DESIGN OF POWER AMPLIFIER (PA) FRONT-END TRANSMITTER FOR CDMA APPLICATION AT 900MHz USING 0.18 MICRON CMOS TECHNOLOGY

Oleh

Muhamad Iskandar Bin Zahari

Disertasi ini dikemukakan kepada
UNIVERSITI SAINS MALAYSIA

Sebagai memenuhi sebahagian daripada syarat keperluan
untuk ijazah dengan kepujian

SARJANA MUDA KEJURUTERAAN (KEJURUTERAAN ELEKTRONIK)

Pusat Pengajian Kejuruteraan
Elektrik dan Elektronik
Universiti Sains Malaysia **Mac 2005**

ABSTRACT

Recent efforts in the design of integrated circuits for RF communication transceivers have focused on achieving higher levels of integration by including more and more analog functional blocks onto a single silicon CMOS chip. One of the final blocks that have yet to be successfully integrated is the power amplifier. The power amplifier is the final functional block in the transmit path and its function is to amplify the signal to be transmitted to the required transmit power level. In general, power amplifiers are difficult to integrate in CMOS because of technology limitations that severely limit the efficiency of the power amplifier. This thesis describes theoretical analysis and circuit techniques for the design and implementation of RF Class C power amplifier in CMOS technologies. There are very few methods exist for designing Class C power amplifier in the past, much of the design process has been empirical. The theoretical work in this thesis attempts to describe a method for designing a Class C power amplifier in CMOS without resorting to blind use of a circuit simulator. A 900MHz CMOS power amplifier was designed using Silterra 0.18μm RF MOSFET. This design is simulated using Cadence Design tool. In this simulation, the peak efficiency of the power amplifier was 37.7%. The power amplifier did meet the spectral mask requirements of CDMA (Code Division Multiple Access) cellular communications system for which it was designed. The power gain that had been achieved is 34.74 dB and meets the power amplifier specification for CDMA application.

ABSTRAK

Pada masa kini, usaha untuk merekabentuk litar bersepadu untuk pemancar frekuensi radio telah diberi perhatian untuk mencapai peringkat yang lebih tinggi dengan menggunakan blok litar analog CMOS. Blok yang terakhir sebelum isyarat dapat dipancarkan adalah blok penguat kuasa. Blok penguat kuasa berfungsi untuk menguatkan isyarat untuk dipancarkan pada peringkat yang dikehendaki. Blok penguat kuasa adalah sukar untuk bersepadu atau direkabentuk dengan menggunakan teknologi CMOS kerana ia akan menghadkan kecekapan bagi blok penguat kuasa. Disertasi ini menerangkan analisis teori dan teknik rekabentuk litar blok penguat kelas C menggunakan teknologi CMOS. Terdapat pelbagai teknik terdahulu yang telah digunakan untuk merekabentuk blok penguat kuasa tetapi kebanyakan rekabentuk tersebut tidak dapat diaplikasikan dengan sempurna. Blok penguat kuasa 900MHz CMOS telah direkabentuk dengan menggunakan peranti yang menggunakan teknologi dari Silterra 0.18µ. Rekabentuk litar bersepadu ini telah disimulasi dengan menggunakan perisian “Cadence”. Keputusan daripada simulasi yang telah dijalankan mendapati kecekapan yang diperolehi adalah sebanyak 37.65%. Daripada keputusan menyeluruh yang diperolehi blok penguat kuasa ini telah menepati keperluan untuk aplikasi CDMA(Code Division Multiple Access). Gandaan kuasa yang diperolehi daripada rekabentuk ialah 34.74 dB dan telah mencapai spesifikasi bagi sebuah penguat kuasa untuk aplikasi CDMA.

ACKNOWLEDGMENT

First and foremost, I would like to thank my advisor, Dr. Tun Zainal Azni Zulkifli, for his support on my thesis progression. His advice on technical matters was invaluable, and very critical to the project's success. I would also like to thank En. Zulfiqar Ali Abdul Aziz , Puan Norlaili Mohd Noh and En. Basir Bin Saibon for their assistance along the way, as their knowledge and experience were also extremely valuable. I would also like the thank Mr. Harikrishnan Ramiah and Mr. K. Selvarajah K.Krishnarajoo for helping me to get through with this project.

Furthermore, I would like to thank again to Dr.Tun Zainal Azni Zulkifli and my second advisor En. Basir Bin Saibon for reading this dissertation. More generally, my colleagues in Dr. Tun Zainal Azni Zulkifli research group were extremely helpful as well. Several graduate students in other research groups were also very helpful along the way. Noor Fadhili Mohd Fadzil and Mohammed Irfan Mustafa deserve great thanks, for being great friends as well as helping technically.

Finally, it is absolutely true that none of this thesis would have been possible without the love and support of my parents especially my late father ,Professor Dr.Zahari Bin Md. Darus, my mother, Wahidah Binti Zainal, my sister, Siti Nur Aini, my brother, Muhamad Azlan and last but not least, my beloved Maisarah Binti Ahmad Ramli. Their love and overwhelming support in the face of any obstacle or adversity I faced was not on instrumental but essential in the completion of this work. I would not be where I am today without their contributions and the foundation that they have provided throughout the years. All of your cooperation and support will be appreciated and always be remembered. Thank you very much.

