same time, proposed quadrature spatial combining effectively eliminates the tradeoff between low insertion loss and high isolation associated with on-chip signal combiners while does not create average spatial directivity. A 60GHz 4I+4Q beamforming transmitter was designed, packaged and characterized. The transmitter delivers a peak EIRP of 22dBm at a power consumption of 382mW. At 6dB back-off when transmitting modulated signals (QPSK/16QAM), the PA achieves an average efficiency of 16.5% with a total transmitter average efficiency of 7%. A mixed-signal filtering approach consisting of

a 4x oversampling FIR and two-fold linear interpolation was used to suppress the spectral images. The measured average PA efficiency at 6-dB back-off is 1.7× better than the state-of-the-art efficiency enhancing mm-wave transmitters.

Bibliography

[1] ERICSSON, "Traffic and market report," 2012. [Online]. Available: http://www.ericsson.com/res/docs/2012/traffic_and_market_report_june_2012.pdf

[2] ITRS, "Itrs reports," 2012. [Online]. Available: http://www.itrs.net/reports.html

[3] Y. Palaskas, A. Ravi, S. Pellerano, B. Carlton, M. Elmala, R. Bishop, G. Banerjee, R. Nicholls, S. Ling, N. Dinur, S. Taylor, and K. Soumyanath, "A 5-ghz 108-mb/s 2 2 mimo transceiver rfic with fully integrated 20.5-dbm power amplifiers in 90-nm cmos," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 12, pp. 2746–2756, Dec. 2006.

[4] G. D. Ewing, "High-efficiency radio-frequency power amplifiers," Ph.D. dissertation, Oregon State University.

[5] N. Sokal and A. Sokal, "Class e-a new class of high-efficiency tuned single-ended switching power amplifiers," *Solid-State Circuits, IEEE Journal of*, vol. 10, no. 3, pp. 168–176, Jun 1975.

[6] M. Acar, A. Annema, and B. Nauta, "Analytical design equations for class-e power amplifiers," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, Dec. 2007.

[7] F. Raab, "Idealized operation of the class e tuned power amplifier," *Circuits and Systems, IEEE Transactions on*, Dec. 1977.

[8] S. Kee, "The class e/f family of harmonic-tuned switching power amplifiers," Ph.D. dissertation, California Institute of Technology.

[9] I. Aoki, S. Kee, D. Rutledge, and A. Hajimiri, "Distributed active transformer-a new power-combining and impedance-transformation technique," *Microwave Theory and Techniques, IEEE Transactions on*, Jan. 2002.

[10] M. Tanomura, Y. Hamada, S. Kishimoto, M. Ito, N. Orihashi, K. Maruhashi, and H. Shimawaki, "Tx and rx front-ends for 60ghz band in 90nm standard bulk cmos," in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, Feb. 2008.

[11] M. Bohsali and A. Niknejad, "Current combining 60ghz cmos power amplifiers," in *Radio Frequency Integrated Circuits Symposium, 2009. RFIC 2009. IEEE*, June 2009.

[12] M. Y. Bohsali, "Millimeter - wave cmos power amplifiers design," Ph.D. dissertation, University of California, Berkeley.

[13] C. Law and A.-V. Pham, "A high-gain 60ghz power amplifier with 20dbm output power in 90nm cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, Feb. 2010, pp. 426–427.

[14] D. Chowdhury, "Efficient transmitters for wireless communications in nanoscale cmos technology," Ph.D. dissertation, University of California, Berkeley.

[15] D. Chowdhury, P. Reynaert, and A. Niknejad, "A 60ghz 1v + 12.3dbm transformer-coupled wideband pa in 90nm cmos," in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, Feb. 2008, pp. 560–635.

[16] C. Marcu, D. Chowdhury, C. Thakkar, J.-D. Park, L.-K. Kong, M. Tabesh, Y. Wang, B. Afshar, A. Gupta, A. Arbabian, S. Gambini, R. Zamani, E. Alon, and A. Niknejad, "A 90 nm cmos low-power 60 ghz transceiver with integrated baseband circuitry," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 12, pp. 3434–3447, Dec. 2009.

