Digital Radio Encoding and Power Amplifier Design for Multimode and Multiband Wireless Communications

by

Jingjing Xia

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Jingjing Xia

Abstract

The evolution of wireless technology has necessitated the support of multiple communication standards by mobile devices. At present, multiple chipsets/radios operating at predefined sets of modulation schemes, frequency bands, bandwidths and output power levels are used to achieve this objective. This leads to higher component counts, increased cost and limits the capacity to cope with future communication standards. In order to tackle different wireless standards using a single chipset, digital circuits have been increasingly deployed in radios and demonstrated re-configurability in different modulation schemes (multimode) and frequency bands (multiband).

Despite efforts and progress made in digitizing the entire radio, the power amplifier (PA) is still designed using an conventional approach and has become the bottleneck in digital transmitters, in terms of low average power efficiency, poor compatibility with modern CMOS technology and limited re-configurability.

This research addresses these issues from two aspects. The first half of the thesis investigates signal encoding issues between the modulator and PA. We propose, analyze and evaluate a new hybrid amplitude/time signal encoding scheme that significantly improves the coding efficiency and dynamic range of a digitally modulated power amplifier (DMPA) without significantly increasing design complexity. The proposed hybrid amplitude/time encoding scheme combines both the amplitude domain and the time domain to optimally encode information. Experimental results show that hybrid amplitude/time encoding results in a 35% increase in the average coding efficiency with respect to conventional time encoding, and is only 6.7% lower than peak efficiency when applied to a Wireless Local Area Network (WLAN) signal with a peak to average power ratio equal to 9.9 dB. A new DMPA architecture, based on the proposed hybrid encoding, is also proposed.

The second half of this thesis presents the design, analysis and implementation of a CMOS PA that is amenable to the proposed hybrid encoding scheme. A multi-way current mode class-D PA architecture has been proposed and realized in 130 nm CMOS technology. The designed PA has satisfied the objectives of wide bandwidth (1.5 GHz - 2.7 GHz at 1 dB output power), and high efficiency (PAE 63%) in addition to demonstrating linear responses using the proposed digital encoding. A complete digital transmitter combining the encoder and the multi-way PA was also investigated. The overall efficiency is 27% modulating 7.3 dB peak to average power ratio QAM signals.

Acknowledgements

It is beyond words to convey my deepest gratitude to my supervisors, Professor Slim Boumaiza and Professor Siddharth Garg, for their support, inspiration and encouragement throughout my research. Without them, this work would not have been possible.

I also like to thank my colleagues and friends from Emerging Radio System Group (EMRG), particularly, Hassan Sarbishaei, Daniel Frebrowski and Dr. Foad Arfaei Malekzadeh. Thank you for sharing your experiences, thoughts, ideas and giving me those valuable feedbacks.

I also like to thank the generous financial support from Graduate Studentship and the department of ECE. In addition, I wish to thank Agilent Technologies and CMC for the software and fabrication support.

Dedication

This thesis is dedicated to my parents and my wife.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the last two decades, the growing demand for wireless connectivity between different parties, devices, applications and regions has led to a proliferation of wireless standards. Take cellular network as an example. As the biggest powerhouse behind the development of wireless technology, users' requirements have grown from telephony in the early days to data and on-demand services nowadays. These requirements have fueled the evolution of cellular networks from 2nd generation GSM, to 3G WCDMA and, most recently, 4G LTE. However, the introduction of a new standard does not mean the obsoleteness of previous generation standards. On the contrary, old standards still co-exist with new technologies and must be supported in the devices for compatibility and global mobility.

At present, mobile devices such as smart phones or tablets are configured as one centralized processor connected to multiple radio frequency (RF) front-end chipsets. These chipsets operate at predefined sets of frequency bands, bandwidths, output power levels and usually only support one particular wireless standard. As RF chipsets are, in general, the most expensive and power consuming components, this multi-chip configuration leads to higher component counts, increased cost and limits the capacity of the system to cope with future communication standards.

Meanwhile, the demand for high data-rate communications using spectrally efficient modulation leads to signals with high peak-to-average-power-ratio (PAPR) and stringent linearity requirements. For example, the new 802.11ac protocol supports over a 1 Gbps data rate wireless link. The PAPR can be as high as 12 dB and the maximum constellation error is -32 dB in 256-QAM. Maintaining high signal quality and satisfying the tight spectrum emission mask become challenges to the transmitter design. Linear power amplifiers (PAs) operating at back-off can be employed in order to achieve the required linearity. However, a linear PA results in low power efficiency and high energy dissipation, which in turn limits the portability of communication devices [2, 10].

Supporting multiple communication standards within one chipset, while maintaining power efficiency and linearity, has become a critical challenge in transmitter design. Motivated by this challenge, researchers have moved toward software defined radio (SDR) transmitter platforms that are reconfigurable, wideband and energy efficient.

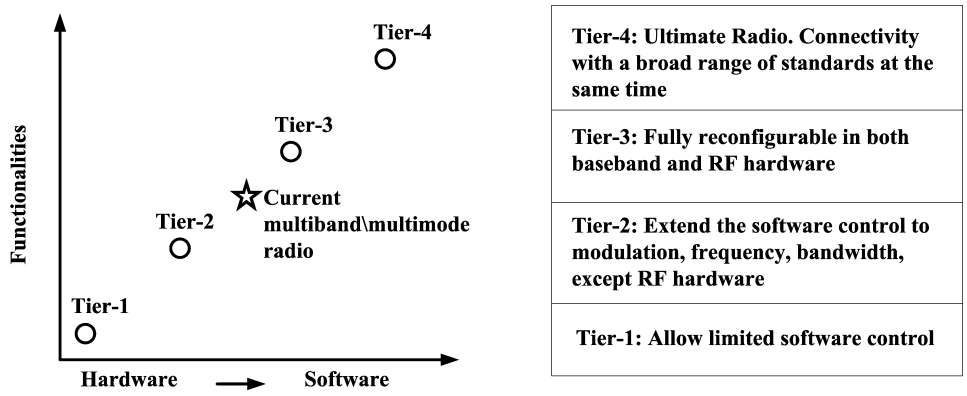

Figure 1.1: Four tiers evolution in SDR

Starting as a concept in the 1990s, SDR is envisioned as an ultimate radio that is fully reconfigurable in software. A SDR radio transceiver directly samples the receiver path through an analog-to-digital (ADC) converter, and transmits directly from a digital-to-analog (DAC) converter [11]. This idealistic vision of SDR imposes unrealistic requirements on the resolution, bandwidth and dynamic range of ADCs/DACs even when operating only at the Nyquist rate. In the last two decades, research activities have been very active exploring practical SDRs and we have witnessed significant advances on the road toward a fully SDR. Based on prior and ongoing activities, SDR has been categorized into four tiers by the SDR forum, as described in Figure 1.1. Tier-2 requires re-configurability in the baseband but allows multiple RF modules for multiband operations. Tier-3 extends the re-programmability to the hardware.

State-of-the-art multimode/multiband radios are between Tier 2 and 3 [11], most of which have achieved re-configurability in the baseband but are striving to extend re-configurability to the analog/RF domain using digital circuits and digital signal processing (DSP). In particular, the greatest hurdle in the digitalization of radio chipsets is the PA, which is required to be extremely linear, efficient and robust at large voltage swing. At this moment, multiple narrow-band PAs are integrated on a single IC die in order to achieve multiband operation. A more area- and cost-efficient approach is to employ a broadband PA (as authors did in [12]) which covers the frequency bands of popular wireless standards. Despite the support for multiband PAs, these approaches are analog intensive, non-reconfigurable and, therefore, less amenable to the digital re-programmability feature of SDR. Some recent work investigated the digital PA concept [11] by exploiting switch mode power amplifiers (SMPAs) combined with digital encoding (*e.g.* pulse-width modulation or delta sigma modulation). These architectures are capable of realizing a fully digital transmitter by extending the re-programmability to the PA stage. However, the practicality of a digital PA is limited by its low efficiency, large quantization noise and frequency spurs.

This work will introduce the author's researches in relation to PA-included multiband/multimode digital transmitters. The thesis explores signal encoding

techniques and digital transmitter architectures that could include SMPA as an integral part of the transmitter in order to achieve better overall efficiency and signal qualities. The organization of this thesis is as follows:

Chapter 2 reviews state-of-the-art digital transmitter architectures. These architectures can be categorized into cartesian- and polar-based at a high level, and each kind can be further divided into smaller groups. We will discuss strengths and limitations of each approach.

In Chapter 3, we intend to address limitations of existing time-encoding techniques by proposing a hybrid approach. Analysis and measurement results are provided to support the theory and demonstrate the advantages of this new scheme.

Chapter 4 presents the design and analysis of an inverse class-D digital PA that supports the proposed encoding scheme. The circuit has been implemented in a mixed signal $130nm$ CMOS process.

Finally, Chapter 5 concludes this thesis with some remarks on future work.

Chapter 2

Background: Digital Transmitters

This chapter provides an overview of key technologies in multimode and multiband radios. With the increasing processing speed of CMOS technology, digital circuits has begun taking over tasks that were traditionally carried out in the analog domain. Therefore we have narrowed down the scope of reviews to digital-intensive or fully digital designs. These designs have demonstrated promising results in reconfigurability, energy efficiency and practicality. In the following section, we refer to them generally as digital transmitters.

2.1 Cartesian-based Digital Transmitter Architecture

In the conventional analog transmitters, a baseband signal can be represented either in cartesian coordinates as

$$S(t) = I(t) + jQ(t) \tag{2.1}$$

or in polar coordinates as

$$S(t) = a(t)\angle\theta(t) \tag{2.2}$$

The former results in a quadrature modulated transmitter and the latter is known as a polar transmitter. Similar concepts have migrated to digital transmitters, yielding designs based on both coordinates. This section will discuss cartesian-based digital transmitters.

2.1.1 Direct Digital RF Modulator

In a conventional quadrature modulated transmitter, the baseband signals need to go through DAC, intermediate frequency (IF) filter, mixer before reaching the PA. As shown in Figure 2.1, such a transmitter has significant analog components, which are in general more power hungry, susceptible to noise, not reconfigurable and occupy more area [13].

The concept of a direct digital RF modulator (DDRM) has been proposed in [14]. The core of DDRM comprises two digital-to-RF converters (DRFC, *a.k.a.*

RF-DAC). Compared to an analog I/Q modulator, Figure 2.1 indicates that using the RF-DAC has significantly reduced the number of analog components and moved the boundary between the digital and analog domain closer to the antenna. Moreover, the LO filter, when implemented as a FIR filter, is versatile and can be reconfigurable for multimode operations.

To better understand the difference between the two architectures, simplified circuit schematics are provided in Figure 2.2. In the analog transmitter, digital samples are converted to analog waveforms through a current steering DAC. Output of the DAC, usually constructed as non-return-to-zero waveforms, has aliased images centring around the DAC's sampling frequency. An IF filter removes the images before the up-conversion.

The idea of RF-DAC originates from a current steering DAC. Since both the DAC and mixing take place in the form of current, rather than voltage, sequence of this three-step process (*i.e.* DAC, filtering and mixing) is altered in an RF-DAC. The DAC and mixing can be realized in the same circuit, rather than separately, as shown in Figure 2.2. By eliminating the voltage-to-current conversion stage in a traditional mixer, an RF-DAC operates purely by steering different current sources, hence removing one major source of non-linearity in the traditional transmitter. Moreover, energy and chip area are saved by reusing the current.

Note that in Figure 2.2, output of the RF-DAC before the reconstruction filter is sampling and holding the envelope of the RF carrier. This creates aliased images at $f_c \pm k f_{sampling}$ and imposes challenging requirements in the reconstruction filter. Take an IEEE 802.11ac 20 MHz baseband signal as an example. Assuming an over-sampling-ratio (OSR) of 4, the nearest image is only 80 MHz away from the carrier frequency. When operating at 2.4 GHz, the bandpass filter needs to achieve fractional bandwidth of as small as 6%. Either a higher OSR or filter order is required for practical applications.

2.1.2 Delta-Sigma DDRM

For wireless standards having a high dynamic range (*e.g.* WCDMA) a large number of current sources are required in the RF-DAC to achieve the amplitude bit resolution. As the number of bits increases, impairments due to mismatches are unavoidable. In an effort to improve the signal-to-noise ratio (SNR), delta-sigma modulator (DSM) has been introduced in DDRM [15]. Figure 2.3 describes the block diagram of a DSM DDRM. The use of noise shaping before the RF-DAC could improve the in-band SNR, at the expense of out-of-band noise. For example, a 1-bit 4th-order DSM could achieve a high SNR. However, the adjacent noise is too steep to be filtered without violating the spectral masks. In [15], a 3-bit quantizer is implemented in the DSM and only requires moderate filtering for 160-MHz OFDM signals at 5.2 GHz.

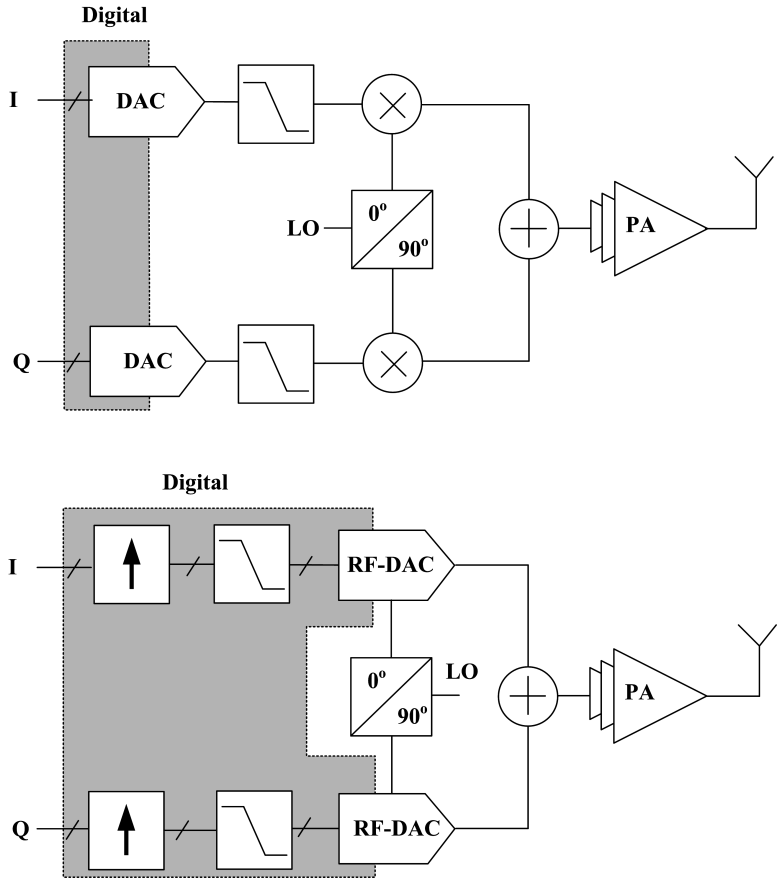

Figure 2.1: Comparison between a conventional direct conversion transmitter and a direct digital RF modulator.

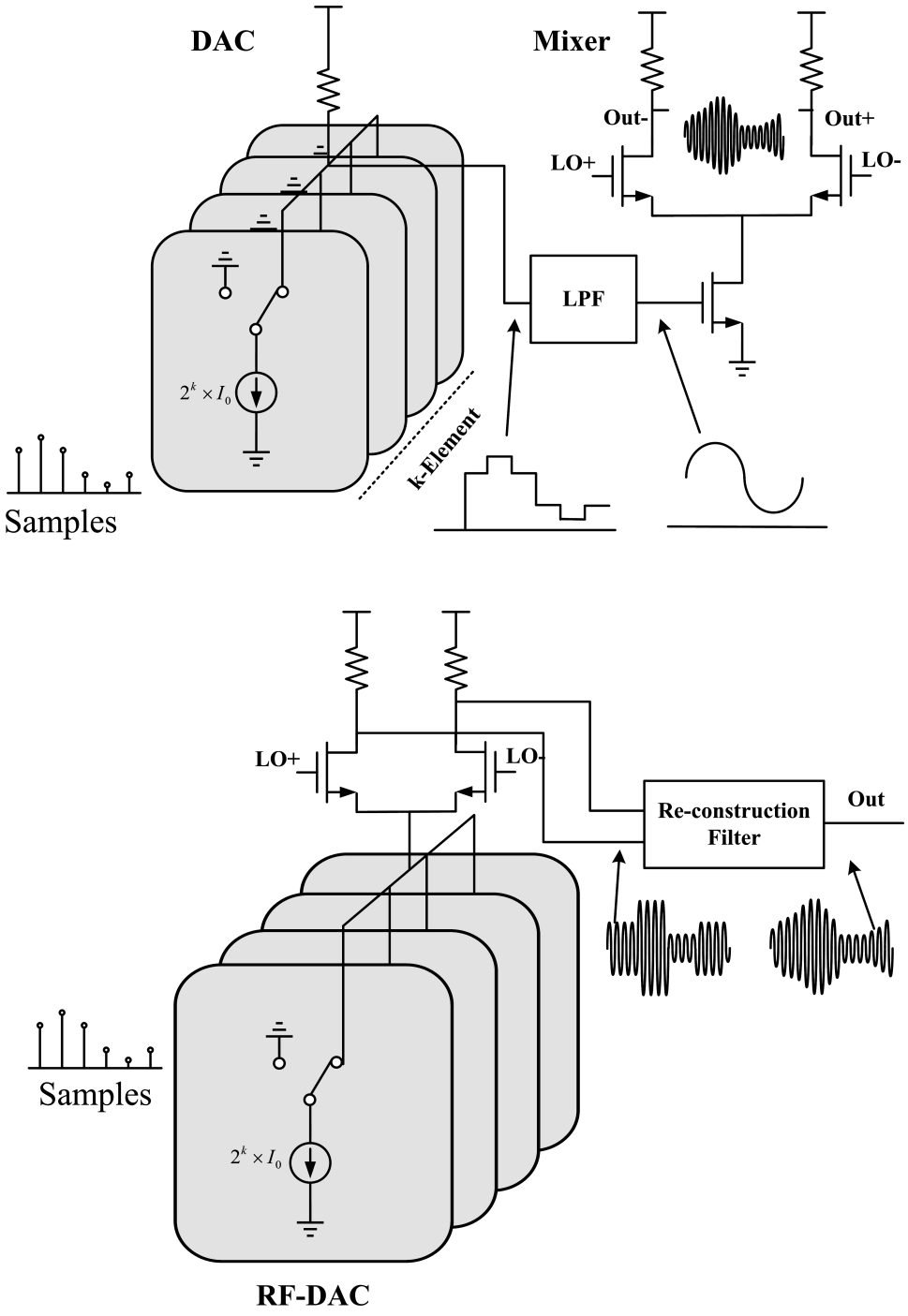

Figure 2.2: Comparison between an analog current steering DAC and a RF-DAC.

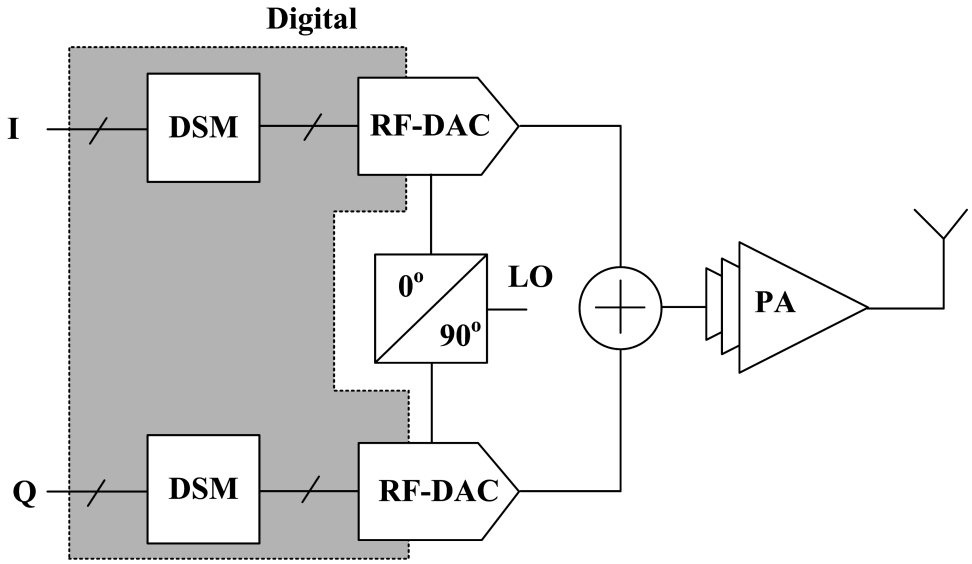

Figure 2.3: A DSM direct digital RF modulator

2.1.3 Fully Digital I/Q Modulator

Revisiting Figure 2.2, it is observed that the RF-DAC stage is still analog-intensive: the LO signal and RF-DAC's output are still in continuous wave. There are two significant drawbacks to using analog LO: 1. the output is more susceptible to noise and I/Q imbalance, especially in a short channel technology node with only hundreds of millivolts voltage headroom; 2. analog processing such as multiplication is power-hungry compared to digital multiplication.

In any I/Q transmitter, the complex $I(t) + jQ(t)$ signals must be summed orthogonally for distortionless modulation and demodulation

$$s(t) = I(t)c_I(\omega_c t) + Q(t)c_Q(\omega_c t) \tag{2.3}$$

where $c_I(\omega_c t)$ and $c_Q(\omega_c t)$ are the LO modulating signals satisfying

$$\int_t^{t+T} c_I(\omega_c t) \times c_Q(\omega_c t) = 0 \tag{2.4}$$

Simply using two 50% duty cycle digital clocks with $T/4$ time delay is not orthogonal and will cause distortion after signal modulation. A recent work proposed a walk-around approach and demonstrated a fully digital I/Q transmitter [1].

Figure 2.4 shows the block diagram of the proposed fully digital modulator. The basic principle is to use two 25% duty cycle pulses with $T/4$ time delay. Denoting the LO clock in digital samples as $c_I[n]$ and $c_Q[n]$ and assuming the clock is running at $4X$ the carrier frequency, the modulating clocks are

$$c_I[n] = \{1, 0, 0, 0\} \tag{2.5}$$

$$c_Q[n] = \{0, 1, 0, 0\} \tag{2.6}$$

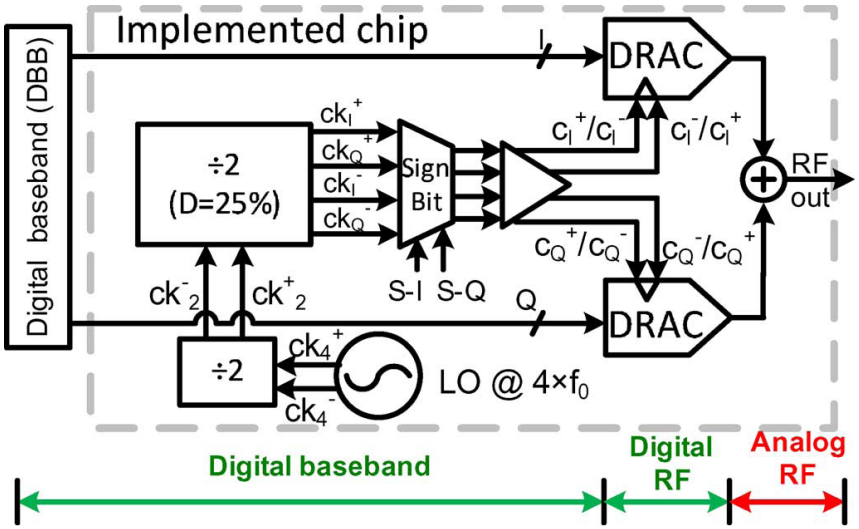

Figure 2.4: A fully digital I/Q transmitter [1]

Generating LO vectors in (2.5) and (2.6) requires a high speed clock, since the 25% duty cycle pulse is derived from dividing down a clock running at $4f_c$. Another fully digital I/Q modulator in [16] avoids the $4f_c$ clock using the following LO vectors

$$c_I[n] = \{1, 1, -1, -1\} \tag{2.7}$$

$$c_Q[n] = \{-1, 1, 1, -1\} \tag{2.8}$$

Their proposed digital I/Q modulator has achieved 20 dB margin in meeting the 802.16e WiMAX spectral masks and -36 dB EVM.