TABLE OF CONTENT

CHAPTER 4: CLASS C THEORETICAL ANALYSIS

CHAPTER 5: POWER AMPLIFIER IMPLEMENTATION

CHAPTER 6: SIMULATION STRATEGY

CHAPTER 7: RESULTS AND DISCUSSION

LIST OF FIGURE

CHAPTER 1 – INTRODUCTION

In recent years, the level of interest in wireless communication links and the devices which can support those links has exploded. New standards are being both approved and designed in order to tap into the exploding market; many of these new standards attempt to connect devices and appliances in the home using lower performance radio transceivers. In addition, the desire for internet connectivity using cellular phones or Personal Digital Assistants (PDA) has dramatically increased the demand for universal wireless connectivity. In order for these standards and the companies that support them to stay competitive, low-cost, small form-factor portable wireless communications devices are a critical component. In order to attract a large group of users, cost is a very important consideration. As such, the desire for a low-cost wireless device with reasonable performance has exploded as well.

1.1 Background of Wireless Communication.

Current implementations of wireless communications devices, such as cellular phones or cordless phones, employ several chips implemented in different semiconductor technologies in order to implement the analog radio-frequency (RF) components and lower frequency components as well as the digital components. Several of these chips are normally designed in different fabrication and device technologies, where each of the chips is designed in a technology which is best suited for the particular function being implemented. This multi-chip solution to implementing these portable wireless devices limits the minimum cost and size of the final devices.

Further adding to the cost and size is the fact that passing signals between chips is generally more complex than passing signals around on a single-chip, especially at radio frequencies. Issues of matching and driving signals between chips add to the number of discrete components required and also increase the power consumption of the overall implementation, reducing the performance of the overall device.

To that end, the amount of investigation into the feasibility of using CMOS processes in the implementation of the circuitry used in these devices has dramatically increased as well. While CMOS processes are already used for many, if not all, of the lower frequency and digital functions, their acceptance as a reliable and feasible alternative on which to build the high-frequency high-performance analog blocks has been limited. In general, CMOS is an extremely poor technology for high-frequency, high-performance analog functions. For reasons that will be detailed in this work, and have been detailed in others, CMOS is much better suited to low-frequency analog implementations or digital functions. A CMOS transistor models is extremely useful for digital functions or switched-capacitor analog functions, but it in general has poor current drive and large associated parasitic capacitances, reducing its usefulness in higher frequency analog functions.

However, what CMOS does provide is an extremely low-cost technology in which to implement circuit blocks, along with the added benefit of potential large-scale integration. Because the base-band analog and digital chips are generally CMOS chips, building the high-frequency, high-performance chips in CMOS would allow those functions to be included in comprehensive, single-chip solutions. Currently, most commercially-available wireless solutions are multi-chip implementations, in which several different technologies, including silicon CMOS (Si-CMOS), silicon bipolar, Gallium Arsenide (GaAs), and Silicon Germanium (SiGe). In general, the higher frequency blocks are done in one of the listed technologies, with the exception of CMOS. A CMOS implementation of the high-frequency analog section would facilitate the integration of all of the radio functions (including high-frequency analog, base-band analog, A/D conversion, and DSP) onto a single-chip.

1.2 Transceiver Architecture.

In order to accomplish this, there must be an understanding of what functions must be implemented in the high-frequency analog portion of the RF transceiver, or radio. Figure1.1 show the basic architecture of a radio transceiver for CDMA application,