[17] T. LaRocca and M.-C. Chang, "60ghz cmos differential and transformer-coupled power amplifier for compact design," in *Radio Frequency Integrated Circuits Symposium, 2008. RFIC 2008. IEEE*, 17 2008-April 17, pp. 65–68.

[18] I. Aoki, S. Kee, D. Rutledge, and A. Hajimiri, "Fully integrated cmos power amplifier design using the distributed active-transformer architecture," *Solid-State Circuits, IEEE Journal of*, vol. 37, no. 3, pp. 371–383, Mar. 2002.

[19] C. A. Balanis, *ANTENNA THEORY ANALYSIS AND DESIGN*. John Wiley & Sons, 2005.

[20] A. Natarajan, B. Floyd, and A. Hajimiri, "A bidirectional rf-combining 60ghz phased-array front-end," in *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, 2007, pp. 202–597.

[21] A. Natarajan, S. Reynolds, M.-D. Tsai, S. Nicolson, J.-H. Zhan, D. G. Kam, D. Liu, Y.-L. Huang, A. Valdes-Garcia, and B. Floyd, "A fully-integrated 16-element phased-array receiver in sige bicmos for 60-ghz communications," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 5, pp. 1059–1075, 2011.

[22] A. Valdes-Garcia, S. Nicolson, J.-W. Lai, A. Natarajan, P.-Y. Chen, S. Reynolds, J.-H. Zhan, D. Kam, D. Liu, and B. Floyd, "A fully integrated 16-element phased-array transmitter in sige bicmos for 60-ghz communications," *Solid-State Circuits, IEEE Journal of*, vol. 45, no. 12, pp. 2757–2773, 2010.

[23] K. Raczkowski, W. De Raedt, B. Nauwelaers, and P. Wambacq, "A wideband beamformer for a phased-array 60ghz receiver in 40nm digital cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, 2010, pp. 40–41.

[24] H. Wang and A. Hajimiri, "A wideband cmos linear digital phase rotator," in *Custom Integrated Circuits Conference, 2007. CICC '07. IEEE*, 2007, pp. 671–674.

[25] K. Doris, E. Janssen, C. Nani, A. Zanikopoulos, and G. Van Der Weide, "A 480 mw 2.6 gs/s 10b time-interleaved adc with 48.5 db sndr up to nyquist in 65 nm cmos," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 12, pp. 2821–2833, 2011.

[26] D. Stepanovic and B. Nikolic, "A 2.8gs/s 44.6mw time-interleaved adc achieving 50.9db sndr and 3db effective resolution bandwidth of 1.5ghz in 65nm cmos," in *VLSI Circuits (VLSIC), 2012 Symposium on*, 2012, pp. 84–85.

[27] E. A. Firouzjaei, "mm-wave phase shifters and switches," Ph.D. dissertation, University of California, Berkeley.

[28] M. Tabesh, A. Arbabian, and A. Niknejad, "60ghz low-loss compact phase shifters using a transformer-based hybrid in 65nm cmos," in *Custom Integrated Circuits Conference (CICC), 2011 IEEE*, 2011, pp. 1–4.

[29] L. Kong, "Energy-efficient 60ghz phased-array design for multi-gb/s communication systems," Ph.D. dissertation, University of California, Berkeley.

[30] M. Tabesh, J. Chen, C. Marcu, L. Kong, S. Kang, A. Niknejad, and E. Alon, "A 65 nm cmos 4-element sub-34 mw/element 60 ghz phased-array transceiver," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 12, pp. 3018–3032, 2011.

[31] J. Chen and A. Niknejad, "A compact 1v 18.6dbm 60ghz power amplifier in 65nm cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, 2011, pp. 432–433.

[32] J.-W. Lai and A. Valdes-Garcia, "A 1v 17.9dbm 60ghz power amplifier in standard 65nm cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, 2010, pp. 424–425.