2.2 Polar-based Digital Transmitter Architecture

The Cartesian-based digital transmitter has demonstrated its capability to generate standard-compliant signals for almost any kind of wireless protocol: GSM/EDGE, WCDMA, WLAN, LTE, etc. [16]. However, revisiting Figures 2.1-2.4, it can be observed that the digital circuit is still unable to tap into the PA in these designs and, with few exceptions, they rely on external linear amplifiers. It is well known that conventional linear PAs are inefficient when amplifying wideband signals with a high PAPR. Moreover, conventional PAs are in general narrow-band and non-reconfigurable. As the most power hungry and expensive component in a radio, it is highly desirable to bring high efficiency and re-configurability to the PA for the future SDR.

A switch-mode power amplifier (SMPA) is a highly efficient architecture that could potentially achieve 100% efficiency. Operating the transistors as switches, rather than current sources, it is also amenable to integration with CMOS technology. A Cartesian-based digital transmitter is, by its nature, not compatible with

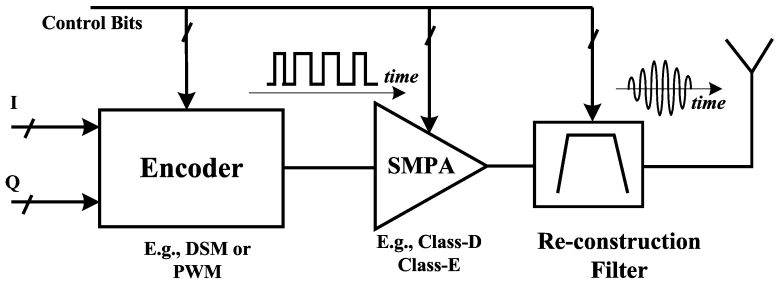

Figure 2.5: Generic diagram of a digitally modulated power amplifier (DMPA)

SMPA, since output of the digital I/Q modulator has varying magnitude and will cause significant distortions when amplified through a SMPA.

On the contrary, an important advantage of a polar-based digital transmitter is the ability to separate the amplitude from the phase and only allow constant envelope signals into the SMPA. More commonly known as a digitally modulated power amplifier (DMPA), this architecture ensures that power efficiency is always at the maximum in the SMPA. Figure 2.5 shows a generic block diagram of DMPA, consisting of an encoder, an array of SMPAs and a reconstruction filter. The encoder generates digital pulse waveforms that embed the desired information in the frequency band of interest. The SMPAs utilize Class-D or Class-E PAs that can achieve 100% efficiency in theory, hence greatly improving the power efficiency of the overall system. Finally, the filter removes the quantization noise and frequency spurs that arise from the digital encoding. This section will present a detailed review of various DMPAs.

2.2.1 Amplitude-encoded DMPA

In an analog polar transmitter (*a.k.a.* envelope elimination and restoration) envelope $a(t)$ and phase $\theta(t)$ travel along two different paths. In the envelope path $a(t)$ controls the supply voltage of PA in order to vary the output power. In the phase path $\theta(t)$ is a phase modulated signal with a constant envelope, hence will always saturate the PA at maximum output power and efficiency. The operation is described in Figure 2.6.(a).

An amplitude-encoded DMPA builds upon this idea. Rather than varying the PA's supply voltage, it divides a large PA into smaller units, each being enabled/disabled by a decoder, as shown in Figure 2.6.(b).

Research in amplitude-encoded DMPA has been very active in recent years [17, 2, 18, 19, 3]. Figure 2.7 shows a DMPA with 21.8 dBm output power, 44% peak efficiency and 800 MHz 1-dB bandwidth from only 1 V supply [3]. Digital filtering has been implemented to remove images from up-sampling. These promising results have further demonstrated that high efficiency, multiband/multimode operations and high levels of function could be realized in digital transmitters.

One drawback of amplitude-encoded DMPA results from the power combining. In order to achieve a high dynamic range, a large number of unit-PAs are required.

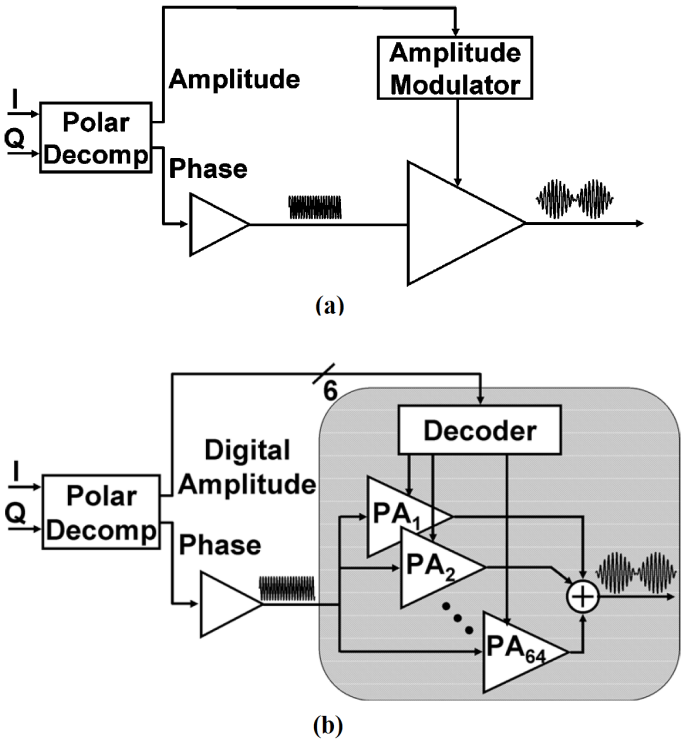

Figure 2.6: Block diagram of a amplitude-encoded DMPA [2]

These unit-PAs are usually binary-weighted, having different output impedance and electrical delays. The current from different unit-PAs have different phase and will reduce the combining efficiency. In addition, the SMPA efficiency drops very quickly at low amplitude levels due to the high on-state resistance of unit-PAs [3].

2.2.2 Envelope Pulse Width Modulation

As previously discussed, amplitude-encoded DMPA employs a 2^n-PA array in the process of digital-to-RF conversion. When the number of bits, n, increases, the required resolution reduces at a ratio of 2^n, creating issues due to mismatch and process temperature variation. Instead of encoding in the amplitude, an alternative approach is to encode the signal envelope, $a(t)$, in time. With the aggressive channel length scaling in CMOS technology, digital circuits are gaining considerably in speed but losing in voltage headroom. Time-encoded operations are therefore promising solutions for future SDR.

A well-known time encoding technique is pulse width modulation (PWM), which can be further categorized into two kinds: envelope (or low-pass) and RF PWM. Figure 2.8 describes the principle of envelope PWM when modulating a complex signal. The envelope, $a(t)$, is being sampled at a rate of f_p . Each sample, $a[n]$, is then converted into a pulse of period $1/f_p$ and width d/f_p, where the duty cycle d

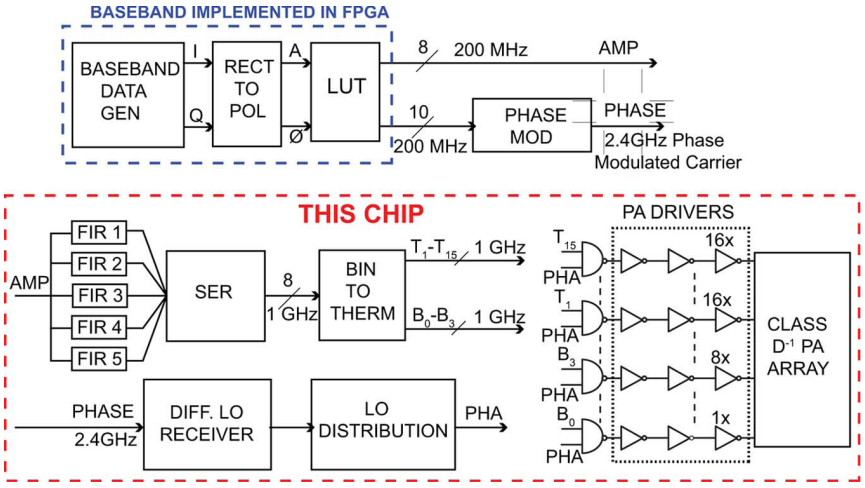

Figure 2.7: A mixed-signal DMPA with digital filters and high efficiency inverse class-D PA array [3]

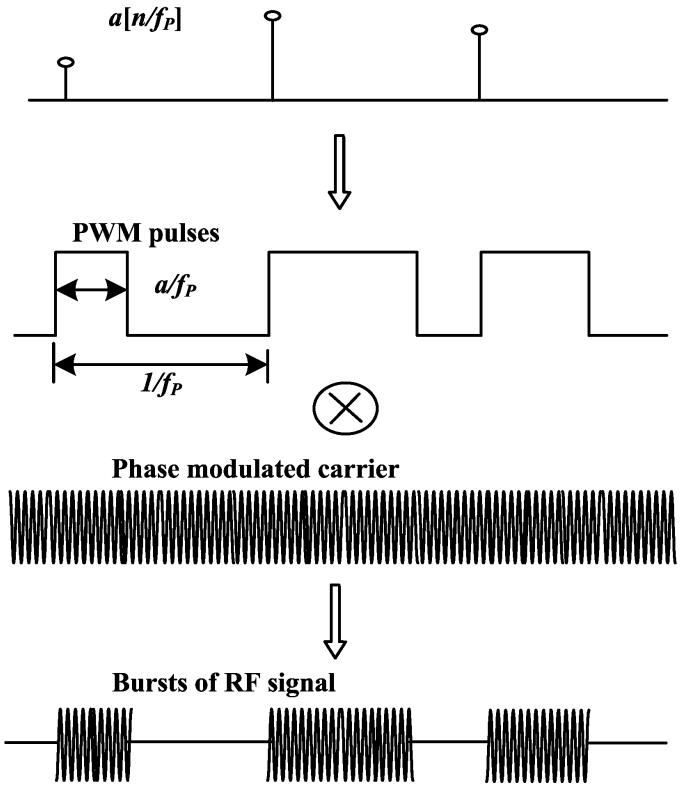

Figure 2.8: Block diagram of an envelope PWM digital transmitter

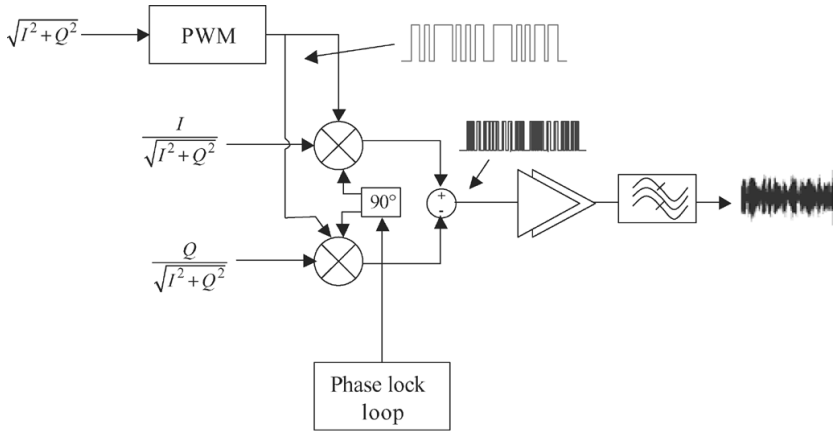

Figure 2.9: A practical implementation of an envelope PWM transmitter [4]

is equal to the normalized amplitude, or

$$d = a[n] \tag{2.9}$$

When the PWM pulse is multipled with the phase modulated RF signal, Figure 2.8 illustrates that bursts of RF signals are generated. The RF burst only has two levels, cut-off or constant envelope. Therefore, burst mode operation ensures that the SMPA will only operate at maximum efficiency. In Figure 2.9, an implementation of envelope PWM digital transmitter in [4] is provided. Note that the desired signals are reconstructed after the bandpass filter.

Other recent works in envelope PWM include [20, 21, 22, 23, 24, 25] and have all achieved very promising result. In particular, [25] implements envelope PWM in the supply modulator of an inverse class-D PA and has demonstrated a total efficiency of 51% with 31 dBm output power at 0.75 GHz. Recently, [26] reported a truly digital envelope PWM transmitter. The PWM is based on an asynchronous digital delay-line approach and has -24 dB EVM when modulating 20 MHz WLAN signals at 2.4 GHz.

Despite its simplicity, envelope PWM still has issues such as low dynamic range and coding efficiency. This will be further elaborated in Chapter 3.

2.2.3 RF Pulse Width Modulation

The concept of RF-PWM was firstly introduced in [27]. Unlike envelope PWM, in which the encoding takes place at burst-level, RF-PWM is encoded at the pulse-level. The basic idea of RF PWM is shown in Figure 2.10. Amplitude, a, of a modulated signal is mapped to the width, or equivalently the duty cycle d, of a pulse with a period of $T = 1/f_c$. A wider pulse indicates larger amplitude and vice versa. Phase ϕ is mapped to the position t_ϕ as $t_p = \frac{\phi}{\pi}T$. Mathematically, a RF-PWM modulated signal can be defined as

$$S_{RF-PWM}(t) = \Pi\left[\frac{d(t-t_\phi)}{T}\right] \tag{2.10}$$

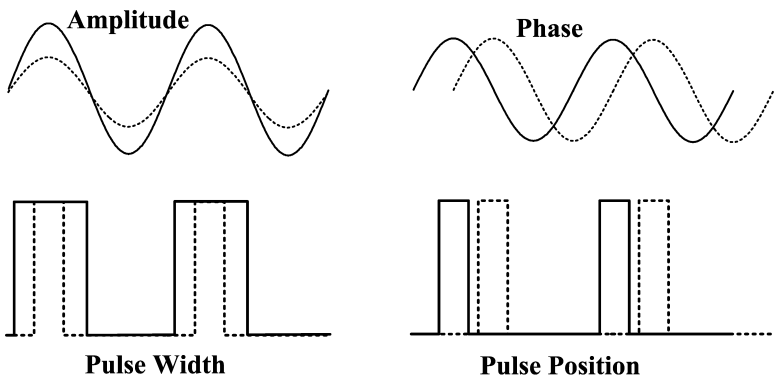

Figure 2.10: Basic principle of RF-PWM

where $\Pi(t)$ is our defined square waveform

$$\Pi(t) = \begin{cases} 1 & t \in [-1, 1] \\ 0 & t \; elsewhere \end{cases} \tag{2.11}$$

In order to better understand RF-PWM, we express $S_{RF-PWM}(t)$ in Fourier series [1] as

$$S_{RF-PWM}(t) \quad = \quad a_0 + \sum_{k=1}^{\infty} Re\left[a_k e^{jk\omega_c(t-t_\phi)}\right] \tag{2.12}$$

where $a_k \forall k \in [1, \infty)$ is the Fourier series coefficients of $\Pi(\frac{dt}{T})$.

A number of observations can be made from (2.12):

1. The RF-PWM signals only comprise signals at DC, fundamental frequency and its harmonics. Unlike envelope PWM, the aliased frequencies of RF-PWM are separated by at least ω_c, hence greatly reducing the filtering requirement.

2. After the bandpass filtering, only the fundamental component is preserved with magnitude of

$$a_1 = \frac{2}{\pi} sin(d\pi) \tag{2.13}$$

Equation (2.13) reveals that RF-PWM is a highly non-linear encoding process. Output of the encoder modulates the input, d, by the $sin()$ function. When applying any predistortion, the duty cycle is limited to the range of $[0, 50\%]$.

3. RF-PWM imposes stringent requirements on PA bandwidth. The pulse width is required to be extremely narrow for achieving reasonable dynamic range. As an example, for $20dB$ dynamic range at $1GHz$, the pulse needs to be $16ps$.

One notable work on a digital transmitter applying RF-PWM is from [5]. The RF-PWM signal generation block diagram is shown in Figure 2.11. Denoting the

[1] The underlining assumption is a periodic signal in order to apply Fourier series. In reality, the baseband signal varies at a much lower speed than the clock rate, hence Fourier series anaylsis could still be a good approximation.

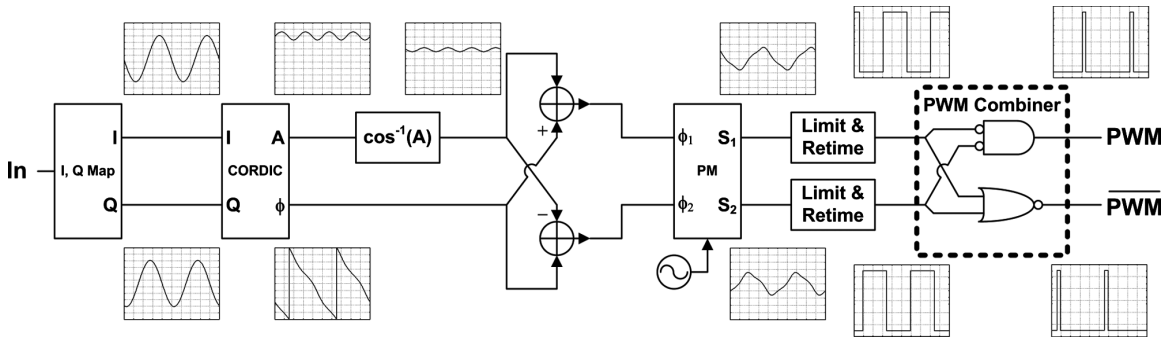

Figure 2.11: RF-PWM signal generation through outphasing technique [5]

baseband signal $I+jQ$ signal in polar form, $A\angle\phi$, the basic ideal here is to construct $A\angle\phi$ through outphasing as

$$A\angle\phi = S_1\angle(\phi+\theta) + S_1\angle(\phi-\theta) \tag{2.14}$$

where S_1 has constant magnitude and the outphasing angle $\theta = cos^{-1}(A)$. This digital outphasing technique is able to generate arbitary pulse width and pulse phase, by varying the phases of two 50% duty cycle clocks.

Recently, a fully digital WLAN transmitter in 32nm CMOS was announced by Intel [28]. The design used similar digital outphasing technique and reported an average efficiency of 22% at 20 dBm average power in 802.11g signals.

2.2.4 Delta-Sigma Modulation

As mentioned earlier in section 2.1.2, in addition to PWM, another widely used time-encoding technique is DSM. Figure 2.12 shows the block diagram of a first-order DSM. The basic idea here is that the sampling speed is much faster than the input speed, a process known as over-sampling. In each sample, the quantization noise of a previous sample is stored in the accumulator and subtracted from the quantization noise of the current sample. This process partially removes the quantization noise, at least in-band, and improves the SNR.

Similar to PWM, DSM can be implemented either as envelope-DSM or RF-DSM. Envelope-DSM is constructed similarly to Figure 2.9, except by replacing the PWM with DSM in the modulator. Figure 2.13 shows the block diagram of an RF-DSM digital transmitter [7]. Its operations are summarized as follows. The baseband I/Q signal is up-converted in the DSP digitally. Given a carrier frequency f_c of 800 MHz in this design, the clock needs to run four times the carrier frequency, or 3.2 GHz, for optimal SNR in a RF-DSM [7]. After digital up-converter, the samples are then processed in the DSM before passing to the SMPA.

In general, DSM produces better signal quality than PWM because the aliasing in PWM could increase the in-band noise floor [29]. The main disadvantage of a DSM-based digital transmitter is the coding efficiency. In order to generate a binary-level signal, significant amounts of out-of-band energy exist as quantization noise. The noise is not only difficult to filter, but more importantly, will reduce the

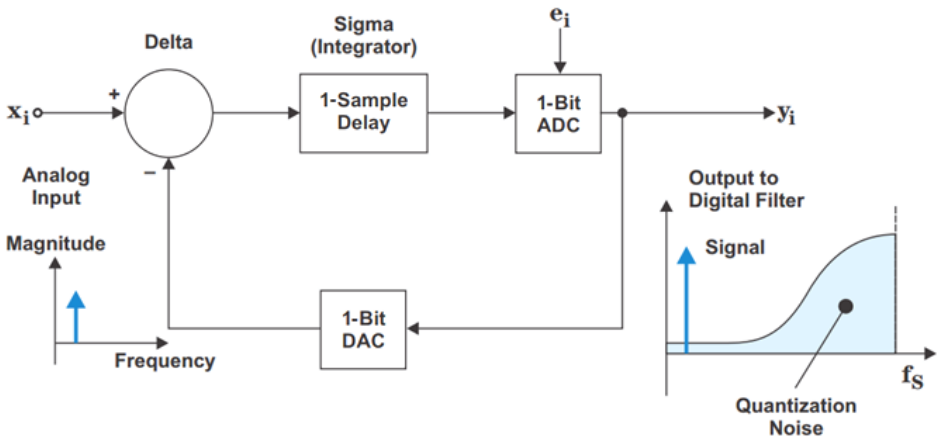

Figure 2.12: Block diagram of a first-order DSM [6]

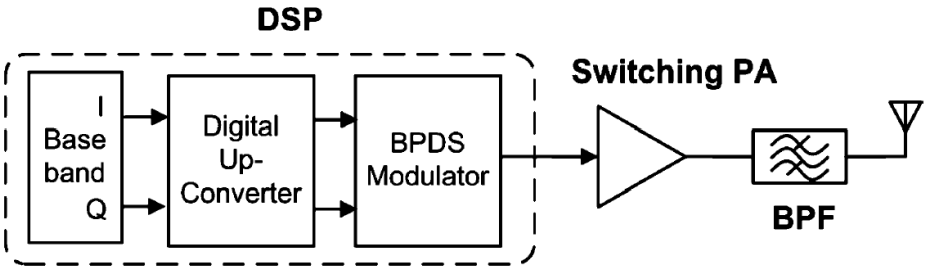

Figure 2.13: An RF-DSM digital transmitter [7]

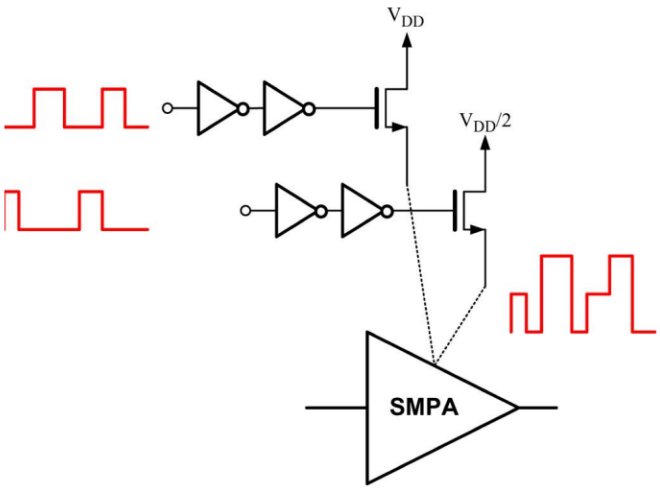

Figure 2.14: A three-level envelope-DSM [8]

overall transmitter efficiency. Even with narrow band signals such as EDGE, the coding efficiency is 25% in a BP-DSM. For wideband signals with a high PAPR such as WiMAX, the coding efficiency reduces to only around 6% [30].