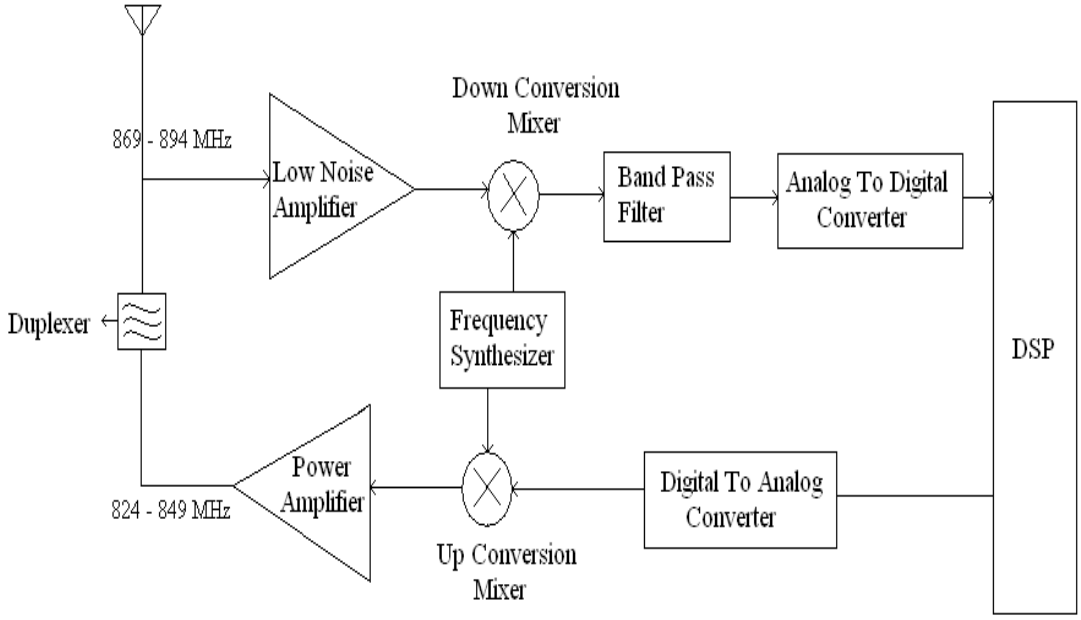


Figure 1.1 - Transceiver Architecture

which contains two main functions that is receiver and transmitter. The function of the analog portion of the receive chain is to take a radio-frequency (RF) analog signal and convert that into the digital bit stream which it represents. The basic blocks in the receive path are a low-noise amplifier that is critical to amplify the incoming signal without contributing any significant noise, a frequency translation block that is also known as a down conversion mixer and a block which demodulates the signal and outputs the desired digital bits that is call analog to digital converter. A block that is required but which is not directly in the signal path is the frequency synthesizer or local oscillator (LO), which is used by the mixer to frequency translate the signal to low-frequency.

On the transmit side, the digital bit stream is modulated and converted to an analog signal using digital to analog converter. The resulting analog signal is frequency translated by an up-conversion mixer, and the final signal is amplified to the power level that the antenna needs to radiate, as required by the standard. This final block is generally known as a power amplifier.

1.3 Power Amplifiers.

In RF transceivers, the power amplifier takes a small-amplitude signal from the up-conversion mixer as its input and drives a high power representation of the input into a low impedance load (generally the antenna, which is nominally 50Ω). Because the peak power levels required will be significantly higher than the mixer can supply on its own, a separate functional block is required. So the key purpose of the block is high frequency amplification as mentioned earlier where CMOS is not ideally suited to this function. While many of the other functional blocks have been implemented in CMOS reliably, the power amplifier is one of the final blocks to be implemented and integrated with the other blocks onto a single CMOS chip.

One of the key reasons for the general industry-wide reluctance to move the implementation of the power amplifier to CMOS is the fact that while the power amplifier is on, it can dominate the power consumption of the entire transceiver. Since most of these portable wireless devices will be powered by a battery with a finite energy reserve, the amount of power is consumed by any given block can be a critical factor in determining the usability of a particular part. As a result, optimizing the power-performance of the power amplifier has been a far more important goal than that of integrating the power amplifier into a transceiver. While the CMOS realization of the power amplifier might reduce the cost itself, if it significantly reduces the battery-life of the wireless device, there is really no gain (and there might be a potential loss) in that implementation.

Although the implementation of the power amplifier in CMOS has not reached reliable production levels yet, there is a great deal of ongoing research into making the CMOS power amplifier a realizable goal. This is because the cost and form-factor benefits from a high performance CMOS power amplifier are dramatic. If one can avoid having to use a GaAs or SiGe power amplifier in favor of a CMOS power amplifier, there is an inherent cost reduction in that switch. Second, the potential for including power amplifier in a single-chip CMOS transceiver, it would serve as the entire analog *and* digital processing unit for a cellular phone would dramatically reduce the cost and form factor of the PA, allowing for the placing of cellular phone transceivers in almost anything, including something as small as a watch.

1.4 Research Goals.

The focus of this research, and this dissertation, is the realization of a high performance, integrated power amplifier in CMOS. Through the use of power amplifier architecture, or class, as well as circuit design techniques, this work will try to show the feasibility of a CMOS power amplifier implementation, as well as its usefulness in an integrated transceiver. More generally, a key goal of this research is the development of a design methodology for the design of Class-C power amplifier in CMOS technologies. The key for a result of this research are listed here:

- A 900MHz power amplifier has been designed in this work in a 0.18μm CMOS process, using Class-C architecture. Experimental results indicated that the power amplifier actually delivered 630mW of output power to the load·

CHAPTER 2 – CDMA (Code Division Multiple Access)

Cellular services are now being used every day by millions of people worldwide. The number of customers requiring such services is increasing exponentially, and there is a demand for integration of a variety of multimedia services. The range of services includes short messaging, voice, data, and video. Consequently, the bit rate required for the services varies widely from just 1.2 kbps for paging up to several Mbps for video transmission.

2.1 Review

The CDMA is a digital modulation and radio access system that employs signature codes (rather than time slots or frequency bands) to arrange simultaneous and continuous access to a radio network by multiple users. Contribution to the radio channel interference in mobile communications arises from multiple user access, multi-path radio propagation, adjacent channel radiation and radio jamming.

The spread spectrum system's performance is relatively immune to radio interference. Cell sector and voice activity used in CDMA radio schemes provide additional capacity compared to FDMA and TDMA. However, CDMA still has a few drawbacks, the main one being that capacity (number of active users at any instant of time) is limited by the access interference. Furthermore, near-far effect requires an accurate and fast power control scheme. The first cellular CDMA radio system has been constructed in conformity with IS-95 specifications and is now known commercially as cdmaOne.