[33] F. Wang, A. Yang, D. Kimball, L. Larson, and P. Asbeck, "Design of wide-bandwidth envelope-tracking power amplifiers for ofdm applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, no. 4, pp. 1244–1255, 2005.

[34] D. Cox, "Linear amplification with nonlinear components," *Communications, IEEE Transactions on*, vol. 22, no. 12, pp. 1942–1945, 1974.

[35] H. Chireix, "High power outphasing modulation," *Radio Engineers, Proceedings of the Institute of*, vol. 23, no. 11, pp. 1370–1392, 1935.

[36] S. Moloudi and A. Abidi, "The outphasing rf power amplifier: A comprehensive analysis and a class-b cmos realization," *Solid-State Circuits, IEEE Journal of*, vol. 48, no. 6, pp. 1357–1369, 2013.

[37] C. Liang and B. Razavi, "Transmitter linearization by beamforming," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 9, pp. 1956–1969, 2011.

[38] D. Zhao, S. Kulkarni, and P. Reynaert, "A 60-ghz outphasing transmitter in 40-nm cmos," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 12, pp. 3172–3183, 2012.

[39] A. Kavousian, D. Su, M. Hekmat, A. Shirvani, and B. Wooley, "A digitally modulated polar cmos power amplifier with a 20-mhz channel bandwidth," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 10, pp. 2251–2258, 2008.

[40] D. Chowdhury, L. Ye, E. Alon, and A. Niknejad, "An efficient mixed-signal 2.4-ghz polar power amplifier in 65-nm cmos technology," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 8, pp. 1796–1809, 2011.

[41] L. Ye, "Design and analysis of digitally modulated transmitters for efficiency enhancement," Ph.D. dissertation, University of California, Berkeley.

[42] C. Lu, H. Wang, C. Peng, A. Goel, S. Son, P. Liang, A. Niknejad, H. Hwang, and G. Chien, "A 24.7dbm all-digital rf transmitter for multimode broadband applications in 40nm cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, 2013, pp. 332–333.

[43] A. Niknejad, D. Chowdhury, J. Chen, J. Park, and L. Ye, "mm-wave quadrature spatial power combinig: A proposal," July 2010. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-110.pdf

[44] J. Chen, L. Ye, D. Titz, F. Gianesello, R. Pilard, A. Cathelin, F. Ferrero, C. Luxey, and A. M. Niknejad, "A digitally modulated mm-wave cartesian beamforming transmitter with quadrature spatial combining," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, 2013, pp. 232–233.

[45] A. Babakhani, D. Rutledge, and A. Hajimiri, "Transmitter architectures based on near-field direct antenna modulation," *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 12, pp. 2674–2692, 2008.

[46] S. Emami, R. Wiser, E. Ali, M. Forbes, M. Gordon, X. Guan, S. Lo, P. McElwee, J. Parker, J. Tani, J. Gilbert, and C. Doan, "A 60ghz cmos phased-array transceiver pair for multi-gb/s wireless communications," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, 2011, pp. 164–166.

[47] V. Vidojkovic, G. Mangraviti, K. Khalaf, V. Szortyka, K. Vaesen, W. Van Thillo, B. Parvais, M. Libois, S. Thijs, J. Long, C. Soens, and P. Wambacq, "A low-power 57-to-66ghz transceiver in 40nm lp cmos with -17db evm at 7gb/s," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*, 2012, pp. 268–270.

[48] T. Tsukizawa, N. Shirakata, T. Morita, K. Tanaka, J. Sato, Y. Morishita, M. Kanemaru, R. Kitamura, T. Shima, T. Nakatani, K. Miyanaga, T. Urushihara, H. Yoshikawa, T. Sakamoto, H. Motozuka, Y. Shirakawa, N. Yosoku, A. Yamamoto, R. Shiozaki, and N. Saito, "A fully integrated 60ghz cmos transceiver chipset based on wigig/ieee802.11ad with built-in self calibration for mobile applications," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, 2013, pp. 230–231.