2.2.5 Hybrid Approaches

As discussed previously, amplitude, PWM and DSM encoding have their advantages as well as limitations. Any combination of the three encoding approaches could be chosen to produce a hybrid encoding scheme. In recent years an increasing popularity in hybrid encoding has been witnessed, due to the potential to avoid the limitations of any one particular method.

One popular approach is to use a multi-level quantizer in DSM. The benefit of doing so is reduced quantization noise, given the same filter response. Figure 2.14 shows one digital polar transmitter [8], which uses a three-level quantizer in the envelope-DSM. Note that a multi-level quantizer is not supported by SMPA by nature. The multi-level waveform must be decomposed into multiple binary-level pulses, as shown in Figure 2.14. Using the same concept, hybrid approaches such as multi-level PWM [20, 31] and hybrid DSM-PWM [32] appear in the literature.

2.3 Conclusion

In general, prior work primarily focused on only amplitude- or time-domain encoding and has yielded sub-optimal overall performance due to sacrificing at least one of the following design objectives: design complexity, coding efficiency and dynamic range. Amplitude-encoding requires a large number of parallel SMPAs and a power combining stage resulting in greater design complexity. On the other hand, time-encoding results in out-of-band power dissipation because of spurs and harmonics

and, therefore, low coding efficiency.

This research takes a synergistic approach that accounts for the PA complexity in the early stage of the encoder design. The objectives of this research are:

1. to exploit existing digital encoding techniques and propose a new scheme that improves coding efficiency as well as dynamic range and is amenable to PA integration;

2. to present a high performance digital PA based on the proposed encoding scheme and demonstrate its advantages.

Chapter 3

A Hybrid Amplitude/Time Encoding Scheme for Enhancing Coding Efficiency and Dynamic Range in Digitally Modulated Power Amplifiers

In this chapter, a new approach is proposed based on hybrid amplitude/time encoding, which provides higher coding efficiency when compared to digital time-encoding, while simplifying the design of the SMPA stage compared to amplitude-encoding. We provide a framework to analyze the design space of the proposed hybrid amplitude/time encoding scheme and provide a new DMPA architecture, based on that encoding scheme, to illustrate its feasibility. Experiments were carried out to demonstrate the significant performance improvements of the novel approach in terms of coding efficiency and signal quality.

3.1 Binary Pulse Width Modulation

We begin with a discussion of the conventional binary PWM technique to provide context for the rest of this paper. An RF signal can be represented, in general, by its amplitude and phase, *i.e.*

$$s(t) = a(t)cos(\omega_c t + \varphi(t)) \tag{3.1}$$

where $a(t)$ is the instantaneous normalized amplitude ($a \in [0, 1]$), $\varphi(t)$ is the instantaneous phase and ω_c is the carrier frequency. A binary waveform has only two amplitude levels: low and high. A binary PWM generates bursts of RF signals, the width of which is proportional to the signal amplitude $a(t)$. Despite the simplicity of PWM, there are two fundamental limitations associated with this encoding technique, as discussed below.

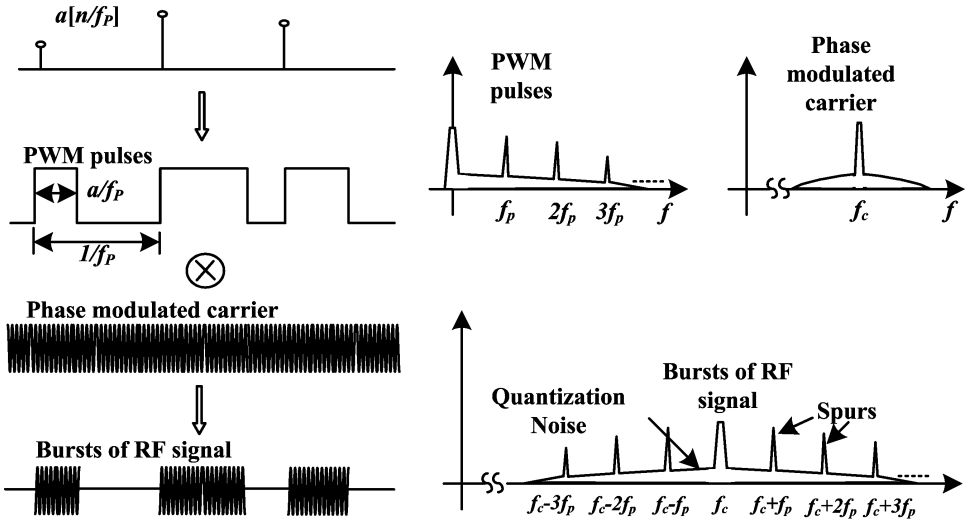

Figure 3.1: Operation principles of binary PWM

3.1.1 Coding Efficiency

The coding efficiency of an encoder is defined as the in-band power over total signal power, both in-band and out-of-band, and is an important figure of merit for the digital transmitter. Encoding techniques such as binary PWM result in an output spectrum that has both the desired in-band and undesired out-of-band components, as shown in Figure 3.1.

The coding efficiency for a signal with a constant envelope a (*i.e.*, $a[n] = a \; \forall n$) is defined as $\eta(a)$. In binary PWM, coding efficiency $\eta(a)$ is equal to the amplitude a and the duty cycle d [23]

$$\eta(a) = a = d \tag{3.2}$$

Equation (3.2) indicates that the coding efficiency $\eta(a)$ is 100% at maximum output power, when a is 1. However, $\eta(a)$ reduces linearly for smaller output amplitudes.

The mean coding efficiency, η_{mean}, for a varying envelope signal depends on the probability distribution function (*pdf*) of its amplitude, $f(a)$, and can be written as

$$\eta_{mean} = \frac{P_{ave}}{P_{total}} = \frac{\int_0^1 a^2 f(a) da}{\int_0^1 \frac{a^2 f(a)}{\eta(a)} da} \tag{3.3}$$

Figure 3.2 shows the *pdf* of amplitude for three modulation techniques [1] with increasing PAPR. We can observe that a significant fraction of amplitude density is concentrated at low amplitude values, which results in poor mean coding efficiency, as shown in Table 3.1. The mean coding efficiency is only 40% for 802.11a signals and up to 70.5% for WCDMA signals.

[1]The word *encoding* refers to the design of the encoder in a digital transmitter. The word *modulation* here refers to its conventional meaning of digital modulation.

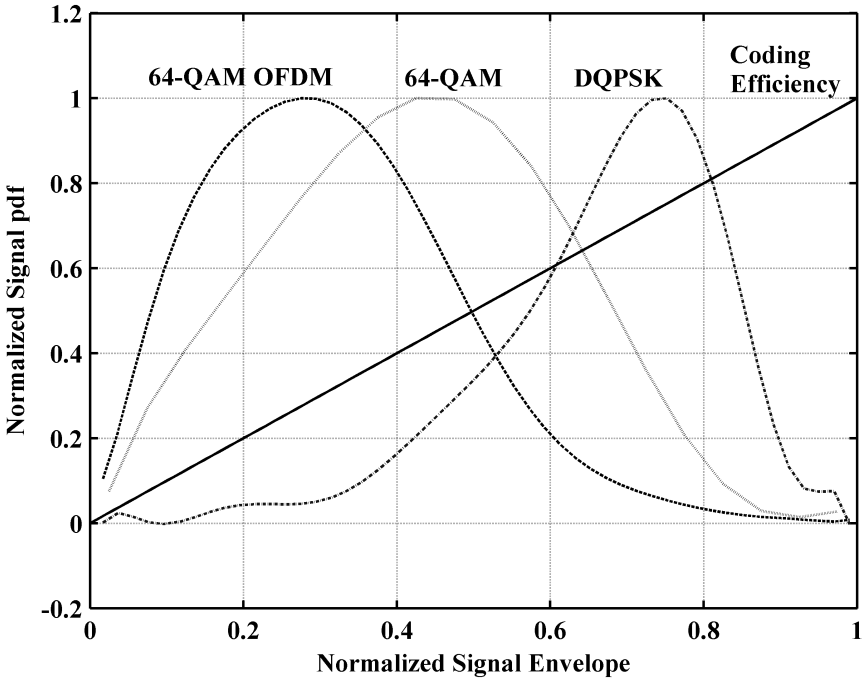

Figure 3.2: Normalized distribution of signals with different PAPR and coding efficiency of binary PWM

Table 3.1: Average Coding Efficiency of Binary PWM in Different Signals

	802.11a (64-QAM OFDM)	64-QAM	WCDMA (DQPSK)
a_{mean}	0.32	0.43	0.66
PAPR(dB)	9.9	7.3	3.6
η_{mean}	40%	51.8%	70.5%

3.1.2 Dynamic Range

The dynamic range (DR) of an encoding technique is defined as the ratio between the maximum and minimum amplitude values that it can encode. For binary PWM, the DR is limited by the smallest pulse-width, w_{min}, that can (i) be generated in a particular technology node, and (ii) be faithfully amplified by the SMPA, depending on its bandwidth. The DR in decibels can be written as

$$DR\,(dB) = 20log_{10}(\frac{1/f_p}{w_{min}}) \tag{3.4}$$

Note that, although the sampling frequency f_p can be reduced to increase the DR, it creates spurs that are closer to the carrier frequency and requires a high quality-factor filter with sharp roll-off. Recent works in the literature reported a dynamic range of at most 25 dB using binary PWM alone [4]. The limited dynamic range using binary PWM is hardly suitable for RF transmitters [24].

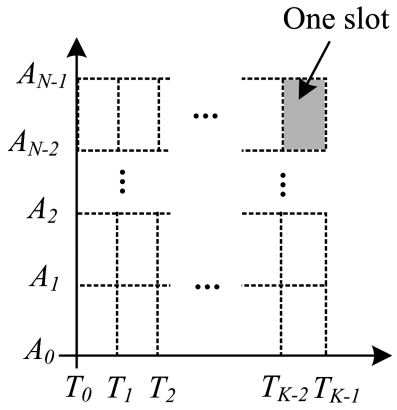

Figure 3.3: The concept of hybrid amplitude/time design space in a DMPA

3.2 Proposed Hybrid Amplitude/time Encoding Technique

Encoding with multiple amplitude levels can be used to address the low coding efficiency of binary PWM. However, to achieve a high dynamic range with an amplitude encoder, a large number of parallel SMPAs are required in the front end of the digital transmitter, along with a power combiner stage to add together the output from each SMPA. This results in significant area overhead and, more importantly, the efficiency of the power combiner becomes the new bottleneck in terms of overall transmitter efficiency.

To simultaneously achieve both high coding efficiency and high dynamic range, we are proposing a hybrid amplitude/time encoding approach that inherits the best features of time and amplitude encoding. The proposed hybrid encoding scheme encompasses previously proposed multi-level PWM schemes as special cases [31, 23, 20]. As we will discuss, hybrid amplitude/time encoding represents a large design space of alternatives, both in terms of constructing the multi-level signal and its decomposition into binary-level PWM signals that feed the SMPAs. The goal of this paper is to theoretically and empirically explore this large design space.

3.2.1 Hybrid Amplitude/Time Encoding

In hybrid amplitude/time encoding, the samples $a[n]$ are first quantized into L discrete levels. The L levels are then encoded using N amplitude steps and K time steps. N and K are chosen such that $L = (N-1) \times (K-1)$.

This results in a two dimensional design space, as shown in Figure 3.3, where one dimension is the number of amplitude steps, N, and the other dimension is the number of time steps, K. Note that this representation encompasses both binary PWM and amplitude-encoding as special cases. In the binary PWM case, $N = 2$ and $K = L + 1$ and in the amplitude-encoding case, $N = L + 1$ and $K = 2$.

A sample, $a[n] = l$ ($l \in [0, L]$), is encoded by filling up l time slots in the

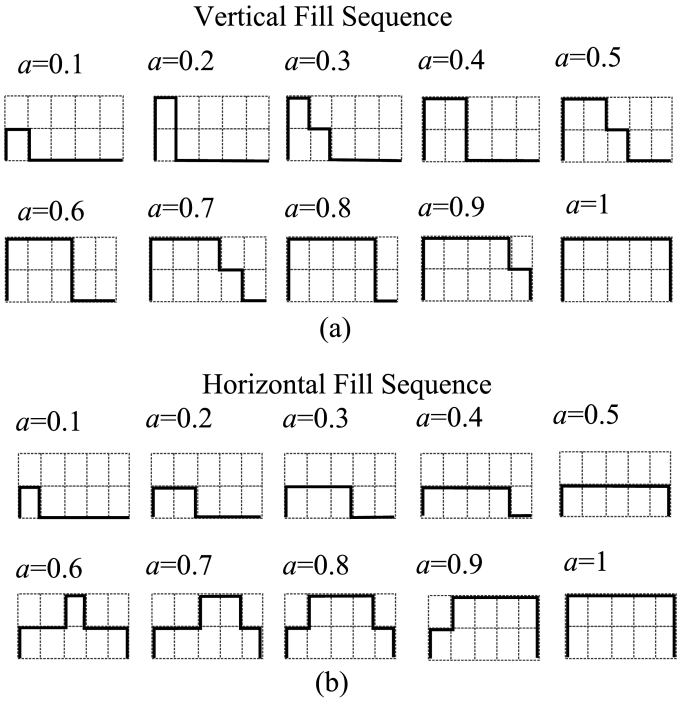

Figure 3.4: An example to illustrate the impact of fill sequences in the design space. (a). Vertical Fill Sequence. (b). Horizontal Fill Sequence.

$(N-1)\times(K-1)$ lattice. There are many ways of achieving this objective that yield different coding efficiencies. This is illustrated in Figure 3.4 for a ($N = 3,\ K = 6$) hybrid amplitude/time encoder using two different fill sequences.

Sequence 1, which we will refer to as a Vertical Fill Sequence (VFS), fills slots vertically as the amplitude is increased. This sequence is similar to previously proposed designs [31, 20]. In contrast, *Sequence 2*, which we will refer to as a Horizontal Fill Sequence (HFS), fills the slots horizontally as the amplitude is increased, *i.e.*, slots in the same amplitude level are filled before slots in the next higher amplitude level.

While both sequences result in the same in-band spectral content, they differ markedly in their coding efficiency. In fact, in the Appendix, we have shown that HFS has the highest coding efficiency while that of VFS is similar to binary PWM. Therefore, for the remainder of this paper, we focus only on HFS. We now derive an analytical expression for the coding efficiency, $\eta(a)$, of HFS.

3.2.2 Coding Efficiency for Horizontal Fill Sequences

To formally derive the coding efficiency, the analysis starts with a hybrid amplitude/time encoder with *N=3*, which will be generalized to higher levels. Figure 3.5.(a) depicts a multi-level pulse in one period $[-T_p/2, T_p/2]$. The time-step is assumed to be small enough to approximate an analog PWM for a closed-form solution, *i.e.*, we assume $K = \infty$ in this derivation.

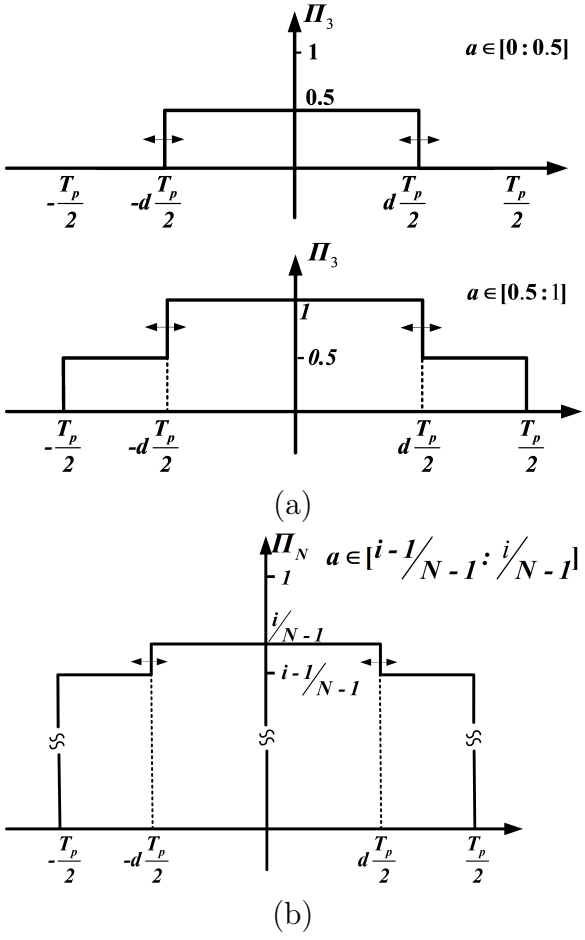

Figure 3.5: One period of (a). three-level PWM; (b). generalized *N+1* level PWM.

Duty cycle $d(a)$ is modulated as a function of the amplitude a

$$d(a) = \begin{cases} 2a & if\ a \in [0, 0.5] \\ 2(a - 0.5) & if\ a \in [0.5, 1] \end{cases} \tag{3.5}$$

The total power (assuming 1 Ω load) is

$$P_{total}(a) = \frac{1}{T_p} \int_{-T_p/2}^{T_p/2} \Pi_3(t, a)^2 dt \tag{3.6}$$

where $\Pi_3(t, a)$ is the multi-level pulse with *N=3*. The in-band power after up-conversion will correspond to the DC power at baseband and can be computed as

$$P_{in}(a) = \frac{1}{T_p^2} \left[\int_{-T_p/2}^{T_p/2} \Pi_3(t, a) dt \right]^2 \tag{3.7}$$

The coding efficiency for a given amplitude level can be computed as the ratio between the in-band power and total power

$$\eta(a) = \frac{P_{in}(a)}{P_{total}(a)} = \frac{\left[\int_{-T_p/2}^{T_p/2} \Pi_3(t, a) dt \right]^2}{T_p \int_{-T_p/2}^{T_p/2} \Pi_3(t, a)^2 dt} \tag{3.8}$$

The coding efficiency can be expressed as a piecewise non-linear function as follows

$$\eta(a) = \begin{cases} 2a & if\ a \in [0, 0.5] \\ \frac{a^2}{1.5a - 0.5} & if\ a \in [0.5, 1] \end{cases} \tag{3.9}$$

The same argument is generalized to a hybrid amplitude/time encoder with N levels, as shown in Figure 3.5.(b). The multi-level pulse can be analytically expressed by

$$\Pi_N(t, a) = \frac{\lfloor a(N-1) \rfloor}{N-1} rect(\frac{t}{T_p}) + \frac{1}{N-1} rect(\frac{t}{d(a)T_p}) \tag{3.10}$$

where $\lfloor \rfloor$ is the floor function and $d(a)$ is the duty cycle

$$d(a) = \frac{a}{1/(N-1)} - \lfloor \frac{a}{1/(N-1)} \rfloor \tag{3.11}$$

Finally, the coding efficiency, $\eta(a)$, can be expressed as a piecewise non-linear function as follows

$$\eta(a) = \begin{cases} a(N-1) & a \in [0, \frac{1}{N-1}] \\ \vdots & \vdots \\ \frac{a^2(N-1)^2}{i - i^2 - a(N-1) + 2a(N-1)i} & a \in [\frac{i-1}{N-1}, \frac{i}{N-1}] \\ \vdots & \vdots \\ \frac{a^2(N-1)^2}{(N-1)[2 - N + a - 2a(N-1)]} & a \in [\frac{N-2}{N-1}, N] \end{cases} \tag{3.12}$$

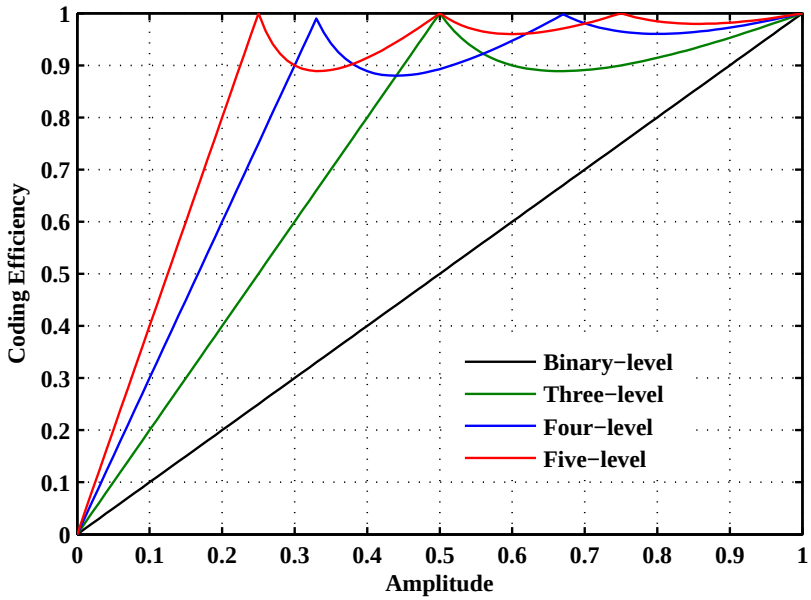

Figure 3.6: Coding efficiency of proposed hybrid encoding technique compared with binary PWM

The coding efficiency of a three- to five-level hybrid amplitude/time encoding is plotted in Figure 3.6, along with the coding efficiency of binary PWM. The figure indicates that $\eta(a)$ is greatly improved by moving to more than two levels. Even with three-level hybrid amplitude/time encoding, the coding efficiency is 100% for 6 dB back-off from peak power. For the same back-off, the corresponding coding efficiency for a conventional binary PWM is only 50%.

In the case of modulated signals (see Table I), the average coding efficiency η_{mean} with increasing amplitude levels, N, is plotted in Figure 3.7. The Figure indicates that the benefits of increasing N grow significantly between $N{=}2$ to $N{=}5$, but saturates quickly beyond that point.

3.3 Hybrid Amplitude/Time Waveform Decomposition

In the previous section, a hybrid amplitude/time encoding scheme was proposed to maximize coding efficiency for a given number of amplitude levels N and time steps K. This multi-level waveform, however, cannot be used to directly drive an SMPA since it requires a binary waveform. Therefore, the hybrid amplitude/time encoded waveform must be decomposed into *N-1* binary waveforms that are used to drive *N-1* parallel SMPAs. The SMPAs' outputs are combined using a power combining stage.