2.2 Description

CDMA is a wireless communications technology that uses the principle of spread spectrum communication. The intent of CDMA technology is to provide an alternative method for increasing bandwidth in limited frequency systems. Another advantage to CDMA is extended range and more secure communications. With CDMA, a narrowband message signal is multiplied by a spreading signal, or PN-code, which is a pseudo-noise code sequence that has a rate much greater than the data rate of the message. CDMA uses these code sequences as a means of distinguishing between individual conversations.

All users in the CDMA system use the same carrier frequency and may transmit simultaneously. A standard CDMA call starts out at 9600 bits per second (bps). After applying the PN-code, the rate of the signal jumps to around 1.25 kbps. This rate is shared by all the other users on that cell as well. As the signal is received, the PN-codes are removed from the desired signal which separates the users and returns the call to a rate of 9600 bps.

2.3 Media

CDMA is enabling new products and services, from palm-size phones with internet capabilities to satellite communications. The most prevalent application for CDMA is cellular phones. With CDMA, you can leave your phone on due to its low power consumption. The technology allows this by using a power control to monitor the amount of power your system and handset need at any time. CDMA handsets typically transmit at the lowest power levels in the industry, allowing for longer battery life which results in longer talk time and standby time. CDMA handsets can also incorporate smaller batteries, resulting in smaller and light-weight phones.

2.4 CDMA and Competing Technologies

TDMA: Is a common multiple access technique employed in digital cellular systems. It divides conventional radio channels into time slots to obtain higher capacity and not using separate codes over one time slot as CDMA does. Its standards include North American Digital Cellular, GSM, and PDC. No other conversations can access an occupied TDMA channel until the channel is vacated.

FDMA: Divides radio channels into a range of radio frequencies and is used in the traditional analog cellular system. As with TDMA, only one subscriber is assigned to a channel at a time. Other conversations can access this channel only after the subscriber's call has terminated or after the original call is handed off to a different channel by the system. AMPS is included within the FDMA cellular standards.

W-CDMA: Oddly enough, this technology is not compatible with regular CDMA. The debate about cdma2000 and W-CDMA convergence has been based on the fact that these CDMA-based proposals have certain parameter definitions that present an opportunity for compromise. The most discussed and debated parameter is the system chip rate. W-CDMA uses a chip rate value of 4.096 Mbps. cdma2000 uses 3.6864 Mbps. W-CDMA supporters claim as much as a 10% capacity improvement over that of cdma2000.

2.5 Spread Spectrum

CDMA is a "spread spectrum" technology, which means that it spreads the information contained in a particular signal of interest over a much greater bandwidth than the original signal. A CDMA call starts with a standard rate of 9600 bits per second (9.6 kilobits per second). This is then spread to a transmitted rate of about 1.23 Megabits per second. Spreading means that digital codes are applied to the data bits associated with users in a cell. These data bits are transmitted along with the signals of all the other users in that cell.

Traditional uses of spread spectrum are in military operations. Because of the wide bandwidth of a spread spectrum signal, it is very difficult to jam, difficult to interfere with, and difficult to identify. This is in contrast to technologies using a narrower bandwidth of frequencies. Since a wideband spread spectrum signal is very hard to detect, it appears as nothing more than a slight rise in the "noise floor" or interference level. With other technologies, the power of the signal is concentrated in a narrower band, which makes it easier to detect. Increased privacy is inherent in CDMA technology. CDMA phone calls will be secure from the casual eavesdropper since, unlike an analog conversation, a simple radio receiver will not be able to pick individual digital conversations out of the overall RF radiation in a frequency band.

2.6 CDMA Benefits

When implemented in a cellular telephone system, CDMA technology offers numerous benefits to the cellular operators and their subscribers. The following is an overview of the benefits of CDMA.

- Capacity increases of 8 to 10 times that of an AMPS analog system and 4 to 5 times that of a GSM system.
- Improved call quality, with better and more consistent sound as compared to AMPS systems.
- Simplified system planning through the use of the same frequency in every sector of every cell.
- Enhanced privacy.
- Improved coverage characteristics, allowing for the possibility of fewer cell sites.
- Increased talk time for portables.

CHAPTER 3 – OVERVIEW OF POWER AMPLIFIER

In this chapter, a little background information will be provided, both on transmitters and on power amplifiers themselves. First, in order to be able to integrate a non-linear power amplifier in a transmitter that attempts to satisfy a cellular standard, a transmitter architecture that is amenable to power amplifier integration must be used. While the architecture advancements are not apart of the effort detailed in this dissertation (D.Yee, 2000), the ability to integrate a non-linear power amplifier is contingent on the use of an architecture such as the one described in the next section. Because it is a necessary condition to the use of this non-linear power amplifier, a basic explanation of the concepts is presented here.

The rest of the chapter will discuss power amplifier in further detail. A basic background on power amplifier and their function will first be given, followed by a brief introduction to the different classes of power amplifier. Finally, the final section will fully investigate what is known about Class-C power amplifier, and provide an idea of what else needs to be done to implement a Class-C power amplifier today.

3.1 Transmitter Basics

In today's world, the end goal for many research efforts in implementing low cost wireless transceivers is the single-chip radio. Currently, most two-way radios, especially those designed for high-performance wireless systems like cellular phone systems, are implemented with multi-chip solutions. The RF, IF, and base-band sections may be comprised of different chips in different technologies, and there may even be further subdivisions within those groups. In order to implement the single-chip radio, a single technology must be chosen, and architectures that enable the integration of all the different blocks onto a single chip must be used.

On the receiver side, there has already been significant research into such architectures, including well-known architectures such as direct-conversion. Newer architectures have also found significant interest, including Low-IF and Wideband- IF-with-Double-Conversion (WIFDC) (J.C. Rudell, 1997). On the transmitter side, the level of interest has lagged behind that of the receiver side. However, more and more work is being focused on the transmitter.

One of the key issues in transmitter integration is that significant amounts of filtering are generally needed in the signal path in order to ensure that the output of the power amplifier is free of both spurs and spectral re-growth that would violate either the spectral mask or the level of spurs which a given standard specifies. This is especially true if the power amplifier to be used in the transmit chain is a non-linear power amplifier, in which the level of distortion in the power amplifier input will generally be amplified. If two simple transmitter architectures are examined, it will be apparent that neither is particularly well-suited to full-scale integration. The first of these architectures

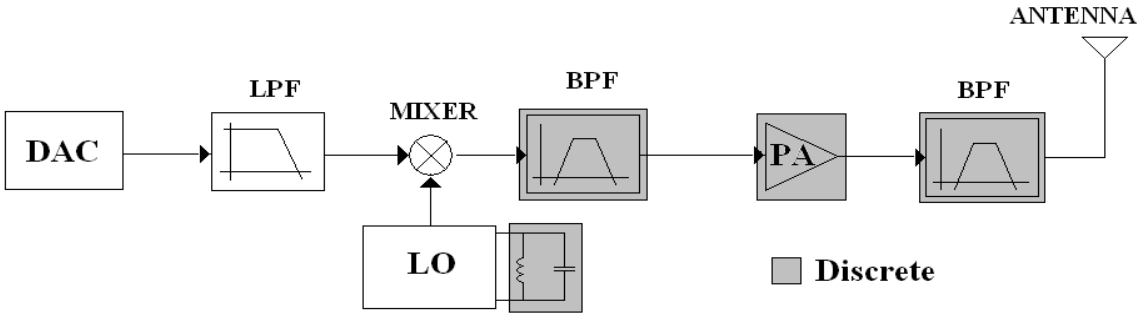


Figure 3.1 – Direct Conversion Transmitter

is the transmitter implementation of the direct-conversion architecture. In the direct-conversion transmitter, the base-band digital bits representing the in-phase and quadrate channels are converted to analog signals in a digital-to-analog converter (DAC), filtered, then up-converted directly to the RF carrier frequency in one frequency translation step. In general, at this stage, the signal needs to be filtered before passing through the power amplifier. This is for two reasons: first, the noise outside the transmit band must be

reduced, which reduces the filtering required after the power amplifier, and second, third-order inter-modulation between frequency components at the input of the power amplifier could mix and sit on top of the desired signal. This filter is difficult to integrate since it must have a narrow pass-band and a large amount of rejection, which cannot be achieved with the low quality-factor (Q) passive components available in a silicon CMOS technology.

Furthermore, the frequency synthesizer used to generate the LO frequency needs to be a tunable frequency synthesizer which can tune between a large number of steps which are several orders of magnitude below the center frequency. As a result, the frequency synthesizer will need a high quality-factor LC tank in order to generate a clean LO signal (i.e. one with very low phase-noise); like the filter, this high-Q tank will not be able to be integrated.

A second popular architecture is the transmitter version of the super-heterodyne receiver, which can be thought of as a transmitter using two steps to convert the frequency from base-band to the RF carrier frequency. In this architecture, once the digital bits are converted to analog signals in the DAC, the signal is frequency-translated