The transformation of a multi-level signal into multiple binary PWM signals can be done using *Thermometer Decomposition (TD)*. In this case, a separate SMPA is assigned to each of the *N-1* amplitude levels (excluding level-zero). Figure 3.8(a)

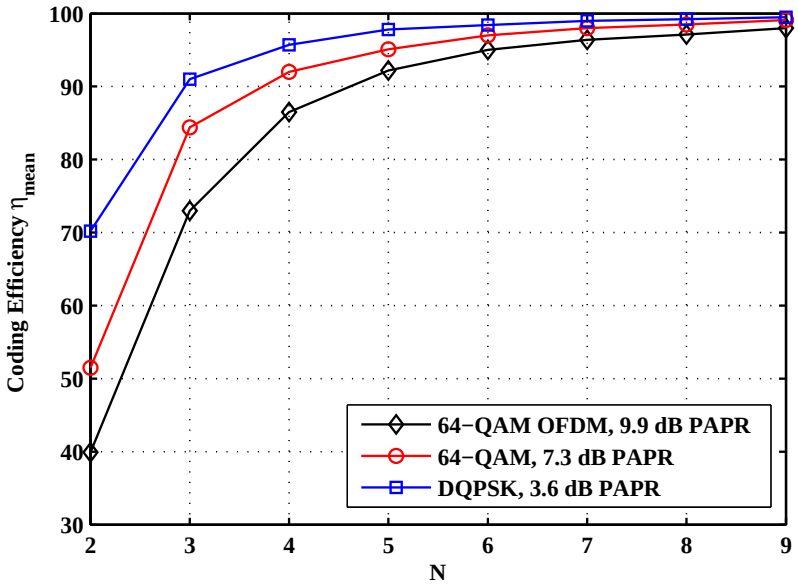

Figure 3.7: Average coding efficiency η_{mean} using the hybrid encoding technique. *N=1* is the conventional binary PWM

illustrates the *TD* method using an example waveform of *N=4* and *K=7*.

In this paper, we propose a new decomposition that we refer to as *Interleaved Decomposition (ID)*. The *TD* method generates *N-1* binary PWM waveforms with *comparable* duty-cycles that are phase shifted with respect to each other, as shown in Figure 3.8.(b). We note that the decomposed binary pulses using *ID* are conceptually similar to multi-phase encoding [33]. In multi-phase encoding, the duty cycle of every binary PWM is *identical*, because of which the time resolution required in every path is the same as that of binary PWM. By relaxing this restriction, *ID* is able to achieve a factor of *N-1* higher dynamic range than multi-phase encoding for a given minimal pulse width. Alternatively, to achieve the same dynamic range, the minimum pulse width required in multi-phase encoding is $T_p/\left[(N-1)(K-1)\right]$, while *ID* only requires pulses of minimum width $T_p/(K-1)$.

In general, if l slots are occupied in the hybrid amplitude/time waveform, and assuming that l_i represents the number of filled slots in the i^{th} binary decomposition, the following relationships hold for *ID*:

$$\sum_{i=1}^{N} l_i = l,\ l \in [0, (N-1)(K-1)] \tag{3.13}$$

$$|l_i - l_j| \leq 1\ \ \forall i, j \in [1, N-1] \tag{3.14}$$

Figure 3.8 demonstrates the principle of *ID* using an example waveform of *N=4* and *K=7*. The three binary PWM waveforms are offset by $T_p/3$. We can write the following expressions for $l_i\, \forall i \in [1, 3]$:

$$l_1 = q(l, N-1) + H[r(l, N-1) - \xi] \tag{3.15}$$

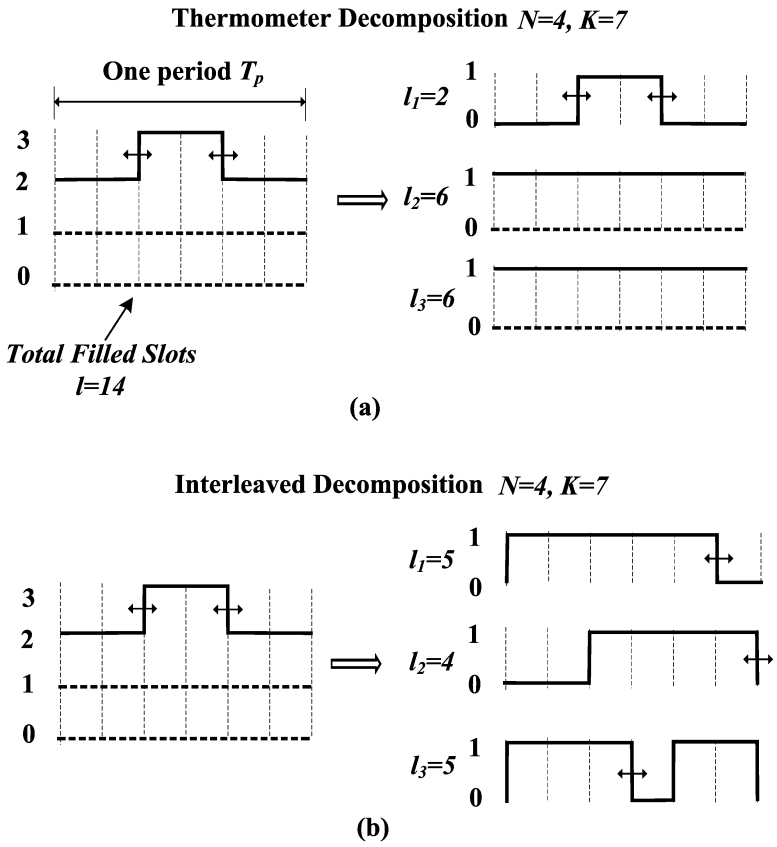

Figure 3.8: Two multi-level to binary decomposition methods. (a). thermometer decomposition. (b). interleaved decomposition.

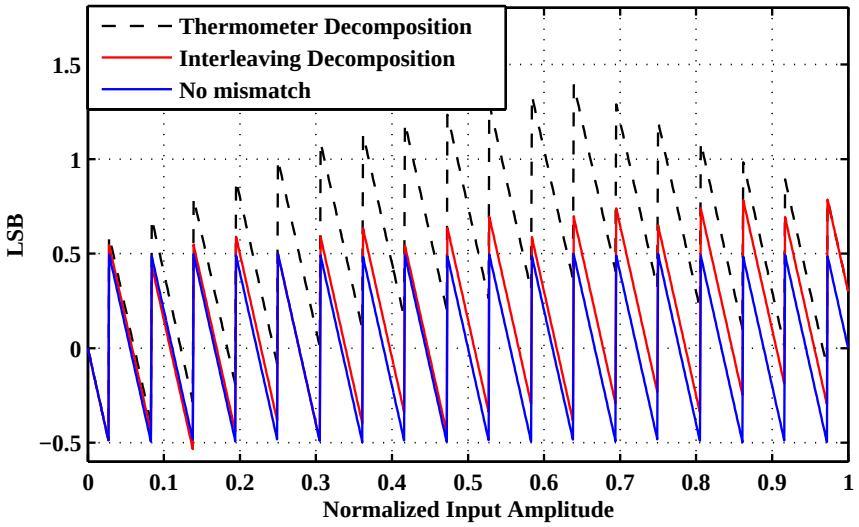

Figure 3.9: Comparison of quantization error with mismatch assuming +10%, 5% and -10% mismatch from ideal value in $l_1 - l_3$.

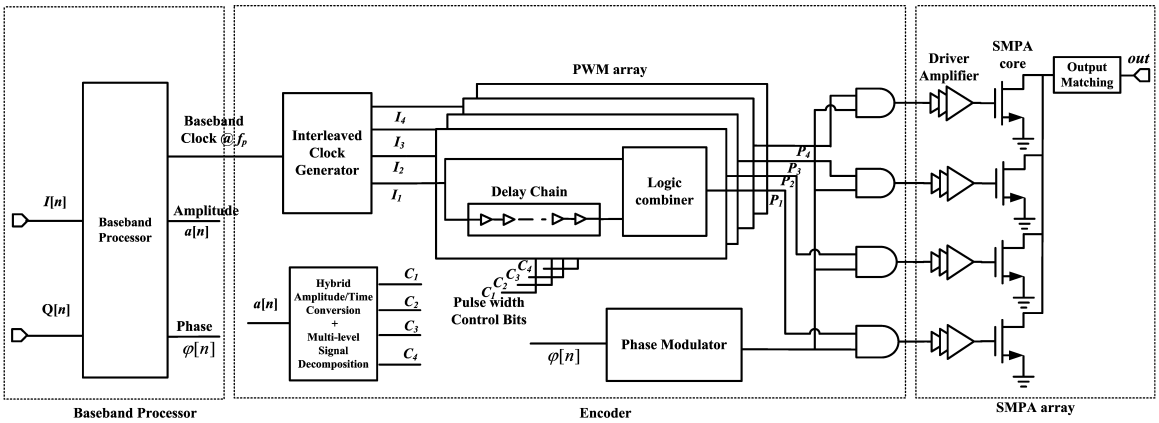

Figure 3.10: Potential implementation of the proposed hybrid amplitude/time encoding scheme.

$$l_2 = q(l, N-1) + H[r(l, N-1) - 2 - \xi] \tag{3.16}$$

$$l_3 = q(l, N-1) + H[r(l, N-1) - 1 - \xi] \tag{3.17}$$

where $q(a, b)$ and $r(a, b)$ represent the quotient and remainder obtained from dividing integers a and b, $H[n]$ is the Heaviside step function and ξ is a tiny number such that $H[0 - \xi]$ =0. In case of $l = 14$, $l_1 = 5$, $l_2 = 4$ and $l_3 = 5$. The quotient in (3.15)-(3.17) ensures that $l_i \, \forall i \in [1, 3]$ has almost identical values and the Heaviside function finely adjusts the individual values of l_i.

In a practical implementation, the decomposed binary signals are subject to random mismatch in their amplitudes and phases. The mismatch results in increased quantization error, which reduces the signal quality. Figure 3.9 shows the quantization error of a hybrid amplitude/time encoder ($N = 4$, $K = 7$) using both *TD* and *ID*. The peak error is 0.78 LSB using *ID* versus 1.4 LSB using *TD*. These results suggest that *ID* is more robust to mismatch compared to *TD*. In our experiments, we used the *ID* scheme to decompose the encoded signals.

The previously described new hybrid amplitude/time encoder and the waveform decomposition were exploited to sketch a complete architecture of a fully digitally modulated amplifier with improved efficiency, shown in Figure 3.10. The resulting DMPA architecture consists of a baseband processor, a signal encoder and an array of SMPAs. The baseband processor converts the input signal ($I[n]$, $Q[n]$) into polar form ($a[n]$,$\phi[n]$) and outputs a baseband clock at f_p that feeds to the PWM array. Signal amplitude $a[n]$ is converted to a multi-level waveform using the proposed hybrid amplitude/time encoding scheme and then decomposed into multiple binary pulses, P_i. As an intermediate step, control signals, C_i, selects the delay along the digital delay line, followed by combinational logic gates, to generate P_i of different pulse width. The resulting four PWM signals, P_i, are multiplied by the phase modulated signal before feeding the four SMPAs.

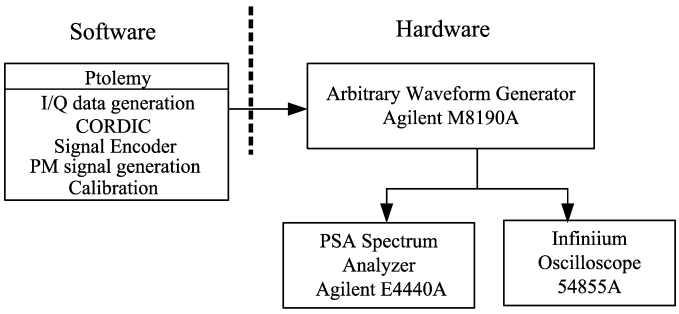

Figure 3.11: Development platform for the validation of proposed encoding scheme

3.4 Validation Results

The validation of the proposed encoding scheme was conducted using a hybrid software/hardware development platform, shown in Figure 3.11. In this platform, the discrete time simulator Ptolemy (Agilent Technology) was used to emulate the encoder and the resulting signal was uploaded to a high speed arbitrary waveform generator (AWG) M8190A (Agilent Technology). A high speed oscilloscope (Infiniium 54855A) was used to analyze the encoded signal properties.

The validation was conducted with a reference design in which $N = 5$ and $K = 40$. The number of levels N was set to 5, since Figure 3.6 indicates that the coding efficiency saturates beyond that point. This configuration resulted in a dynamic range of 44 dB, which is a significant improvement over reported data using binary PWM [4, 5].

We conducted the experiments in three phases. In the first phase, a ramp signal with amplitude $a[n]$ varying from 0 to 1 was used as a test signal and the coding efficiency was assessed. Figure 13 shows the measured efficiency, $\eta(a)$, of the proposed encoder compared to that of binary PWM encoding. The analytically derived efficiency was also shown to agree with measurement. The recorded values of the efficiency $\eta^{SW}(a)$ in Figure 3.12 were obtained using a square-wave carrier. This efficiency was related to the continuous wave one, $\eta^{CW}(a)$, according to the following equation

$$\eta^{SW}(a) = \frac{8}{\pi^2}\eta^{CW}(a) \tag{3.18}$$

According to Figure 3.12, the peak efficiency was bounded by the theoretical limit of 81.6%, as predicted by (18). One can also observe the significant efficiency improvement brought about by the proposed encoder. This improvement increased as $a[n]$ value decreased. As an example, for $a[n] = 0.25$, an improvement of about 60% in coding efficiency was obtained.

In the second phase of validation, a test signal with 10 MHz bandwidth employing 64QAM modulation with pseudo-random data stream was used. The baseband sampling frequency f_p and the carrier frequency f_c were equal to 50 MHz and 480 MHz respectively. We noted that the minimum pulse width used for the binary PWM encoding had to be set four times smaller than that of the hybrid ampli-

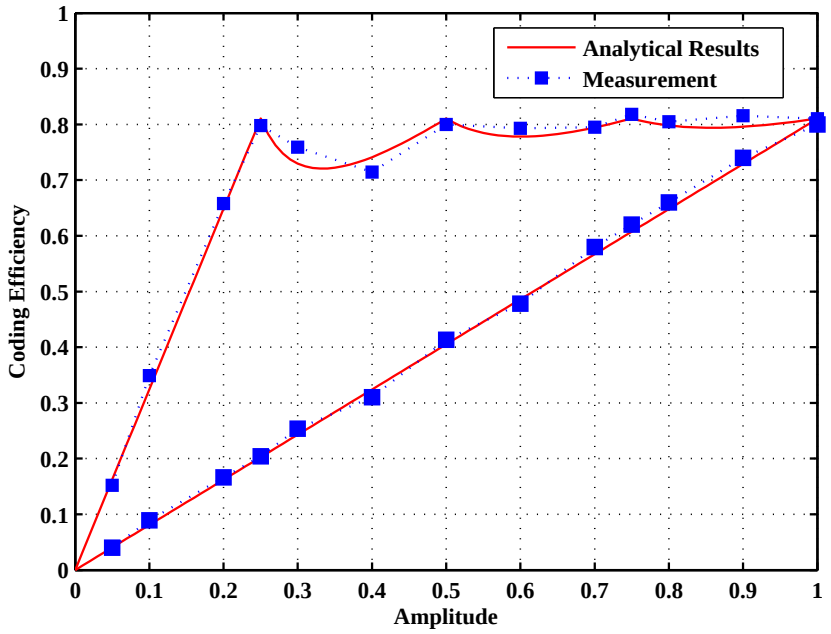

Figure 3.12: Measurement results compared with analytical coding efficiency using square wave carrier

tude/time encoding to ensure the same dynamic range. In a practical implementation, this would have a drastic impact on the timing requirement of the signal encoder and the bandwidth of the PA stage.

The spectrum of the encoder output signal is shown in Figure 3.13 together with the spectrum obtained from binary PWM encoding. One can clearly observe the significant improvement of the quality of signal that manifested in lower spurs and translates into better efficiency. In addition, the adjacent channel power ratio (ACPR) was improved by 7 dB. Figure 3.14 shows the constellation of the measured encoder output signals and reveals a good signal quality, since the error-vector-magnitude (EVM) was limited to 2%.

A fast Fourier-transform (FFT) was applied to a record of the encoder output signal (25-µsec) and used in (3.19) to compute the average coding efficiency, η_{mean}

$$\eta_{mean} = \frac{\sum_{f_C - f_B/2}^{f_C + f_B/2} |A(f)|^2}{\sum_{0}^{f_s/2} |A(f)|^2} \tag{3.19}$$

where $A(f)$ denotes the FFT of the encoder output signal, f_s is the AWG sampling clock rate and f_B is the signal bandwidth. The measured η_{mean} for a 64QAM signal was equal to 78%, which is 27% higher than that of binary PWM encoding.

The third phase of the encoder validation was conducted using a more realistic communication signal synthesized according to the WLAN standard (802.11g). The test signal had a 20 MHz bandwidth and employed 64QAM-OFDM modulation. The baseband sampling frequency f_p and the carrier frequency f_c were equal to 80 MHz and 448 MHz respectively.

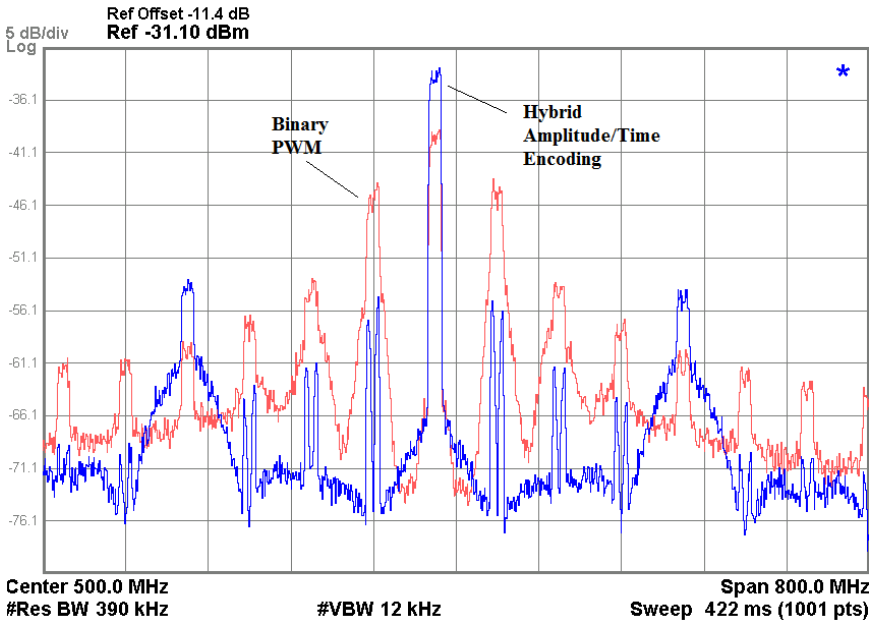

Figure 3.13: Measured output spectrum of a 10 MHz bandwidth 64QAM signal

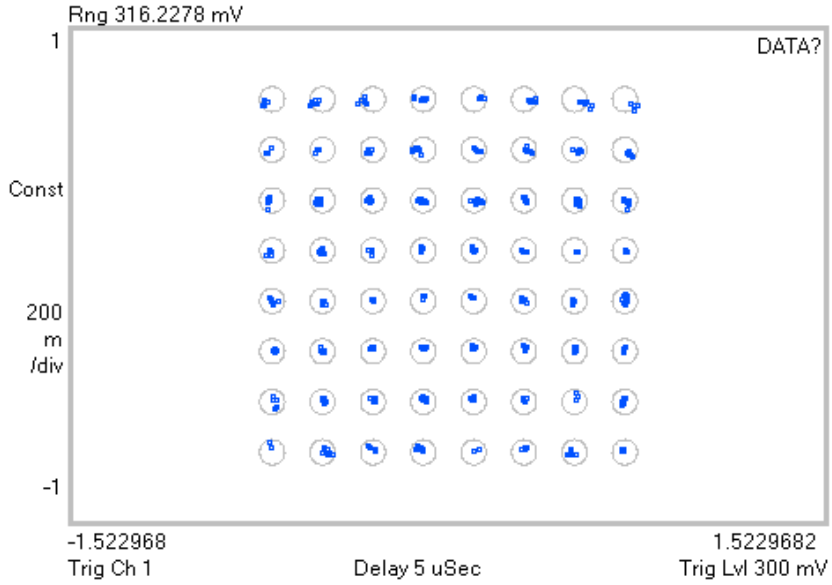

Figure 3.14: Constellation plots for a 10 MHz bandwidth 64QAM signal

Table 3.2: Summary of Test Results

	PAPR (dB)	η_{mean}^{SW}	EVM (dB)	ACPR (dB)
64QAM	7.6	78.1%	-33.5	-33
WLAN	9.9	74.9%	-29.1	-31.6

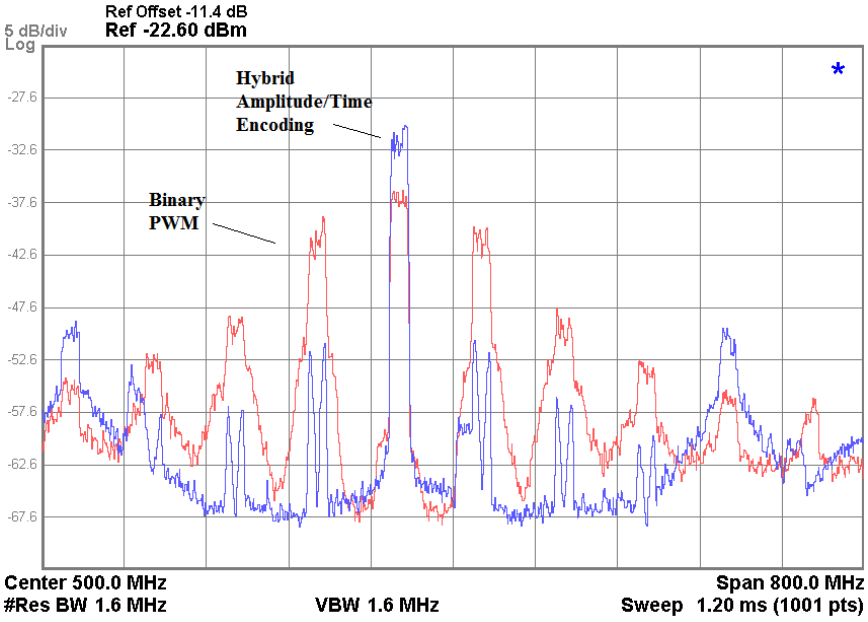

Figure 3.15: Output spectrum of 20 MHz WLAN signals with no re-construction filter.

The measured average coding efficiency η_{mean} was 74.9% for the hybrid amplitude/time encoder, which is 35% higher than that achieved using binary PWM, as reported in Table II. Despite the high PAPR of the WLAN signal (9.9 dB), the proposed encoding scheme allowed for an average efficiency that was only 6.7% lower than the peak efficiency.

The significant improvement of the coding efficiency is reflected in Figure 3.15, which shows the encoder output spectrum along with the binary PWM spectrum. Note from the Figure that the nearest spur to carrier frequency is -22 dBc lower. The measured ACPR was -31.6 dBc, as reported in Table 3.2. The close-in spectrum is shown in the same Figure.