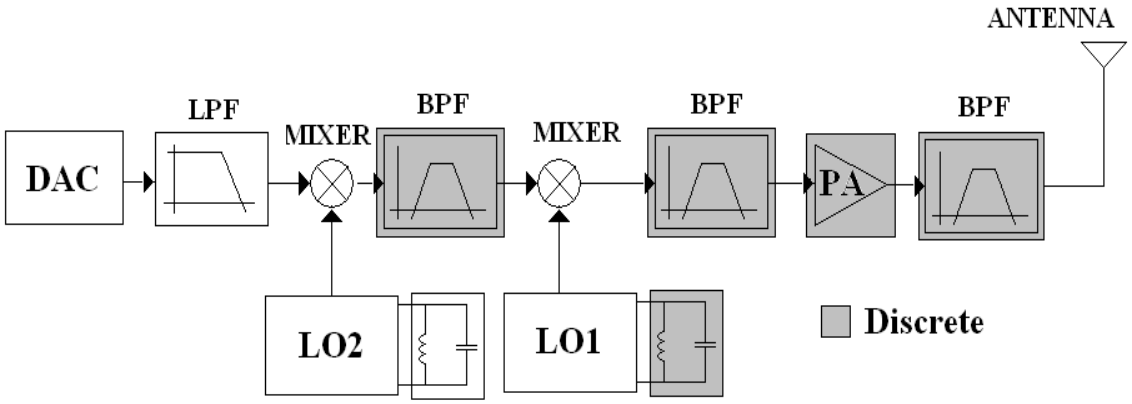


Figure 3.2 – Two Step Transmitter

to a fixed intermediate frequency (IF). There, the signal is filtered by a narrow-band filter, which serves the purpose of removing the harmonics of the lower frequency local

oscillator (LO2). The LO2 frequency is usually relatively low, when compared with the RF carrier frequency, and as a result, the harmonics of LO2 will cause inter-modulation distortion in both the mixers and the power amplifier which can give rise to significant errors in the transmitted signal. Because the LO2 frequency is generally low, the filter will have to have a narrow pass-band as well as lots of suppression in the stop-band, necessitating a high-Q filter. This IF filter is normally implemented as a discrete component. Once the signal has been filtered at IF, it is then frequency translated to the RF carrier frequency.

Again, the high frequency local oscillator used in this second mixing (LO1) is the one used to select the channel, and since it must be able to select between a numbers of closely-spaced channels, the LC tank required for this synthesizer will again likely be discrete. Finally, a pre-power amplifier filter is, like the direct-conversion case, required in the signal path in order to reject the noise and potential for inter-modulation. Finally, the signal is passed through the power amplifier. As in the direct-conversion case, the LC tank used in the frequency synthesizer and the pre-power amplifier filter need to be discrete; furthermore, in this super-heterodyne implementation, another discrete element, the IF filter, is added to the signal path. If anything, this architecture is even less amenable to integrating the entire signal path.

A third popular architecture is the transmitter version of the Harmonic Rejection

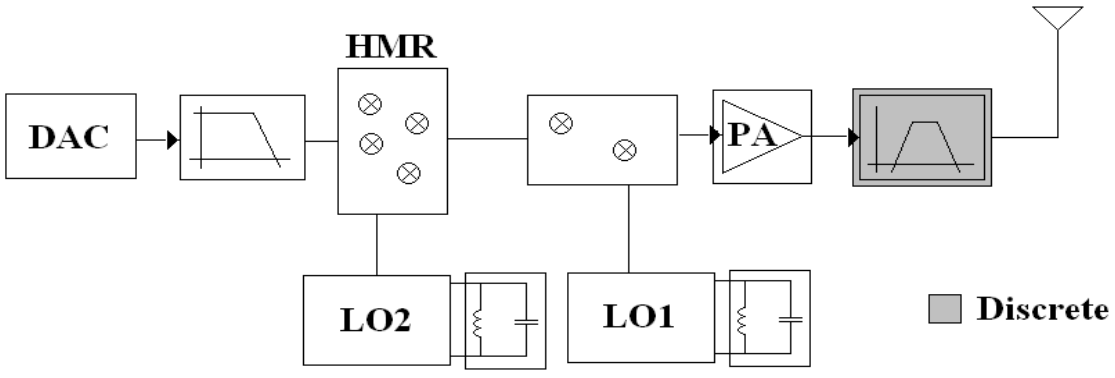


Figure 3.3 – Harmonic Rejection Transmitter

Transmitter (HRT) includes a special set of mixers known as Harmonic Rejection Mixers (HRM). The HRT architecture allows for the elimination of all of the discrete components required in the previously described architectures, except for the post-power amplifier RF filter. As a result, the entire signal path from the DAC to the output of the power amplifier can be included on a single chip.

The two key innovations in the transmitter include the use of the HRM and the roles of the frequency synthesizers used. The HRT performs the frequency translation from base-band to the RF carrier frequency in two steps. However, unlike the super-heterodyne transmitter mentioned previously, the roles of the local oscillators are swapped; the high-frequency LO (LO1) is nominally fixed in frequency, and the lower frequency LO (LO2) is the tunable, channel-select oscillator. The benefit of this arrangement is that it allows the use of a wide-bandwidth PLL in LO1, which can be designed to provide low phase noise even if a low-Q LC tank is used. Furthermore, because the channel-selection is performed at a lower frequency, the output of LO2 will inherently have lower phase noise, again allowing the use of low-Q on-chip components. Thus the need for the high-Q discrete components required in the VCO have been eliminated.

Furthermore, the use of the HRM helps to eliminate the need for the filters used in the signal path before the power amplifier. In standard active mixers (such as the Gilbert Cell-style mixers), the LO signal is applied to the mixers as a square wave. As is well-known, the square wave has large frequency components at the third and fifth harmonics of the fundamental frequency. In the super-heterodyne transmitter, the harmonics of the LO input mix with the base-band signal in the first mixing step. The IF filter is required in the super-heterodyne transmitter in order to prevent those harmonic components from producing distortion in the high-frequency mixers or the power amplifier. In the Harmonic-Rejection Mixers (HRM), however, the LO input is applied not as a square wave but as a staircase function. The staircase function can be thought of as a three-bit, amplitude-quantized sine wave (similarly, the square-wave can be thought of as a one bit

amplitude-quantized sine wave). This staircase function has no third or fifth harmonic component, and thus there will be no mixing products at those frequencies after the first mixing stage. If the LO2 frequency is low (on the order of 10 MHz or less), the seventh harmonic may still be problematic, as it would only be 70 MHz away, and may still produce inter-modulation distortion that mixes with the desired signal.

However, if the frequency plan is designed to use a higher LO2 frequency, the harmonic products existing after the first mixing stage would not exist at frequencies which may distort the output signal, and thus the need for the IF filter is eliminated. Furthermore, with the low phase-noise of LO1 and the removal of the harmonic components from the HRM, the pre-power amplifier filter is also not required, and thus the entire signal-path from beginning to end can be integrated if the Harmonic-Rejection Transmitter is used (J.A.Weldon, J.C.Rudell, L.Lin, R.S.Narayanaswani, 2001)(D.Yee, 2000). The HRT architecture facilitates the use of a non-linear power amplifier in the integrated transmitter, and furthermore facilitates the integration of the entire transmit path, from the DAC all the way through the power amplifier, including the frequency synthesizers used. In the next section, power amplifier will be discussed in significantly more detail, in order to provide a basis from upon which the remainder of this work will be built.

3.2 Power Amplifier Basic

Power amplifiers are used in the transmit chain of communications devices, in order to amplify the signal to the desired power level. That power level is determined by the communications system and it must be high enough such that the amount of power that the receiver is able to sense (taking into account the losses in the communication medium) is adequate to recover the desired signal. For different applications, the order of magnitude of the transmitted power can vary greatly. For base-stations used in cellular systems, this can be on the order of tens to hundreds of watts. For satellite communications, this can be on the order of thousands of watts. For portable wireless

communications devices, the peak transmitted power is often significantly less. It will vary from tens to hundreds of milliwatts in cordless systems to hundreds of milliwatts to a few watts in cellular systems.

Finally, in many of the emerging standards for wireless connectivity in the home (such as Bluetooth or CDMA), the power will vary between tens of milliwatts and hundreds of milliwatts. When speaking of the power output of these blocks, one common unit used is dBm, which is the output power in dB referenced to 1-mW. That is, the output power in dBm is given by:-

$$P_{dBm} = 10\log\frac{P}{0.001W} \tag{3.1}$$

where P is defined in watts.

In the cases where the power output is on the order of hundreds of milliwatts or more, the power that the PA needs to deliver to its load in itself is a large percentage of the total power consumed by the entire transmitter. The power that needs to be delivered will be taken from the source that powers the power amplifier. In essence, the power amplifier converts the DC power from the battery into RF power delivered to the load. Unless that power conversion is lossless, which is possible only as an ideal abstraction, the PA itself will consume power, over and above what it delivers.

The measure of how much power of power amplifier consumes in this conversion is one of the key performance parameters used to gauge power amplifier especially those are used for portable applications. Because power amplifier in portable applications are driven from a source with a finite amount of available energy, power consumed in the power amplifier directly goes to reducing the battery life. This metric is known as the power amplifier *efficiency*, given by:-

$$\eta = \text{efficiency} = \frac{P_{Load}}{P_{Supply}} \tag{3.2}$$

Since the power amplifier is really converting the DC power of the supply into the RF power delivered to the load, the maximum efficiency is 1, or 100%. That is, if there is no power consumed in the power amplifier, all the power from the supply is sent to the load. Both value of numerator and the denominator in equation 2-2 are the same. However, since this is only ideally possible, the biggest issue in power amplifier today is maximizing this metric. Furthermore, there are variations on this metric that give us more information about the power amplifier. The *drain efficiency* is defined as (Behzad Razavi, 1998):-

$$\eta_D = \text{drain efficiency} = \frac{\text{Power Delivered to Load}}{\text{Power Consumed in Final Stage}} \tag{3.3}$$

This tells us how efficient the final stage, often referred to as the *power stage*. The most common efficiency metric used, though, is the Power-Added Efficiency (PAE). The one thing missing in the previous two versions of efficiency is any idea of how much power is needed to drive the input to the power amplifier(Behzad Razavi, 1998).

PAE is defined as;

$$\text{PAE} = \frac{P_{RF_{OUT}} - P_{RF_{IN}}}{P_{DC}} \tag{3.4}$$

So the power amplifier that delivers 1W to the output, consumes 2W, and must be driven by a 100mW input (PAE=45%) is not as good as the PA that delivers 1W, consumes 2W, but needs only 1mW of input drive (PAE=50%). Since most power amplifier on the market is discrete components, the input drive is a critical concern for designers who plan to drop these discrete components into their designs. At RF, most discrete components are input-matched and output-matched to 50Ω, potentially requiring a large amount of power to drive the input of one of these components. Conversely, if a component is being designed in a single-chip, integrated environment, the input to the component can be made nominally capacitive (especially in CMOS, where the gate input is capacitive), and very little “real” power needs to be consumed to drive it.