Figure 3.16 shows the constellation of the demodulated signal, which resulted in an EVM of -29 dB. The EVM was limited by the phase resolution that could be generated from the AWG, which was approximately 5-bit in the experiment. The signal-to-noise ratio and EVM could be further improved using an application-specific integrated circuit (ASIC), in which a picoseconds resolution clock could be achieved [34].

Figure 3.16: Constellation plots for a WLAN signal.

3.5 Conclusion

This chapter proposed a theoretical framework to analyze the time- and amplitude-encoding design space for DMPAs. A new hybrid amplitude/time encoding scheme was devised to optimize the coding efficiency which consequently improved the power efficiency and extended the dynamic range of the DMPA. In addition, the new *ID* method was applied to transform the multi-level output signal of the hybrid amplitude/time encoder into multiple binary PWM waveforms used to drive parallel SMPAs. Experimental proof-of-concept results indicated significant improvements in the coding efficiency and the dynamic range achieved by the novel hybrid encoding scheme when compared with the conventional PWM scheme. As an example, the application of the proposed encoding scheme to a WLAN signal characterized by a PAPR equal to 9.9 dB revealed a measured average encoding efficiency of about 75%, which is 35% higher than that achieved using binary PWM. The measured EVM of about -30 dB corroborates the signal quality improvement as theoretically predicted.

Chapter 4

A Current Mode Multi-way Class-D CMOS Power Amplifier for Hybrid Amplitude/Time Encoded Digital Transmitters

Switch mode power amplifiers (SMPA s) are attractive in digital transmitters due to their higher efficiency and wider bandwidth as compared to a conventional transconductive power amplifier. Moreover, implementing SMPAs in CMOS technology is advantageous because CMOS is optimized for high speed switching operations.

The hybrid encoding scheme outlined in the last chapter requires multi-way SMPAs. In this chapter, we present a current mode class-D power amplifier that is amenable to implementing the proposed hybrid amplitude/time encoding scheme. The design has been realized in 130-nm CMOS technology. This chapter starts with a discussion of SMPA topologies, followed by theoretical analyses and circuit implementations. Validation results are presented and demonstrate the advantages of using the proposed encoding scheme in a complete digital transmitter.

4.1 PA Topology Selection

4.1.1 Voltage Mode Class-D (VMCD) PA

Schematic of a voltage model Class-D (VMCD) PA is shown in Figure 4.1. The pull-down nmos transistor, M_1, and pull-up pmos transistor, M_2, are switched on and off 180^o out-of-phase. A series resonator, formed by L_1 and C_1, filters out the higher harmonics and only allows current at the fundamental frequency to pass. The voltage and current waveform are plotted in the same Figure. In an ideal case, the voltage waveform only consists of the fundamental and odd harmonics, and the half-wave rectified sine wave current only consists of the fundamental and even harmonics, hence resulting in no power consumption at harmonic frequency and 100% efficiency.

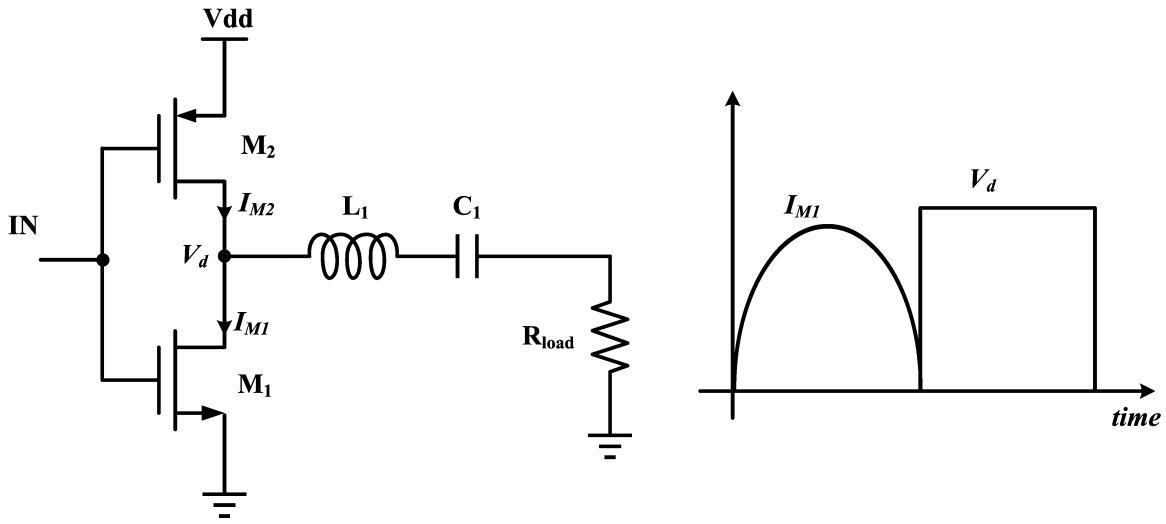

Figure 4.1: Basic schematic of a voltage mode class-D PA

Despite its simplicity, a VMCD is usually limited to operations below the radio frequency. The capacitive switching loss, P_C, as associated with the parasitic capacitance of M_1 and M_2, is expressed as

$$P_C = \frac{1}{2} C_d V_{dd}^2 f \tag{4.1}$$

where f is the fundamental frequency and C_d is the total parasitic capacitance. The loss is identified as the dominant loss mechanism of VMCD beyond hundreds of megahertz [25] and, therefore, makes VMCD an unpopular choice for RF PA. Another limitation of VMCD is the shoot-through current. During the operation of a VMCD, there is a finite period of time that both nmos and pmos are ON, resulting in large current flowing between the V_{dd} and ground [7]. A third limitation is the pmos transistor, which is usually sized two to three times the width of its nmos counterpart, in order to achieve the same ON-resistance. The parasitic capacitance due to a larger device size further reduces the power efficiency and complicates the design of the driver amplifier [3].

4.1.2 Current Mode Class-D PA

Figure 4.2 shows a current mode class-D (CMCD) PA, in which current sources are used to replace the voltage sources of a VMCD PA. The parallel resonator tank, formed by L_1 and C_1, acts as high impedance at the fundamental frequency and short-circuits all higher-order harmonics. The voltage and current waveform of a CMCD, plotted in the same Figure, indicate that the current is a square waveform and the voltage is of a half-rectified sinusoidal shape, resulting in zero power consumption at the harmonic frequencies.

A significant advantage of CMCD over VMCD is that the drain parasitic capacitance can be easily absorbed into the resonator tank. At the instant the transistor is

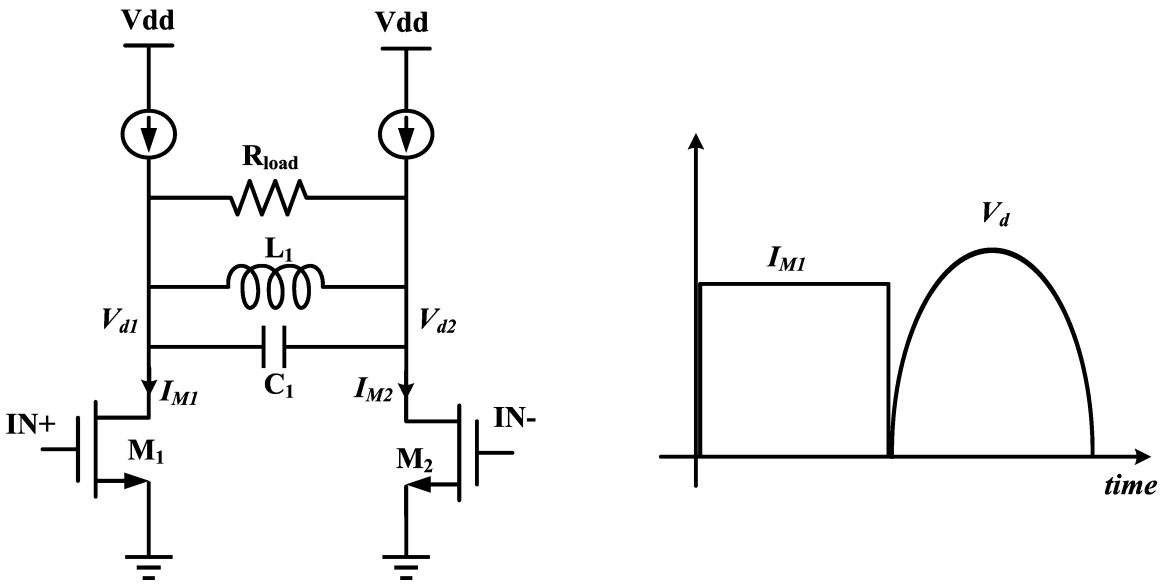

Figure 4.2: Basic schematic of a current mode class-D PA

switched ON, the drain voltage is not zero if the parasitic capacitor exists and takes a period of time to fully discharge, hence violating the zero-voltage-switching (ZVS) condition. However, if the parasitic capacitor is part of the resonator tank, the ZVS condition can still be achieved and results in no overlap between the drain voltage and current. This will be further analyzed in the design methodology section.

4.1.3 Class-E PA

The sharp, rectangular voltage/current waveform assumes an ideal switch in a class-D PA. At radio frequency, a switch is non-ideal and the transition time between the ON and OFF is non-negligible. As pointed out in [35], a class-E offers an alternative between the hard-switched class-D and conventional transconductive operation of a PA.

A schematic of a class-E configuration is shown in Figure 4.3, which consists of a transistor operating as a switch and a shunt capacitor, C_S. The basic principle behind class-E is that current from the choke inductor, L_d, is steered and sinks either through the switch to the ground or through the shunt capacitor/load network to the ground. Due to C_S, the developed drain voltage rises up slowly, rather than abruptly as in the case of a purely resistive load. A series resonator, formed by L_1 and C_1, only allows the current at the fundamental frequency to reach the load. In the original class-E mode [36], two conditions must be satisfied at the time of switching on

$$V_d(t) = 0 \mid_{switch-on} \tag{4.2}$$

$$\frac{dV_d(t)}{dt} = 0 \mid_{switch-on} \tag{4.3}$$

The first condition, ZVS, states that when the switch is on, capacitor C_S should

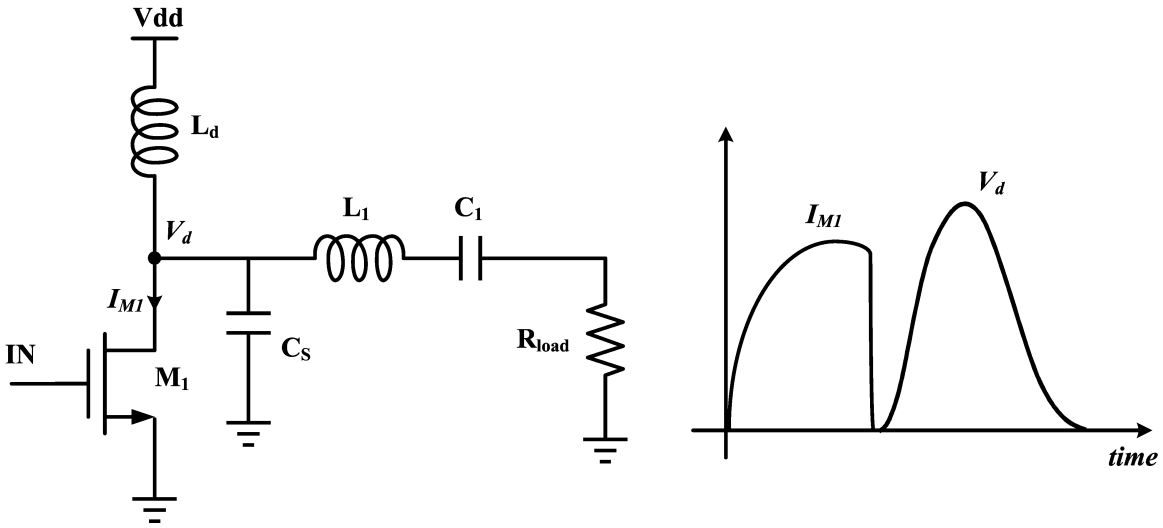

Figure 4.3: Basic schematic of a class-E PA

be fully discharged. Otherwise the remaining charge stored in the capacitor will be discharged through the switch to the ground, thus reducing the efficiency. The second condition, commonly known as $dZVS$, makes the circuit less sensitive to components, frequency and switching instance variations. Most of the time, it is easier to only satisfy the first condition and the power amplifier operates in a quasi-class E mode in RF PA [5].

There are also a number of limitations of Class-E PA in CMOS technology. Firstly, the peak voltage across the drain and source is as large as 3.6 times that of the supply voltage [37]. Voltage stress limits the application of class-E mode in advanced CMOS technology due to reliability concerns. Moreover, class-E requires multiple passive components, which occupy more silicon area and the overall quality factor quickly reduces [3].

4.1.4 Summary of Comparisons

Based on the comparisons between VMCD, CMCD and Class-E PA, we decided to choose CMCD for implementing the multi-way PA. The reasons include:

1. VMCD has large parasitic capacitance, as both nmos and pmos are required. These parasitic capacitance cannot be absorbed reactively and will reduce the efficiency. Simulations also indicate over 20% efficiency improvement using CMCD rather than VMCD PA at 2 GHz;

2. Theoretical class-E mode is difficult, if not impossible, to realize at gigahertz frequency. The peak voltage of class-E is 15% higher than VMCD PA and class-E PA output network requires more passive components.

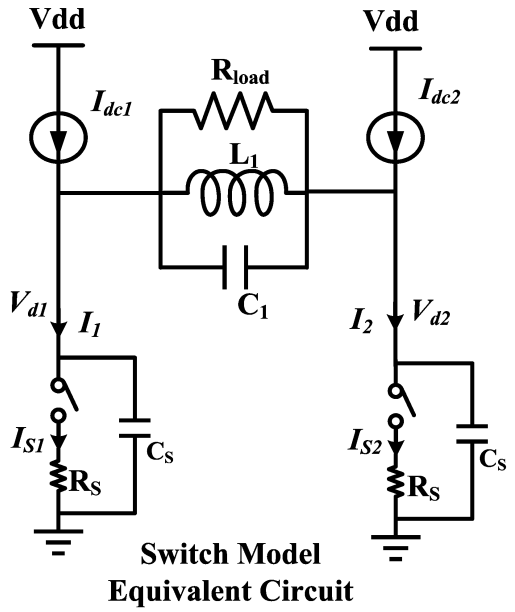

Figure 4.4: Equivalent circuits of CMCD with realistic switch models.

4.2 Design Methodology in Current Mode Class-D PA

We started analyzing the CMCD PA by replacing the transistors with switch models, *i.e.*, an ideal switch with parasitic capacitor C_S in shunt and parasitic resistor R_S in series. Figure 4.4 shows the equivalent circuits of a CMCD PA. In an ideal CMCD PA where $C_S = 0$ and $R_S = 0$, the drain current I_{S1} has a square shape, as the switch steers current between I_{pk} during on-state and zero during off-state. A square wave only consists of odd harmonics and I_{S1} can be expressed in its Fourier series as

$$\begin{aligned} I_{S1}(t) &= I_0 + \sum_{k\ odd} I_k \times sin(k\omega_c t) \\ &= I_{DC1} - \sum_{k\ odd} \frac{4}{k\pi} I_{DC1} sin(k\omega_c t) \end{aligned} \tag{4.4}$$

where I_k is the Fourier series coefficient and $I_{DC1} = \frac{I_{pk}}{2}$ is the supply current from the DC choke.

With an ideal resonator, the output voltage, $V_{d1} - V_{d2}$, across load R_L is a perfect sine wave. V_{d1} and V_{d2} are half-wave rectified sine waves with 180^o phase difference. V_{d1} can be expressed in Fourier series as

$$V_{d1}(t) = V_0 + V_1 \times sin(\omega_c t) + \sum_{k\ even} v_k \times cos(k\omega_c t) \tag{4.5}$$

Important design equations can be found based on (4.4) and (4.5). The details have already been provided in [25] and only the results are summarized here. The

peak drain voltage, V_{pk}, is

$$V_{pk} = \pi V_{dd} \tag{4.6}$$

Equation (4.6) indicates that the CMCD PA's peak voltage is 3.14 times of V_{dd}, which is smaller than a Class-E PA ($3.6V_{dd}$). By extracting the magnitude of $e^{j\omega_c t}$ term in (4.5), the power delivered to R_L at the fundamental frequency would be

$$P_{out} = \frac{\pi^2 V_{dd}^2}{2R_L} \tag{4.7}$$

Equations (4.4)-(4.7) are well known in CMCD PA designs. However, the assumption of infinite impedance at odd harmonics and zero impedance at even harmonics are not necessarily true in modern technologies. The objective in this section was to develop a design methodology based on a realistic CMCD PA.

In order to analyze the operation more accurately, some assumptions made previously become invalid. We shall discuss some non-idealities and their effects on the PA performance.

4.2.1 Switch On-state Resistance

A transistor, when operated as a switch, has non-negligible on-state resistance (R_S). Given the expression of I_{S1} in (4.4), the power loss, P_{R_S}, on the switch is

$$P_{R_S} = 2 \times R_S \left(I_0^2 + \frac{1}{2} \sum_{k\ odd} I_k^2 \right) \tag{4.8}$$

Including R_S in a CMCD PA has changed the values of the Fourier coefficients to

$$I_0 = \frac{1}{2} \frac{\pi^2 V_{dd}}{\pi^2 R_S + 2 \times R_L} \tag{4.9}$$

$$I_{k\,\forall k\ odd} = \frac{1}{k} \frac{2\pi V_{dd}}{\pi^2 R_S + 2 \times R_L} \tag{4.10}$$

The revised output power can be found with $k = 1$

$$\begin{aligned} P_{out} &= I_1^2 \times R_L/2 \\ &= \frac{2\pi^2 V_{dd}^2 R_L}{\left(\pi^2 R_S + 2 \times R_L\right)^2} \end{aligned} \tag{4.11}$$

The dependence of efficiency η_{DE} on R_S can be found by omitting other sources of losses

$$\eta_{DE} = \frac{P_{out}}{I_0 V_{dd}} \tag{4.12}$$

The efficiency and output power are plotted versus increasing R_S in Figure 4.5. The calculation indicates a strong dependence of CMCD PA performance on R_S. With only 0.5 Ohm of switch resistance, the efficiency drops by 10% and the ouput power reduces from the theoretical value of 1.23 watt to 1 watt.

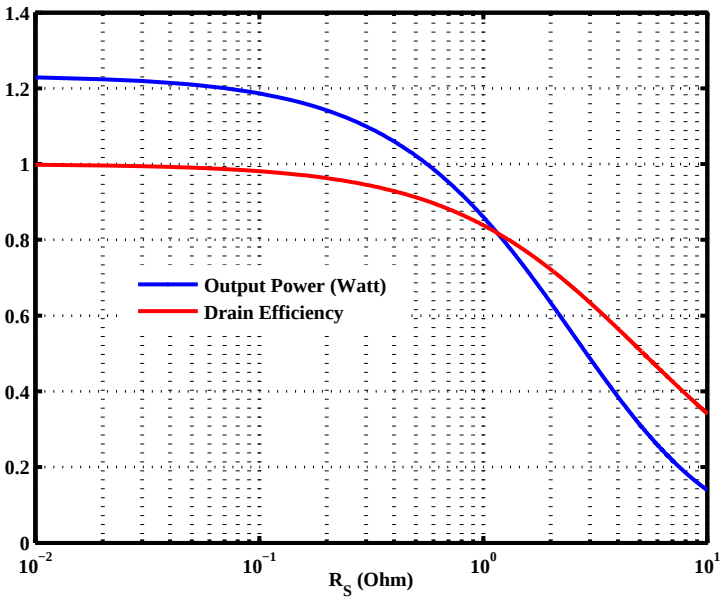

Figure 4.5: Efficiency and output power plot with increasing R_S where R_L is chosen to be 25 Ohm.

4.2.2 Inductor Power Loss

The parasitic resistor, R_{ind}, of inductor (L) has a significant impact on the efficiency. We denote this loss as P_{Q_L}, in which $Q_L = {}^{sL}/_{R_{ind}}$ is the inductor quality factor.

Only power loss at the fundamental frequency needs to be considered, due to its dominant percentage. In order to calculate P_{Q_L}, the serial L/R_{ind} network is transformed to a parallel L'/R'_{ind} network

$$L' = L(1 + Q_L^{-2}) \tag{4.13}$$

$$R'_{ind} = R_{ind}(1 + Q_L^2) \tag{4.14}$$

As R'_{ind} appears in parallel with R_{load}, it is convenient to express the inductor loss as a fraction of the output power or

$$P_{Q_L} = \beta P_{out} \tag{4.15}$$

where $\beta = \frac{Q_L R_{load}}{\omega_c L(1+Q_L^2)}$. The calculated η_{DE} with different Q_L is plotted in Figure (4.6). The plot indicates a strong dependence of efficiency on Q_L. Even with $Q_L = 10$, which is considered high in bulk silicon CMOS, the drain efficiency is only 80% of the peak value. In order to minimize power loss due to the inductor, we decided to use the bonding wire's parasitic inductor, which has a $Q_L \geq 50$.

4.2.3 LC-Tank Leakage

The LC resonator is not truly a zero-impedance at high harmonics. Any current leakage into the load is modelled as P_{leak} or

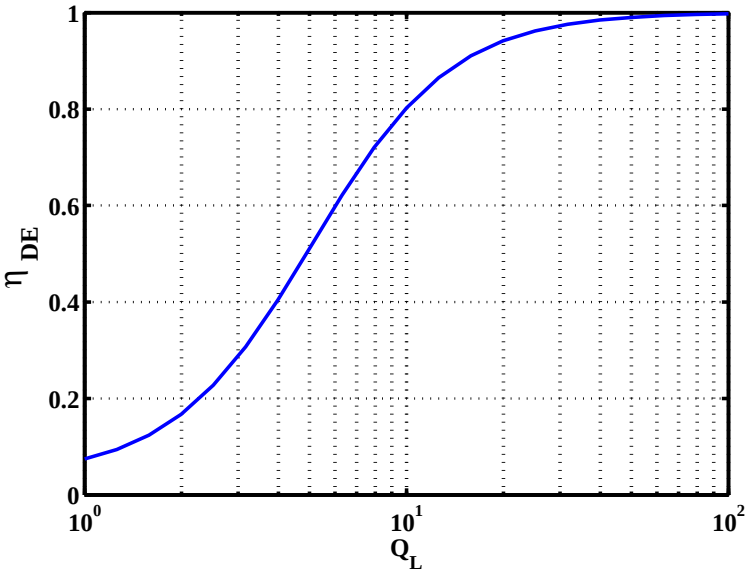

Figure 4.6: η_{DE} with different Q_L. Only inductor loss is modelled here. $L = 0.8nH$ and $f_c = 2GHz$.

$$P_{leak} = \sum_{k\ odd,\ k\geq 3} Re\left(I_k^2 \times Z_{T,k}\right) \tag{4.16}$$

where the $k-th$ harmonic impedance $Z_{T,k}$, is given by

$$\begin{aligned} Z_T &= \frac{sR_{load}L}{2R_{load} + 2sL + 2s^2R_{load}LC} \\ &= \frac{R_{load}}{2} \times \frac{1}{1 - jQ_t\frac{\omega_0}{\omega}(1 - \frac{\omega^2}{\omega_0^2})} \end{aligned} \tag{4.17}$$

and the tank quality factor is

$$Q_t = \frac{R_{load}}{\omega_0 L} \tag{4.18}$$

The calculated efficiency with leakage as the only source of loss is plotted in Figure (4.7). The results indicate that Q_t only contributes a small percentage to the total loss and $Q_t \geq 1$ is adequate for the design.

4.2.4 Parasitic Shunt Capacitor (C_S)

The parasitic capacitors C_S is charged to V_{pk} during off-state and discharged to zero during on-state. Charging and discharging current creates additional energy loss. Equivalent circuits of the CMCD PA at odd- and even-harmonic frequencies are shown in Figure 4.8. At the fundamental frequency, C_S can be absorbed into the resonator tank as part of the resonator, or

$$\omega_c = \frac{1}{\sqrt{\frac{L_1}{2}(2C_1 + C_S)}} \tag{4.19}$$

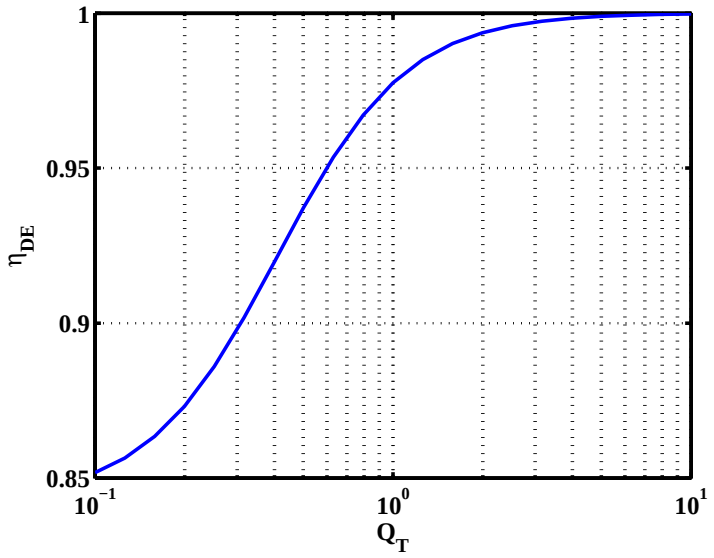

Figure 4.7: Variation of η_{DE} with increasing Q_t.

However, all the even-order harmonics are discharged through the switch and contribute to additional power loss [1]. In order to keep C_S small, the transistor size should be kept minimal. On the other hand, a small transistor has higher on-state resistance R_S, hence increasing the switch loss. For the time being, we neglected the analysis of power loss at the $2nd$-harmonics from C_S, due to the complexity resulting from the highly non-linear switch behaviour as well as its small percentage compared to other power losses.

4.3 Circuit Implementation Details

A block diagram of the proposed multi-way CMCD PA is shown in Figure 4.9. It comprises four major parts: the input buffer, the driver stage, the quad-PA stage and the output network. Detailed operations of each block are discussed in the following sub-sections. The circuits were designed using IBM 130nm CMOS technology and Cadence SpectreRF was used as the simulator.

4.3.1 Input Buffer

The input buffer serves two purposes: 1. it receives the phase modulated (PM) continuous wave coming off-chip; and 2. it converts the sinusoidal wave into pulse position modulated (PPM) digital pulses. As a high speed receiver, the input buffer should provide impedance termination (50 Ohm), have bandwidth up to at least the clock frequency and minimize its parasitic loading due to the bonding wire,

[1] A recent work from [3] resonates C_S at $2\omega_c$. However, in most cases, even-harmonics are consumed as heat, due to difficulty in achieving the desired load impedance at higher frequencies.

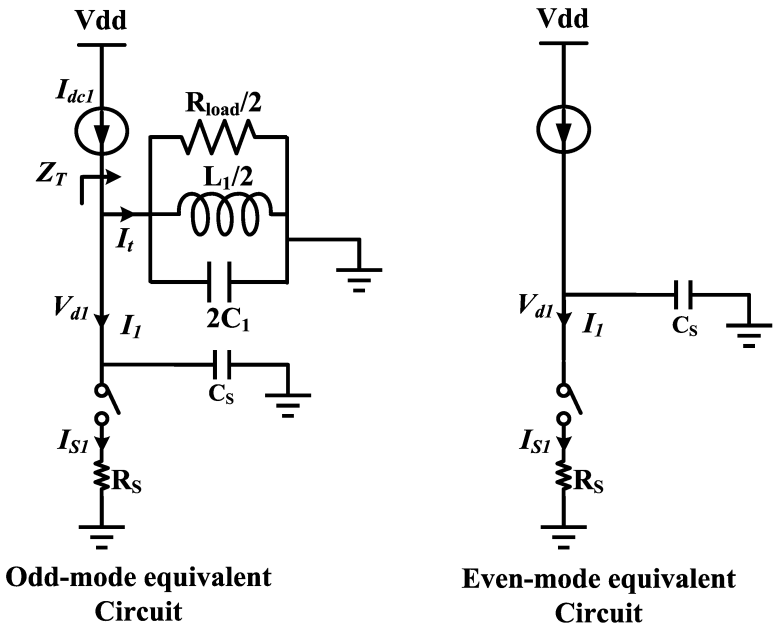

Figure 4.8: Equivalent circuits of the CMCD PA at odd- and even-harmonics

bonding pad and gate oxide. As a sinusoidal-to-square converter, the input buffer needs enough voltage gain to clamp the output from rail to rail.

A schematic of the input buffer is shown in Figure 4.10. Two 50 Ohm off-chips resistors were used to terminate the 50 Ohm transmission lines. To protect the gate oxide from ESD damages, double diodes were employed to provide an alternative discharge path during an ESD event. Four inverters were sized approximately with a ratio of 2. The first inverter, with its size denoted by $1X$, is the unit-inverter and used as a basis to construct the larger devices. The fourth inverter should be large enough to drive four parallel driver stages. Cross-coupled inverters were inserted between the differential branches to ensure that the clock signals were complementary.

4.3.2 Driver Stage

The PPM clock from the input buffer only consists of phase information. The amplitude information of the modulated signals was encoded using the proposed scheme and needed to be combined with the PPM clock at the driver stage. This was done through a 1-bit multiplier using an AND gate. A schematic of the driver stage is shown in Figure 4.11. Encoder bit, E, multiplies with the PPM clock at AND gate. Due to the large size of the PA stage, its parasitic capacitance significantly slowed down the rising and falling time of the digital clocks. The drain efficiency would be reduced if the power transistor spends more time in the linear region and less time in the cut-off and saturation regions. To avoid slow edge-transiting clocks, a total of six inverters were used as the driver amplifier before the PA stage. The size of the inverter chain had a tapering factor of 2 and the

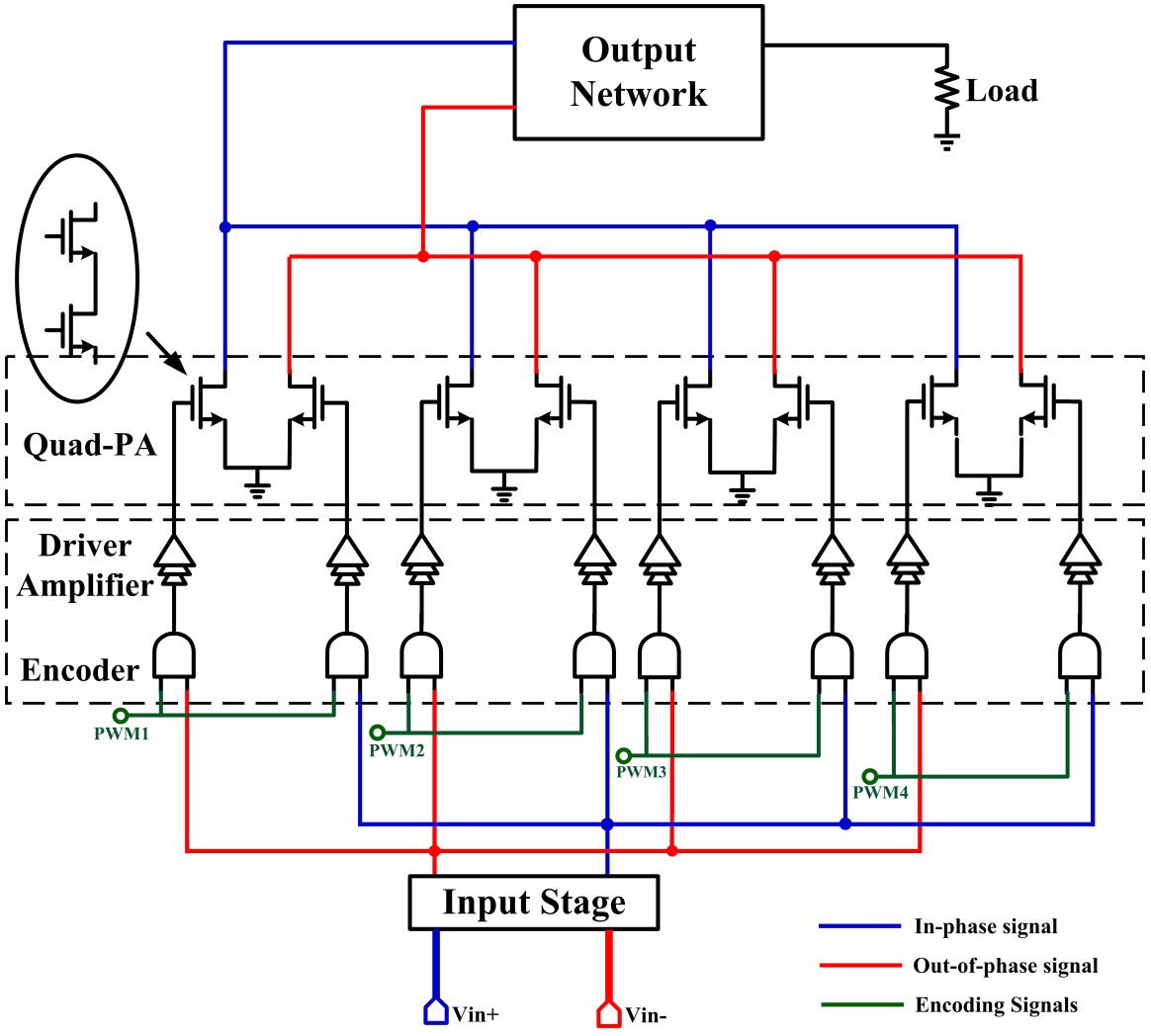

Figure 4.9: Block diagram of the proposed multi-way CMCD PA

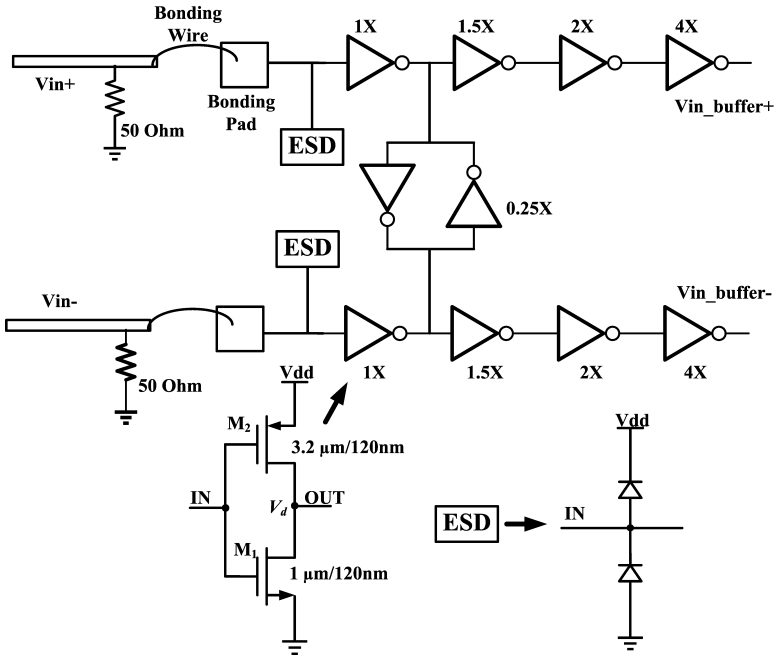

Figure 4.10: Schematic of the input buffer

total width (nmos+pmos) of the last inverter was half of the power transistor it was driving.

The simulated rising and falling edge transition time at the output of each inverter is plotted in Figure 4.12. All the parasitic effects from the layout were extracted in *Calibre* for accuracy. The input clock had a rising and falling time of 35 ps. The transistor was using typical-typical model and the simulation temperature was 45^oC. The plot indicated no sign of slowing down for the edge transition time as the clock propagated through the inverter chain. In other words, the inverter was sized large enough to drive the power transistor. Note the discrepancy between the rising and falling edge at the output of the last inverter. This is due to the non-linear C_{gs} of the n-type power transistor.

4.3.3 Quad-PA Stage

The schematic of the quad-PA stage is shown in Figure 4.13. The structure is similar to our analytical model of CMCD in Figure 4.4, except we divided the one big power transistor into four unit-PAs. The modulated input signals, $V_{i+}\ and\ V_{i-} \forall i \in [1, 4]$, enable or disable the switching of each individual unit-PA separately. When all the unit-PA are switching, the quad-PA is like a conventional current mode class-D. When any of the unit-PA is not switching, it appears as high impedance, thus minimizing its loading effect.

Figure 4.14 plots the simulated power added efficiency (PAE) of the quad-PA stage versus increasing transistor width. Transistor width $W1$ for the thin-oxide devices and $W2$ for the thick-oxide devices were swept for parametric analysis.

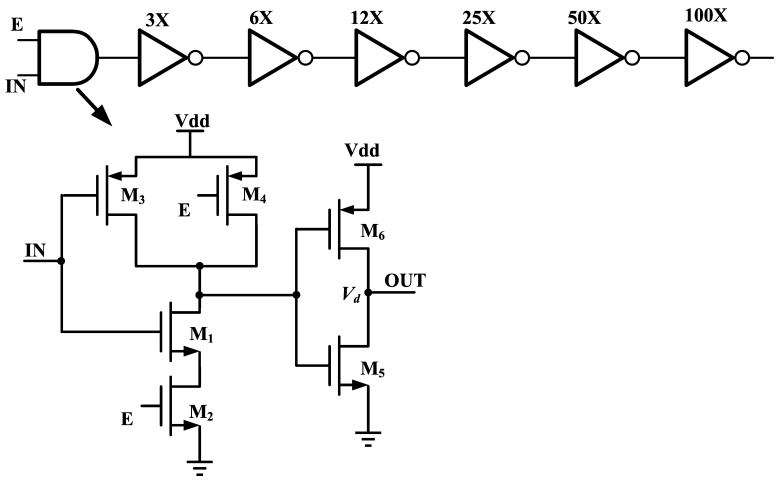

Figure 4.11: Schematic of the driver stage

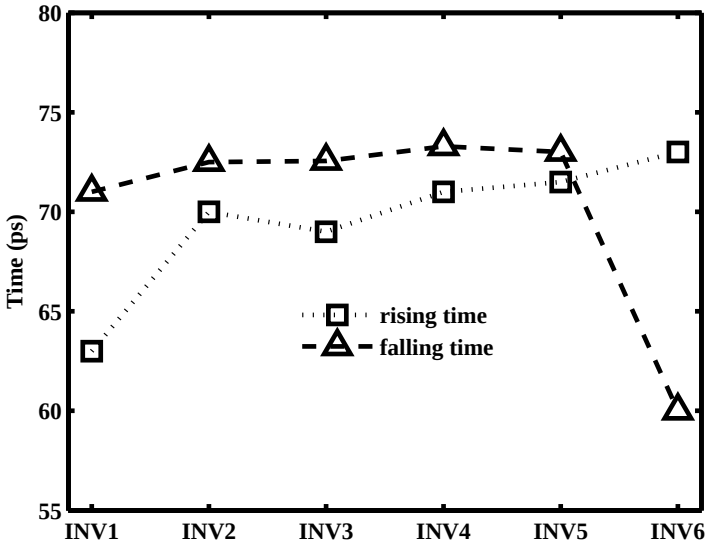

Figure 4.12: Simulated rising and falling time

The figure indicated that PAE almost saturated beyond $W1 = 3000\mu m$ and $W2 = 4000\mu m$. Further increase in the size resulted in marginal improvement in the PAE, but increased parasitics and difficulty in routing. The final transistor sizes are denoted in Figure 4.13.

The peak drain voltage of a CMCD is $3.14Vdd$, which is considered high since the recommended maximum voltage swing should be less than $2Vdd$ for long-term reliability [5]. The thin oxide devices were cascoded with the thick oxide devices to share the voltage stress. The nodal voltages during circuit operation were plotted in Figure 4.14. At any time, the following conditions must be satisfied

$$V_{G2} - V_{Th-thick} \leq 2V_{dd-thin} \tag{4.20}$$

$$3.14V_{dd} - V_{G2} + V_{Th-thick} \leq 2V_{dd-thick} \tag{4.21}$$

where $V_{dd-thin}$ is the rated operation voltage of thin oxide devices, $V_{Th-thick}$ and $V_{dd-thick}$ are the threshold voltage and rated operation voltage of thick oxide devices. The design we chose $V_{G2} = Vdd = 2.5V$, was able to satisfy the voltage stress limit and only required a single power supply.

4.3.4 Output Network

The output network has two purposes: 1. it converts differential signals into a single-ended signal; 2. it transforms the load impedance, R_L, to the desired impedance, Z_{opt}, at the fundamental frequency. A schematic of the output network is shown in Figure 4.16.(a). The differential-to-single-end conversion employed a balun. Due to the large ratio between R_L and Z_{opt}, the impedance transformation was carried out in two steps. The $50Ohm$ load impedance was firstly reduced to an intermediate value through the balun. Given a turns ratio of $1 : n$, the balanced impedance, R_T, of the balun was computed as

$$R_T = R_L/2n^2 \tag{4.22}$$

A lumped $L - element$ matching network, consisting of L_0 and C_0, then matched R_T to Z_{opt}.

The output network was implemented using off-chip components, as the on-chip passives typically have poor quality factors and occupy large silicon areas. The physical implementation of the designed output network is shown in Figure 4.16.(b). Inductor L_0 was realized by absorbing the parasitic inductance of bonding wires and PCB traces. Capacitor C_0 used a SMT RF capacitor and the balun is from Anaren Technologies. The complete output network is modelled by Agilent Momentum as a touchstone file and exported to Cadence for simulation.

4.4 Layout Design

The mixed signal operation of the proposed circuits, including four PAs with driver stage, makes the layout design vital. The layout could heavily influence the circuit performance and, therefore, requires special treatment during the design stage.

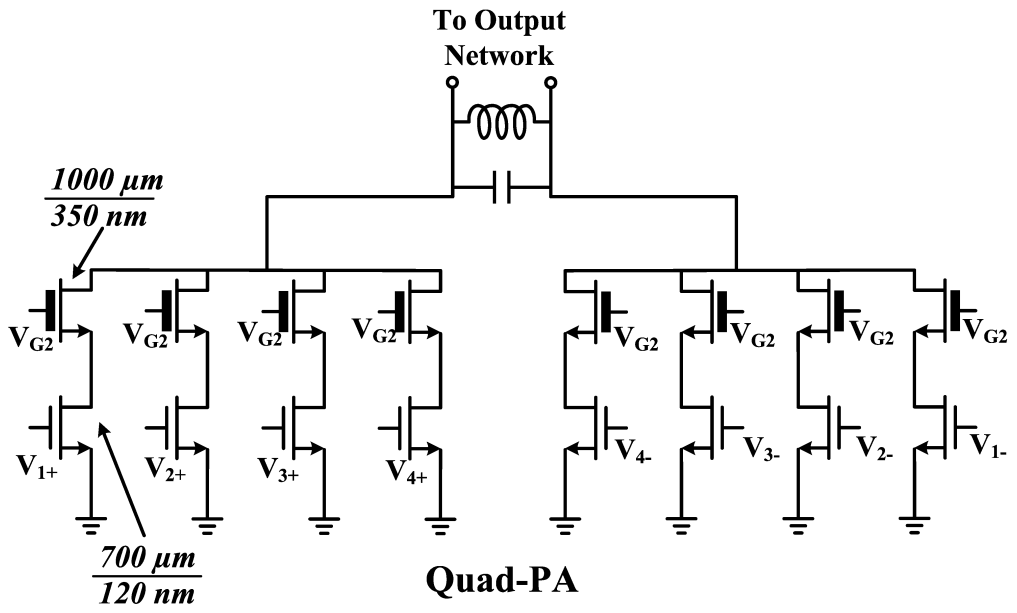

Figure 4.13: Schematic of the quad-PA stage

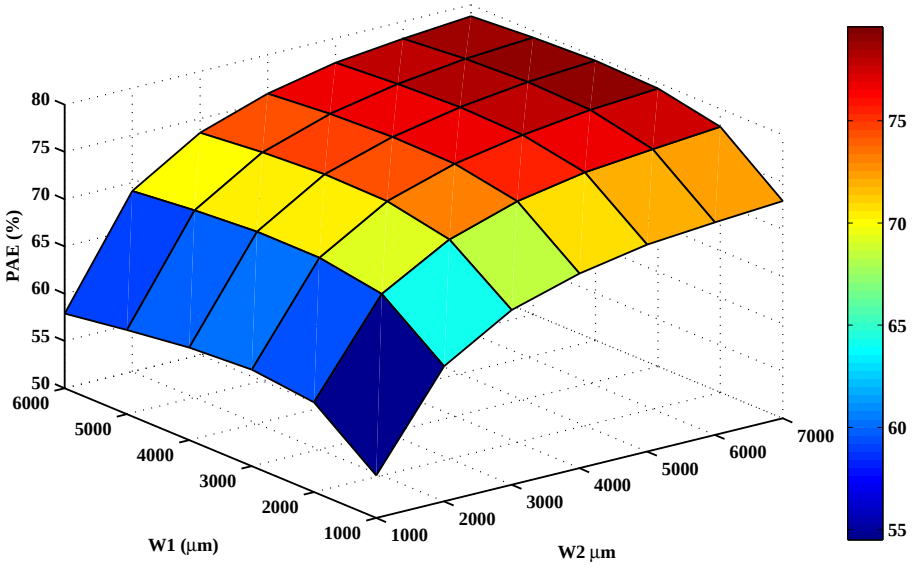

Figure 4.14: Simulated PAE versus transistor width in order to determine the optimal transistor sizes. The resonator was adjusted in each point in order to account for variations in the shunt parasitic capacitors.

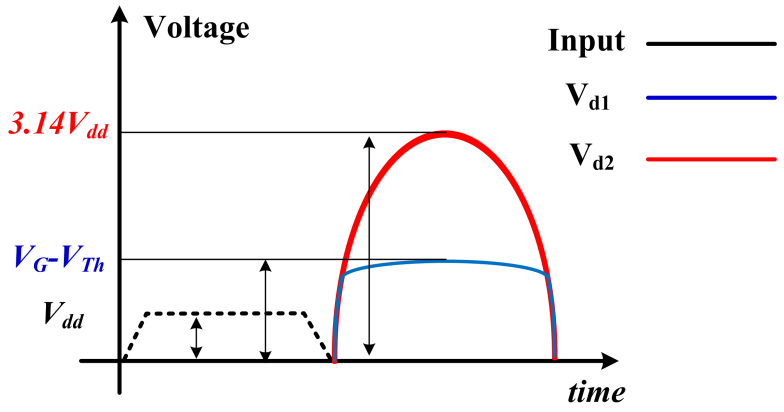

Figure 4.15: Transistor voltage stress during the switching

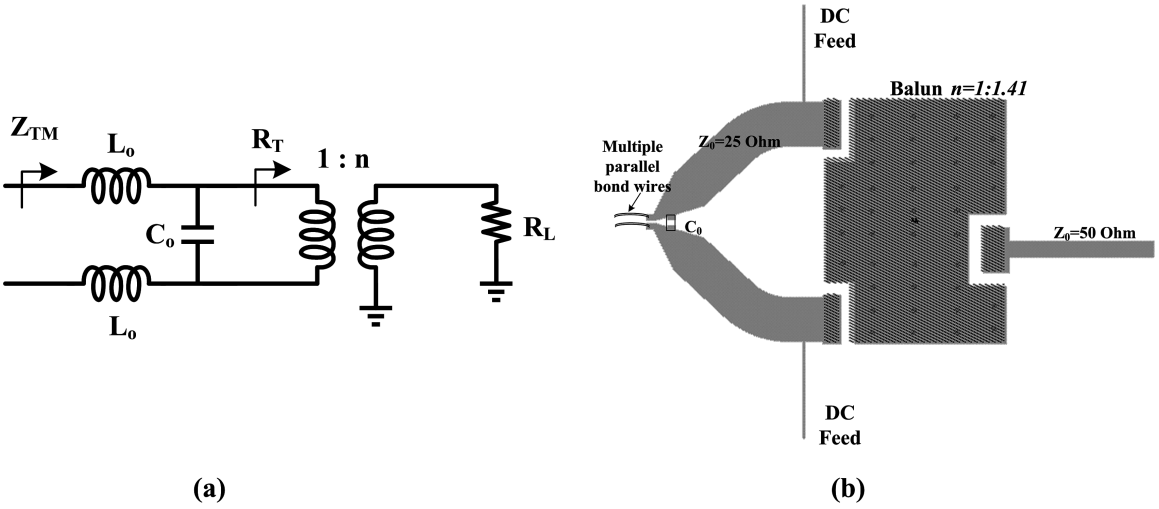

Figure 4.16: (a). schematic of the output network and (b). implementation of the output network using RO4003 dielectric material

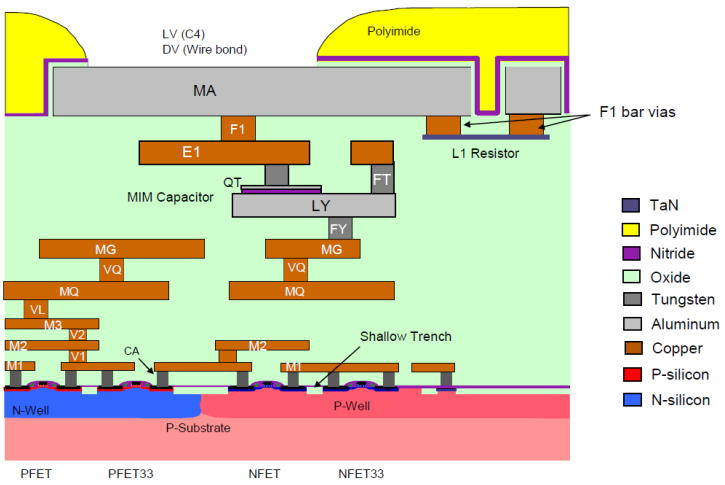

Figure 4.17: IBM CMOS 8RF-DM layer stack-up [9]

One challenge was the layout of the PA itself. The PA needs to be laid out with small parasitic capacitance as well as parasitic resistance. The power combining network, due to large sizes and lossy substrate, was distributed in nature and needed accurate EM simulations. Another challenge arose from the mixed-signal operation. The digital switching noise, generated from the inverter chains, could be easily coupled to the sensitive analog part through the semi-conducting substrate. A number of noise reduction techniques were adopted in this design. Deep submicron-meter CMOS technology presented a third challenge in the layout. Compared to micron-meter CMOS devices, modern CMOS has stringent rules on the metal density, metal-over-oxide ratio (antenna rule) and is more susceptible to electrostatic discharge (ESD) damage.

This section describes how the layout addressed the above challenges based on IBM 130 nm CMOS technology. The IBM 8RF-DM option is a high performance mixed-signal CMOS process that offers both thin and thick oxide transistors, MIM capacitor, precision resistors and inductors. The process includes up to 8 metal layers, of which the top three layers are RF thick metals. A detailed metal layer stack-up is provided in Figure 4.17.

4.4.1 Unit-transistor and unit-PA

Each unit-PA (Figure 4.13) has a gate width of $\sim 1mm$ and instantiating a $1mm$ device directly from the Foundry pcell would result in a transistor with an extremely large aspect ratio, un-equal signal delays and complex routing. Moreover, the Foundry's large-signal model tends to be in-adequate for modelling such a large device. In order to overcome these limitations, a large power transistor was divided into an array of smaller unit-transistors.

In this design, each unit-PA was divided into 14 smaller unit-transistors. In any unit-transistor, the common source transistor has a size of $5\mu m \times 10\ fingers$ and the cascoded transistor has a size of $7.2\mu m \times 10\ fingers$. The routing of a unit-

transistor is shown in Figure 4.18. The input was connected at both sides of the gate, thus reducing the gate resistance by half. To reduce the parasitic resistance, the input and output of the unit cascoded transistor was routed to the top-most RF metal layer through via-stack. To reduce the parasitic capacitance, the distance between signal lines and ground was maximized, thus making the metal layer, M1, as the ground plane.

Based on the unit-transistor layout, a unit-PA was constructed as shown in Figure 4.19. Each sub-PA comprises fourteen unit-transistors cells in a $7\,by\,2$ array. The input signal arrives at the left-most side of the unit-PA and output currents are collected at the right side, hence equalizing the delay mismatch between different unit-transistors.

4.4.2 PA Layout

The quad-PA stage was built from the unit-PA layout. Figure 4.20 describes two ways of laying out the quad-PA stage. The first approach would place differential unit-PAs adjacently, similar to the design in [5]. This arrangement would reduce the coupling capacitance between unit-PAs by half due to the use of differential signalling. However, the output manifold would become asymmetrical due to two different metals layers of different thickness and dielectric materials. The asymmetry would lead to phase difference between the output and different unit-PAs and lower the combining efficiency.

Since a PA's efficiency is directly affected by the output stage, in this design, the layout at the PA's output should be as symmetrical as possible. The layout 2 in Figure 4.20 was a preferred approach, in which identical current combining manifolds are realized. In order to improve the isolation between unit-PAs, double guard rings (n-type and p-type) were placed around each unit-transistor and ground plane are laid around unit-PAs.

Layout of the complete PA is shown in Figure 4.21 [2]. The entire PA occupies $2mm \times 1mm$ including IO pads. A large amount of on-chip decoupling capacitance (~600 pF) was provided between Vdd and Vss to reduce the digital switching noise. Shallow trench isolation was also used between the analog and digital circuits for noise isolation.

[2]The empty space in the chip layout was reserved for a multi-project run

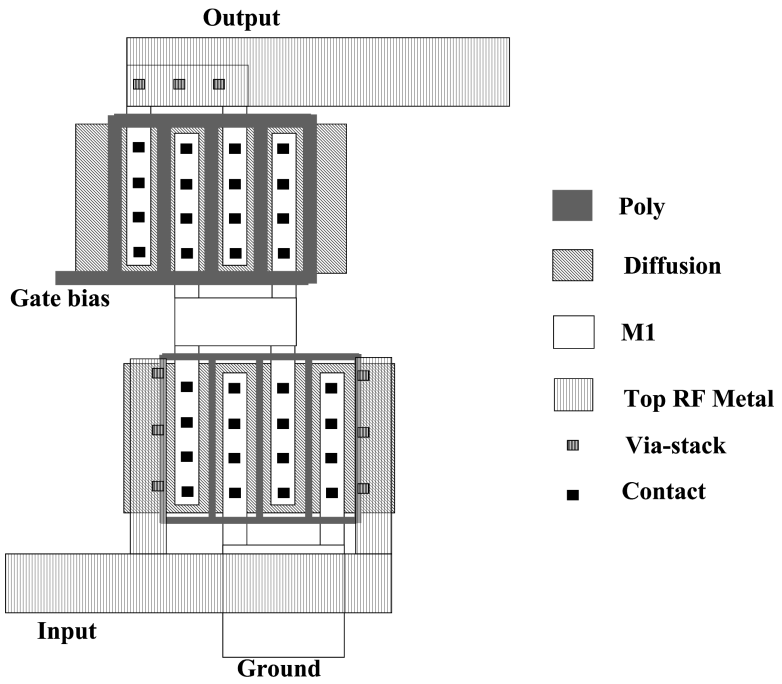

Figure 4.18: Routing of the unit-transistor layout

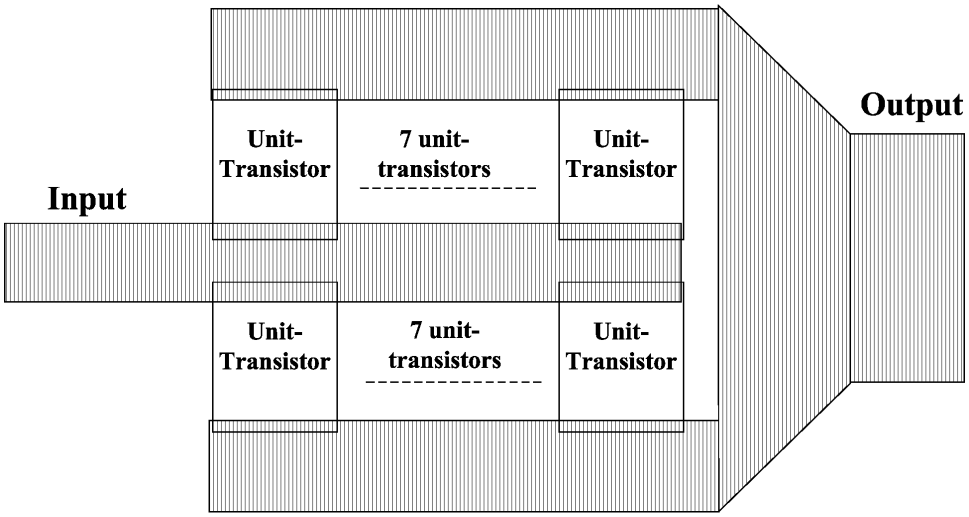

Figure 4.19: Configuration of one unit-PA in the quad-PA stage

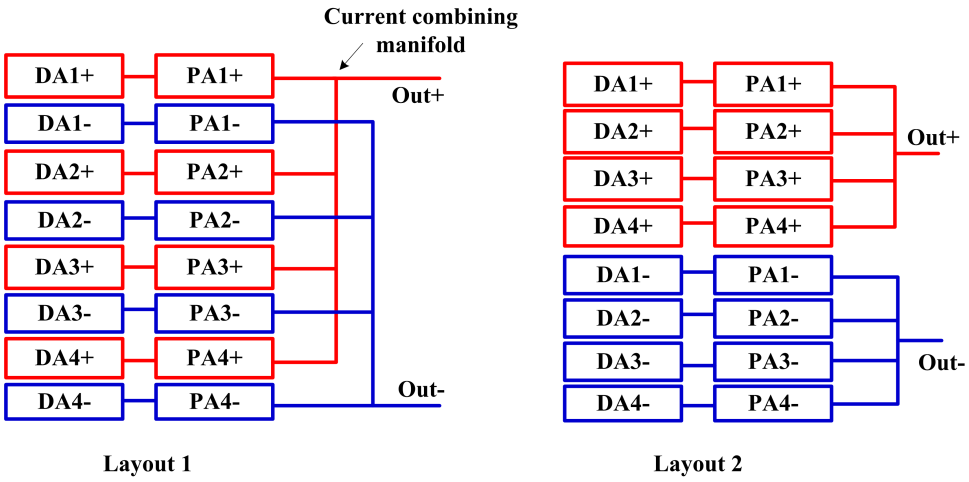

Figure 4.20: Configuration of unit-PAs in the quad-PA

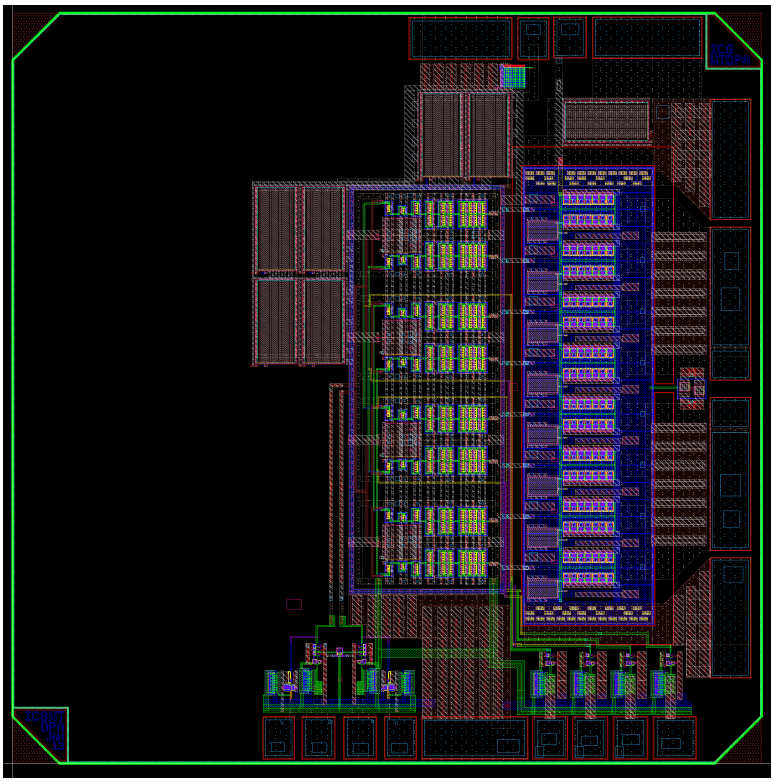

Figure 4.21: Complete chip layout

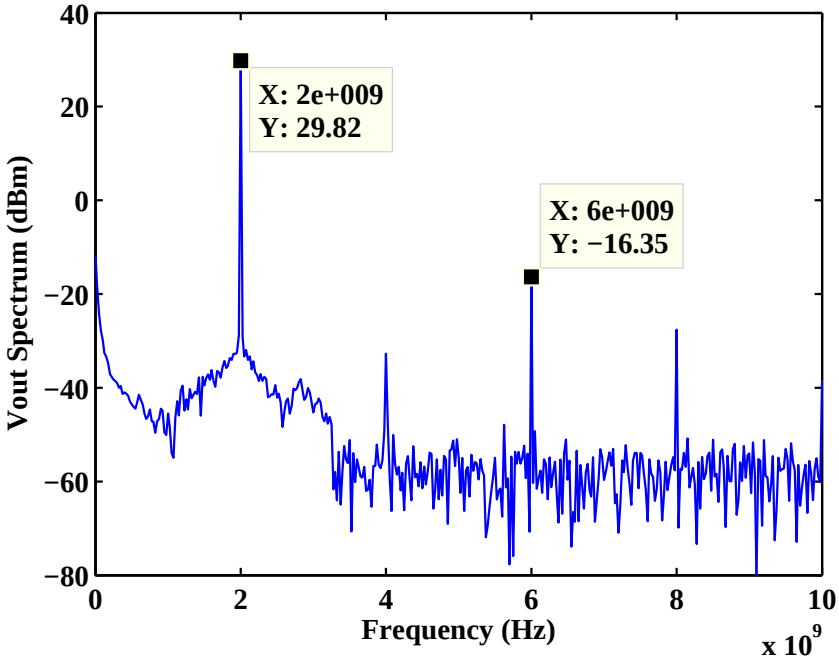

Figure 4.22: Output spectrum (FFT) at saturated output power

4.5 Results and Discussions

In this section, post layout simulation (PLS) results of the designed PA are presented. The output network, including the PCB matching and balun, was fully modelled in S-parameters obtained from Momentum. In order to characterize the PA thoroughly, the simulations were carried out in three stages. The first stage was to test the PA with a continuous wave. No digital encoding was applied to $PMW_1 - PMW_4$.

Figure 4.22 plots the output spectrum of the designed PA. The peak output power was 29.8 dBm at 2 GHz. The driver stage was biased from a 1.2 V supply and consumed 70 mA current. The PA stage was biased from a 2.5 V supply and consumed 594 mA current. The drain efficiency was 68 % and the power added efficiency (PAE) was 64 %. The third-order intermodulation distortion (IM3) was -46 dBc.

The saturated output power and PAE versus frequency are plotted in Figure 4.23. The designed PA had a 1-dB bandwidth from 1.5 GHz to 2.7 GHz. The output power and PAE was relatively stable within the band. The results indicated that the designed CMCD PA would be suitable to be deployed in multiband radios.

The second stage of verification involved digital encoding ($PWM_1 - PWM_4$ in Figure 4.9) in the PA. A continuous wave signal with varying envelope a and a constant phase ϕ was processed first in the baseband. The envelope a was sampled by the hybrid encoder (described in Chapter 3), which ran at a clock rate of 20 MHz and a minimal pulse width of 2.5 ns. Figure 4.24 plots the output power

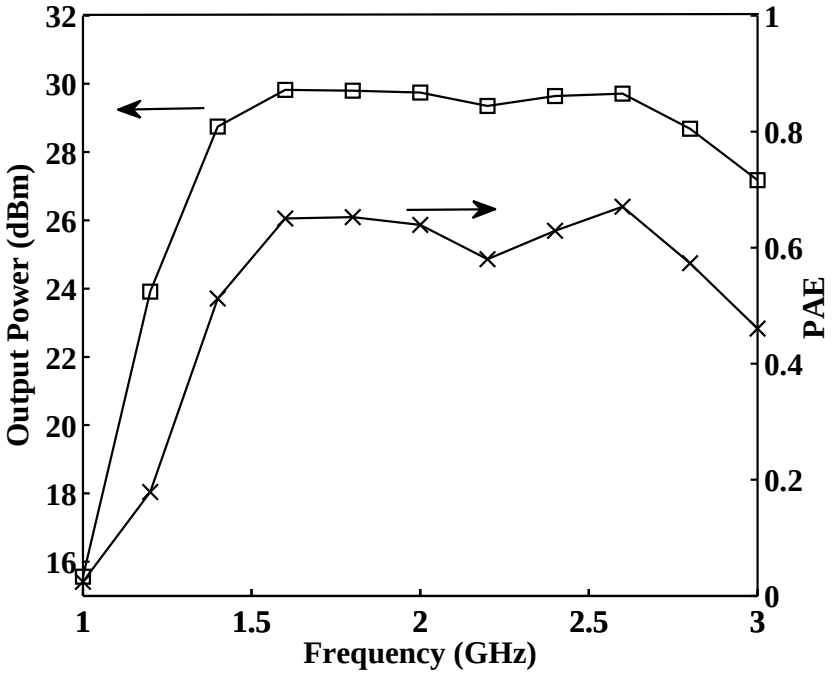

Figure 4.23: Saturated output power and efficiency versus frequency

at different input amplitudes. The maximum output power was 954 mW and the minimal output power was 0.5 mW, corresponding to a dynamic range of 33 dB. A conventional binary PWM encoding scheme, utilizing the same clock rate and minimal pulse width, was also implemented for a comparison. Using binary PWM, the PA had a minimal output power of 1.2 mW and a dynamic range of 29 dB.

PA efficiency at different input amplitude is plotted in Figure 4.25. The plot clearly indicated the efficiency enhancement achieved using hybrid encoding.

Although the proposed encoding scheme begins as fully linear, the non-linear response of the SMPA will alter the overall linearity when the two are combined. This is due to the non-linear switch resistance of the PA as well as imperfect current combining. PA distortion can be characterized in AM-AM and AM-PM. Figure 4.26.(a) plots the AM-AM response, which was fairly non-linear with a shape of compression at large amplitude. A $9th$-order polymonial function is implemented to correct the distortion. The pre-distorted AM-AM response is plotted in the same Figure and can be seen to have improved linearity. Similarly, the AM-PM plot is plotted in Figure 4.26.(b). Note the AM-PM distortion of the proposed PA using hybrid encoding was much smaller than the digital PA in [3], which employed a large array of PAs for amplitude only encoding.

The final stage of validation was to characterize the PA with modulated signals. The block diagram of the entire digital transmitter is shown in Figure 4.27. The I/Q generator in the baseband provided QAM64 modulated signals with 5 MHz bandwidth running at 4× OSR and 7.3 dB PAPR. The I/Q data was then converted into amplitude and phase in the CORDIC before being pre-distorted in the LUT.

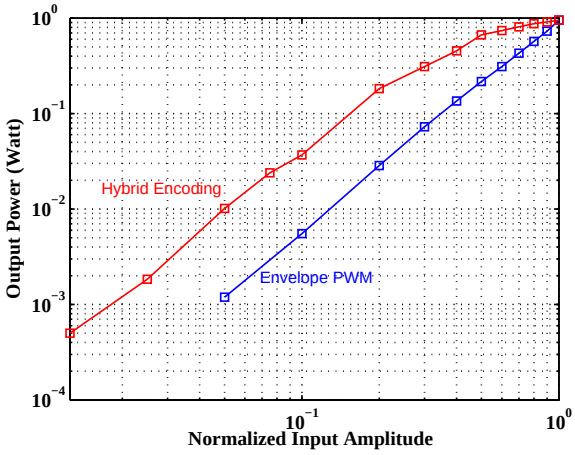

Figure 4.24: PA output power versus input amplitude

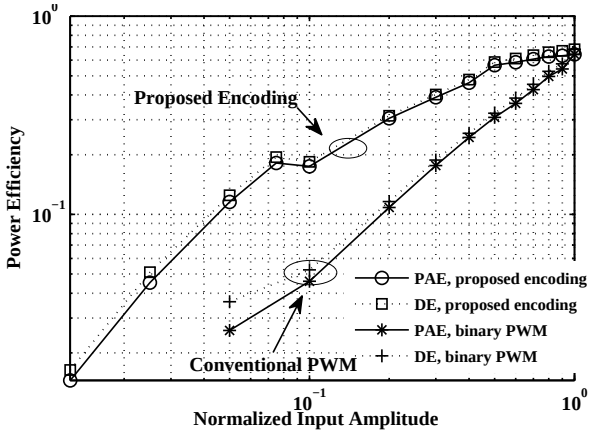

Figure 4.25: PA efficiency versus input amplitude

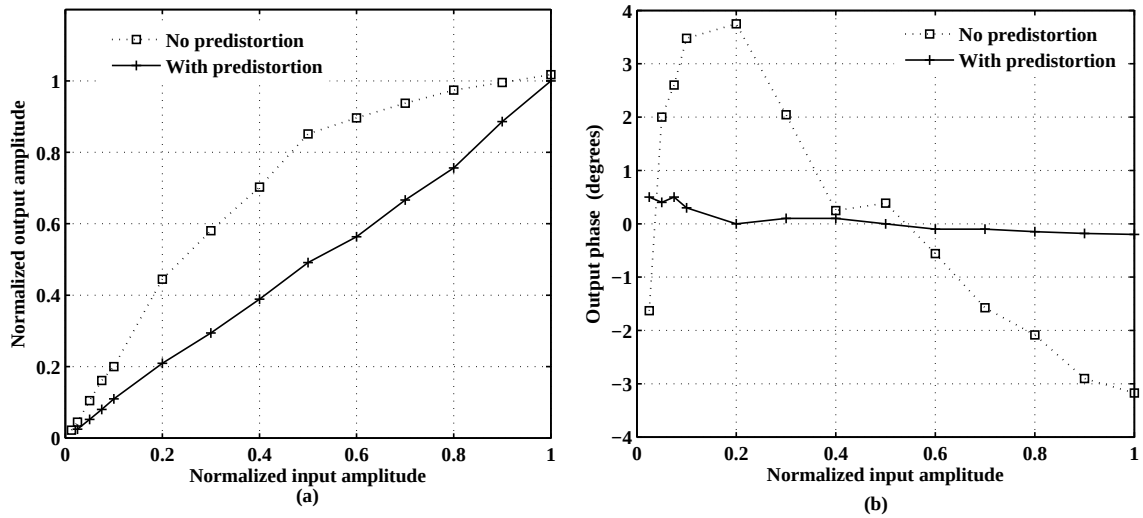

Figure 4.26: (a). AM-AM and (b). AM-PM with and without LUT pre-distortion.

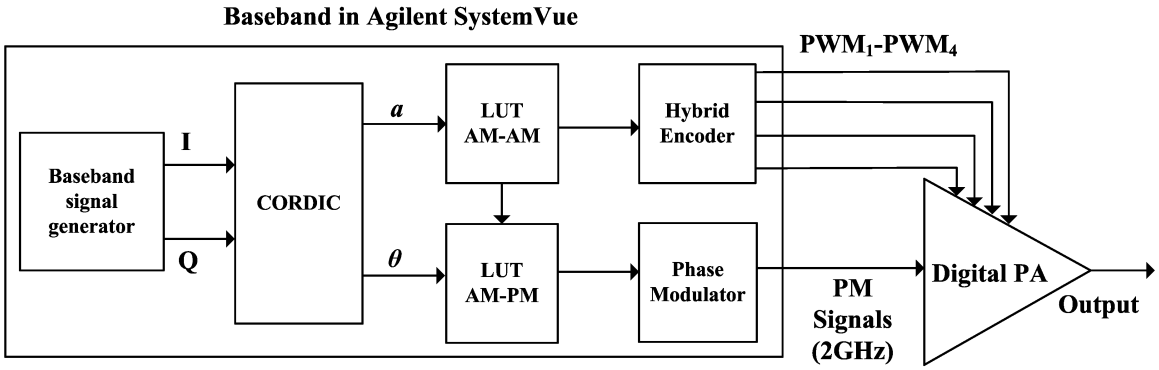

Figure 4.27: Simplified block diagrams of the proposed digital transmitter.

The pre-distorted amplitude signal was further processed in the hybrid encoder described previously. Output of the hybrid encoder, $PWM_1 - PWM_4$, are digital signals controlling the output power of the digital PA. Both the amplitude and phase path were designed with an equivalent bit resolution of 6.3 bits.

Output waveform of the design PA is plotted in Figure 4.28.(a). The envelope displays the features of our hybrid encoding, *i.e.*, variations of pulse width and amplitude levels. FFT spectrum of the digital transmitter output is plotted in Figure 4.28.(b). The spectrum plot indicates a good SNR, with an ACPR of -30.5 dB. No reconstruction filter has been used, hence the spurs can be observed near the carrier frequency. The ratio between the in-band and total output power is 86.7%, which is much higher than that of a conventional PWM or DSM [30]. The average efficiency was calculated by dividing the in-band power over total DC power including the driver stage and was equal to 27%. A comparison to other digital PA was made in Table 4.1. The table indicates that our PA is capable of achieving higher output power and efficiency. The improved performance was due to our synergistic design approach, which involved the system level modelling and analysis of both the encoding and SMPA.

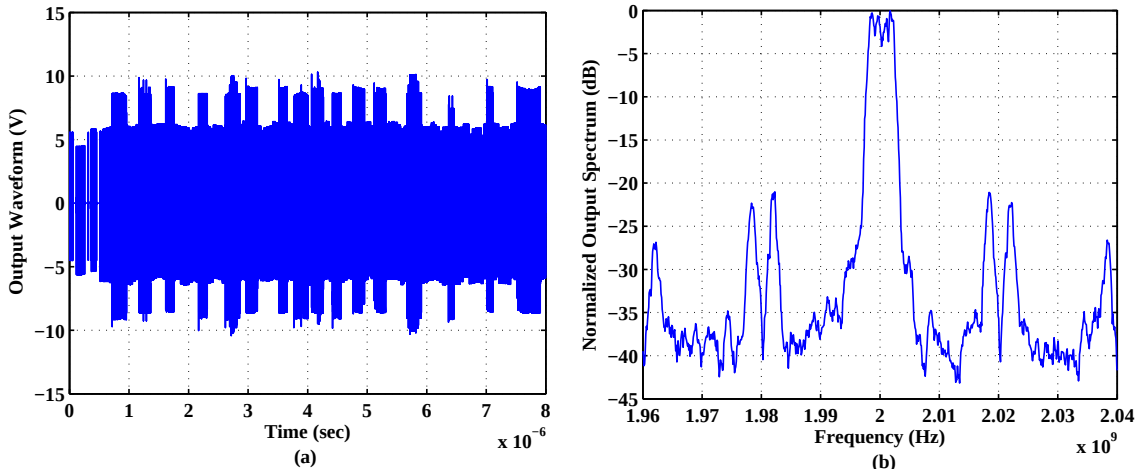

Figure 4.28: (a). PA output waveform and (b). FFT spectrum

Table 4.1: Comparison to Prior Art

Reference	CMOS Technology	Encoding Scheme	Centre Frequency (GHz)	Signal PAPR (dB)	Average Pout (dBm)	Average PAE
[5]	65nm	RF-PWM	2.2	1	27.5	25.3
[18]	130nm SOI	Amplitude	1.92	10	15.3	22
[2]	180nm	Amplitude	1.56	7.8	13.6	6.7
[3]	65nm	Amplitude	2.25	7.8	14	18
This work (PLS Results)	130nm	Hybrid Amplitude/Time	2	7.3	22.5	27

4.6 Conclusion

A multi-way current mode class-D digital PA that implements the hybrid amplitude/time encoding has been presented in this Chapter. The theoretical analysis identified major sources of power losses in order to suggest options to improve the efficiency. Circuit implementations of the PA in 130 nm CMOS were proposed. Post-layout simulations results validated the design methodology. The PA achieved a wide bandwidth (1-dB Pout) of operation from 1.5 GHz - 2.7 GHz. The peak PAE was 64% at 29.8 dBm output power with a 2.5 V supply. Modulated signals with 7.3 dB PAPR were applied to the PA. With simple static predistortion, an ACPR of -30.5 dB was achieved and the average PAE was 27%. The results demonstrated improved efficiency and dynamic range based on hybrid amplitude/time encoding as compared to conventional time-only or amplitude-only encoding.

Chapter 5

Conclusion and Future Work

The exponential growth of wireless technology and its related market has led to an intensified level of research and development in future radios that are energy efficient, low cost and, most importantly, re-configurable. Incorporating digital circuits and digital signal processing (DSP) into the analog and RF domain is a viable solution given CMOS technology's increasing speed, lower power consumption and cost. At present, the power amplifier has become the bottleneck in the digitalization of transmitters, due to low average power efficiency, poor compatibility with nanometer CMOS processes and limited re-configurability. This thesis investigated these key issues and presented the design and implementation of a multimode and multiband digital transmitter.

Among various digital transmitter architectures reviewed, we focussed on studies investigating the integration of SMPA: digitally modulated power amplifier. Design of a DMPA involves two aspects: the encoder and the SMPA. This thesis dealt first with encoder design. We proposed a theoretical framework to analyze the encoding design space for DMPAs and a new hybrid amplitude/time encoding scheme was devised. In addition, interleaved decomposition method was proposed and applied to transform the multi-level output signal into multiple binary PWM waveforms used to drive parallel SMPAs. Significantly improved spectral purity, dynamic range and coding efficiency were observed both analytically and experimentally. The encoder has demonstrated the ability to support different modulations, *e.g.* QAM and OFDM, and can be synthesized as fully digital, thus satisfying the multimode requirement for a digital transmitter.

In order to enable multiband capability, it is desirable to realize a high efficiency and wide bandwidth SMPA. We presented the design and implementation of a multi-way current mode class-D PA in the second half of this thesis. The theoretical analysis provided guidelines for understanding the fsources of power loss and their impacts on efficiency. The multi-way PA has been implemented in 130nm CMOS technology and the post-layout simulations results achieved a high PAE of 64% and high output power of 29.8 dBm. The wide bandwidth (1-dB Pout) from 1.5 - 2.7 GHz demonstrates the capacity for multiband operations. A complete digital transmitter combining the encoder and the multi-way PA was investigated at the end of this study and demonstrated an average efficiency of 27% and ACPR of -30.5

dBc using 7.3 dB PAPR modulated signals.

The results of this research suggest very promising directions in realizing a PA-included multimode and multiband digital radio and have also inspired the author to make some recommendations for future work:

1. One avenue of future work could be to implement an encoder using CMOS combined with the PA design outlined in Chapter 4, hence realizing a system-on-chip (SoC). The main challenge is in realizing a digital delay line with sufficient resolution and robustness against process and temperature variations. Recent work from [26] has demonstrated that such a delay line can be realized with a high speed CMOS process (65 nm or lower).

2. Although the encoding itself achieves significant improvement in the dynamic range and efficiency, the actual result when combined with SMPA is reduced due to variation of the switch resistance. Rather than direct current combining, the multi-way PAs could be combined in an isolated power combiner [38]. It is hypothesized that the efficiency, linearity and dynamic range could be further improved.

Bibliography

[1] M.S. Alavi, R.B. Staszewski, L.C.N. de Vreede, A. Visweswaran, and J.R. Long. All-digital rf i/q modulator. *IEEE Trans. Microw. Theory Tech.*, 60(11):3513 –3526, Nov. 2012. ix, 8, 9

[2] A. Kavousian, D. K. Su, M. Hekmat, A. Shirvani, and B. A. Wooley. A digitally modulated polar CMOS power amplifier with a 20-MHz channel bandwidth. *IEEE J. Solid-State Circuits*, 43(10):2251–2258, Oct. 2008. ix, 1, 10, 11, 59

[3] D. Chowdhury, Ye Lu, E. Alon, and A. M. Niknejad. An efficient mixed-signal 2.4-GHz polar power amplifier in 65-nm CMOS technology. *IEEE J. Solid-State Circuits*, 46(8):1796–1809, Aug. 2011. ix, 10, 11, 12, 36, 38, 43, 56, 59

[4] C. Berland, I. Hibon, J.F. Bercher, M. Villegas, D. Belot, D. Pache, and V. Le Goascoz. A transmitter architecture for nonconstant envelope modulation. *IEEE Trans. Circuits Syst. II, Exp. Briefs*, 53(1):13–17, 2006. ix, 13, 21, 30

[5] J. S. Walling, H. Lakdawala, Y. Palaskas, A. Ravi, O. Degani, K. Soumyanath, and D. J. Allstot. A class-E PA with pulse-width and pulse-position modulation in 65 nm CMOS. *IEEE J. Solid-State Circuits*, 44(6):1668–1678, Jun. 2009. ix, 14, 15, 30, 38, 48, 52, 59

[6] B. Baker. How delta-sigma adcs work, part 1. *Analog Applications Journal*, 3(1):13–16, 2011. ix, 16

[7] T.P. Hung, J. Rode, L.E. Larson, and P.M. Asbeck. Design of h-bridge class-d power amplifiers for digital pulse modulation transmitters. *IEEE Trans. Microw. Theory Tech.*, 55(12):2845 –2855, Dec. 2007. ix, 15, 16, 36

[8] J.S. Choi, J.H. Yim, J.H. Yang, J.O. Kim, J.H. Cha, D.H. Kang, D.S. Kim, and B.M. Kim. A $\delta\sigma$-digitized polar RF transmitter. *IEEE Trans. Microw. Theory Tech.*, 55(12):2679–2690, Dec. 2007. ix, 17

[9] IBM Microelectronics. Design kit and technology training CMOS8RF (CMRF8SF) V1700 IBM. 2010. x, 51

[10] V. Giannini, P. Nuzzo, C. Soens, K. Vengattaramane, M. Steyaert, J. Ryckaert, M. Goffioul, B. Debaillie, J. Van Driessche, J. Craninckx, and M. Ingels. A 2mm^2 0.1-to-5GHz SDR receiver in 45nm digital CMOS. In *Proc. IEEE Int. Solid-State Circuits Conf.*, pages 408–409, Feb. 2009. 1

[11] S. Balasubramanian, S. Boumaiza, H. Sarbishaei, T. Quach, P. Orlando, J. Volakis, G. Creech, J. Wilson, and W. Khalil. Ultimate transmission. *IEEE Microw. Mag.*, 13(1):64–82, Jan. 2012. 2

[12] J.H. Moon, J.H. Son, J.Y. Lee, and B.M. Kim. A Multimode/Multiband envelope tracking transmitter with broadband saturated amplifier. *IEEE Trans. Microw. Theory Tech.*, 59(12):3463–3473, Dec. 2011. 2

[13] P. Eloranta, P. Seppinen, and A. Parssinen. Direct-digital rf-modulator: a multi-function architecture for a system-independent radio transmitter. *IEEE Commun. Mag.*, 46(4):144 –151, Apr. 2008. 4

[14] P. Eloranta, P. Seppinen, S. Kallioinen, T. Saarela, and A. Parssinen. A wcdma transmitter in 0.13 μm cmos using direct-digital rf modulator. In *Proc. IEEE Int. Solid-State Circuits Conf.*, pages 340 –607, Feb. 2007. 4

[15] A. Jerng and C.G. Sodini. A wideband $\Delta\Sigma$ digital-rf modulator for high data rate transmitters. *IEEE J. Solid-State Circuits*, 42(8):1710 –1722, Aug. 2007. 5

[16] V. K. Parikh, P. T. Balsara, and O. E. Eliezer. All digital-quadrature-modulator based wideband wireless transmitters. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 56(11):2487–2497, 2009. 9

[17] R. B. Staszewski, J. L. Wallberg, S. Rezeq, Hung Chih-Ming, O. E. Eliezer, S. K. Vemulapalli, C. Fernando, K. Maggio, R. Staszewski, N. Barton, Lee Meng-Chang, P. Cruise, M. Entezari, K. Muhammad, and D. Leipold. All-digital PLL and transmitter for mobile phones. *IEEE J. of Solid-State Circuits*, 40(12):2469–2482, Dec. 2005. 10

[18] C.D. Presti, F. Carrara, A. Scuderi, P.M. Asbeck, and G. Palmisano. A 25 dBm digitally modulated CMOS power amplifier for WCDMA/EDGE/OFDM with adaptive digital predistortion and efficient power control. *IEEE J. of Solid-State Circuits*, 44(7):1883–1896, 2009. 10, 59

[19] P. Min, M. H. Perrott, and R. B. Staszewski. An amplitude resolution improvement of an RF-DAC employing pulsewidth modulation. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 58(11):2590–2603, Nov. 2011. 10

[20] H.J. Kim, Y.S. Jeon, J.J. Lee, S.H. Lee, and S.W. Nam. A high-efficiency power amplifier using multilevel digital pulse modulation. *Microw. Opt. Technol. Lett.*, 51(8):1921–1924, Aug. 2009. 13, 17, 22, 23

[21] S.L. Chi, C. Vogel, and P. Singerl. The frequency spectrum of polar modulated PWM signals and the image problem. In *Proc. IEEE Int. Conf. Electronics, Circuits, and Systems (ICECS)*, pages 679–682, 2010. 13

[22] Jang S.Y., Ahn K.P., Choi Y.H., Ryu N.S., Park B.Y., Hyun S.B., and Jung J.H. PWM based CMOS supply modulator for LTE envelope tracking transmitter. In *Int. Conf. ICT Convergence*, pages 622–623. 13

[23] S.L. Chi, P. Singerl, and C. Vogel. Coding efficiency optimization for multilevel PWM based switched-mode RF transmitters. In *Proc. IEEE Int. Midwest Symp. Circuits and Systems (MWSCAS)*, pages 1–4, Jul. 2011. 13, 20, 22

[24] K. Hausmair, S.L. Chi, P. Singerl, and C. Vogel. Aliasing-free digital pulse-width modulation for burst-mode RF transmitters. *IEEE Trans. Circuits Syst. I, Reg. Papers*, accepted, 2013. 13, 21

[25] H. Kobayashi, J.M. Hinrichs, and P.M. Asbeck. Current-mode class-d power amplifiers for high-efficiency rf applications. *IEEE Trans. Microw. Theory Tech.*, 49(12):2480 –2485, Dec. 2001. 13, 36, 39

[26] P. A. J. Nuyts, P. Singerl, F. Dielacher, P. Reynaert, and W. Dehaene. A fully digital delay line based GHz range multimode transmitter front-end in 65-nm CMOS. *IEEE J. of Solid-State Circuits*, 47(7):1681–1692, 2012. 13, 61

[27] F. Raab. Radio frequency pulsewidth modulation. *IEEE Trans. Commun.*, 21(8):958 – 966, Aug. 1973. 13

[28] P. Madoglio, A. Ravi, Hongtao Xu, K. Chandrashekar, M. Verhelst, S. Pellerano, L. Cuellar, M. Aguirre, M. Sajadieh, O. Degani, H. Lakdawala, and Y. Palaskas. A 20dbm 2.4GHz digital outphasing transmitter for wlan application in 32nm cmos. In *Proc. IEEE Int. Solid-State Circuits Conf.*, pages 168–170, Feb. 2012. 15

[29] K. Hausmair, Shuli Chi, P. Singerl, and C. Vogel. Aliasing-free digital pulse-width modulation for burst-mode rf transmitters. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 60(2):415–427, Feb. 2012. 15

[30] F. M. Ghannouchi, S. Hatami, P. Aflaki, M. Helaoui, and R. Negra. Accurate power efficiency estimation of GHz wireless delta-sigma transmitters for different classes of switching mode power amplifiers. *IEEE Trans. Microw. Theory Tech.*, 58(11):2812–2819, 2010. 17, 58

[31] J.H. Chen, H.S. Yang, and Y. J. E. Chen. A multi-level pulse modulated polar transmitter using digital pulse-width modulation. *IEEE Microw. Wireless Compon. Lett.*, 20(5):295–297, May 2010. 17, 22, 23

[32] T. Podsiadlik, J. Dooley, A. Canniff, and R. Farrell. Pulse width modulation of multilevel delta-sigma output for class s power amplifier. In *Proc. European Conf. Circuit Theory and Design*, pages 457–460, Aug. 2009. 17

[33] J.H. Chen, H.S. Yang, H.C. Lin, and Y. J. E. Chen. A polar-transmitter architecture using multiphase pulsewidth modulation. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 58(2):244–252, 2011. 27

[34] P. K. Hanumolu, V. Kratyuk, Wei Gu-Yeon, and Moon Un-Ku. A sub-picosecond resolution 0.5-1.5 GHz digital-to-phase converter. *IEEE J. Solid-State Circuits*, 43(2):414–424, 2008. 33

[35] S.C. Cripps. *RF Power Amplifiers for Wireless Communications, Second Edition (Artech House Microwave Library (Hardcover))*. Artech House, Inc., Norwood, MA, USA, 2006. 37

[36] N.O. Sokal and A.D. Sokal. Class e-a new class of high-efficiency tuned single-ended switching power amplifiers. *IEEE J. Solid-State Circuits*, 10(3):168 – 176, Jun. 1975. 37

[37] A. Mazzanti, L. Larcher, R. Brama, and F. Svelto. Analysis of reliability and power efficiency in cascode class-e pas. *IEEE J. Solid-State Circuits*, 41(5):1222 – 1229, May. 2006. 38

[38] G. Liu, P. Haldi, T.J. Liu, and A.M. Niknejad. Fully integrated cmos power amplifier with efficiency enhancement at power back-off. *IEEE J. Solid-State Circuits*, 43(3):600–609, Mar. 2007. 61

Appendix A

Proof of Optimal Coding Efficiency

In the encoding design space, assume l slots are filled

$$l = \sum_{j=1}^{K-1} \sum_{i=1}^{N-1} f_{i,j} \qquad l \in [0, (N-1)(K-1)] \tag{A.1}$$

where $f_{i,j}$ is logic state of the i-th row, j-th column slot. $f_{i,j} = 1$ indicates that a slot has been filled and $f_{i,j} = 0$ indicates an empty slot. The in-band power is

$$P_{in} = \left[\frac{1}{T_p(N-1)(K-1)} \sum_{j=1}^{K-1} \sum_{i=1}^{N-1} f_{i,j} \right]^2 = \left(\frac{\delta}{T_p} l \right)^2 \tag{A.2}$$

where δ is one minimal output step given by

$$\delta = \frac{1}{(N-1)(K-1)} \tag{A.3}$$

We notice that P_{in} is only a function of the number of filled slots and does not depend on the fill sequence. The total power, P_{total}, is given by

$$\begin{aligned} P_{total} &= \frac{\delta}{T_p} \sum_{j=1}^{K-1} \left[\sum_{i=1}^{N-1} f_{i,j} \right]^2 \\ &= \frac{\delta}{T_p} \sum_{j=1}^{K-1} \left[\sum_{i=1}^{N-1} f_{i,j}^2 + 2 \sum_{i=1,q>i}^{N-1} f_{i,j} f_{q,j} \right] \\ &= \frac{\delta}{T_p} \left[\sum_{j=1}^{K-1} \sum_{i=1}^{N-1} f_{i,j}^2 + 2 \sum_{j=1}^{K-1} \sum_{i=1,q>i}^{N-1} f_{i,j} f_{q,j} \right] \\ &= \frac{\delta}{T_p} \left[l + 2 \sum_{j=1}^{K-1} \sum_{i=1,q>i}^{N-1} f_{i,j} f_{q,j} \right] \end{aligned} \tag{A.4}$$

Minimizing P_{total} is equivalent to

$$min \sum_{j=1}^{K-1} \sum_{i=1,q>i}^{N-1} f_{i,j} f_{q,j} \quad subject\ to \sum_{j=1}^{K-1} \sum_{i=1}^{N-1} f_{i,j} = l \tag{A.5}$$

The condition is satisfied by minimizing the occurrences of

$$f_{i,j} \times f_{q,j} = 1 \tag{A.6}$$

Eq. (25) indicates that minimal total power (equivalently the highest coding efficiency) is reached by avoiding filling slots at the same j-th column in the design space.