Another key issue in the design of power amplifier is the issue of linearity. With efficiency being so dominant a concern in the end, designers look to boost the efficiency by any means necessary, even if it comes at the expense of another design parameter, like the linearity. In many systems, however, that is not an unreasonable trade-off; the reason being that many communications systems in use today utilizes modulation schemes that allow for reduced linearity performance in power amplifier. In general, modulation schemes can be separated into two basic categories: constant-envelope and non-constant-envelope. In the former, there is no symbol information contained in the amplitude of the transmitted signal, and thus there does not need to be a linear relationship between the input and the output. In the non-constant envelope case, there is symbol information contained in the transmitted signal, so the power amplifier must accurately amplify the amplitude of the signal that it is driven with. Non-linearity in the power amplifier can cause the transmitted signal to contain the incorrect information, corrupting the communications link.

In the constant-envelope modulation scheme, the symbol information is transmitted in the phase of the transmitted signal. Therefore, what is important is that the signal path not distorts the phase of the signal. Unlike the non-constant-envelope case, the signal to be transmitted will have constant amplitude, since no information is contained therein, and the power amplifier will convert that to a constant amplitude output signal. Even if the power amplifier does not linearly amplify differing input amplitudes, that is acceptable; only one amplitude will be present at any given time. In general, power amplifier for cellular systems must be able to control their output power level, meaning that they will have to supply different levels of power at different times, but as long as the input-output relationship is known, the input level can be varied to reach the desired output level (which will still be constant amplitude).

Phase distortion in cellular power amplifier is a concern among designers. In cellular power amplifier especially those in the mobile units, phase distortion manifests itself as spectral re-growth, which is often characterized by the Adjacent Channel Power

Rejection (ACPR) of the power amplifier. In the frequency domain, the power of the desired signal, which is mostly contained in a specified band, will be spread outside that specified band if the phase information is distorted. Unfortunately, most cellular standards have stringent requirements on the amount of power that can be transmitted in adjacent channels, in order to prevent interference between mobile units. It can be shown, however, that the cause of phase modulation in power amplifier is AM-PM distortion, or Amplitude Modulation- Phase Modulation distortion.

Nonlinear power amplifier will convert possible amplitude modulation in the signals at the input into phase modulation in the output, which will cause phase distortion, but if the input is constant amplitude, the phase information will be converted without distortion. For different amplitudes, the phase delay may be different, but because the amplitude is constant in a given frame, the relative phase between symbols will be unchanged, and the signal will be correctly transmitted. As a result, power amplifier with heavy nonlinearities can be investigated for use in systems using constant-envelope modulation schemes.

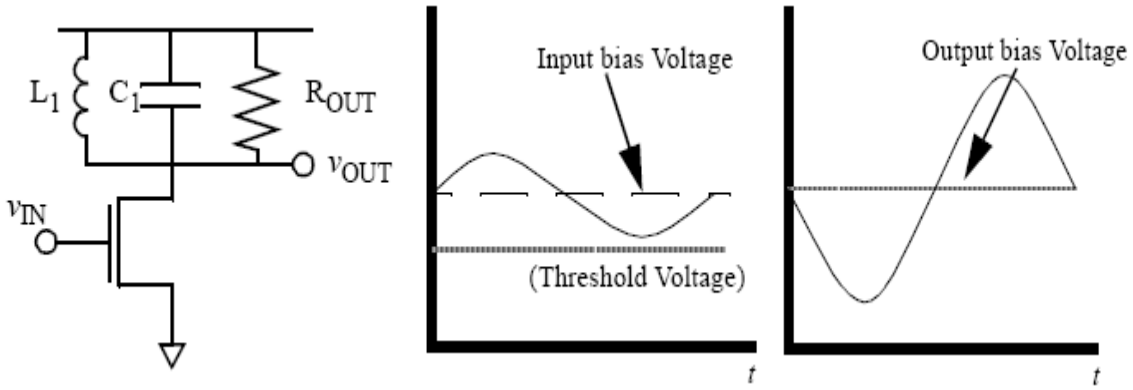
3.3 Power Amplifier Classes

In general, power amplifier can be placed into two different categories: one in which the device nominally acts as a current source (amplifying mode), and one in which the device acts as a switch. One common convention is to refer to the former group as the “linear” class of power amplifier, even though the specific implementation may have a very nonlinear relationship between input and output. The second category is then usually referred to as the “nonlinear” or “switch-mode” class mode of power amplifier. Each of the categories has several different sub-classes, which are used to identify the topology used in a particular implementation.

3.3.1 "Linear" or Amplification-mode Power Amplifier

The term "linear" is enclosed in quotes because the actual input-output relationship of power amplifier in this category does not have to be linear. For device operate in its amplifying region, FET devices should be in the saturation region and for bipolar devices should be the forward-active region. Since the devices are meant to operate in their amplifying region, it should be apparent that there will be some relationship between the magnitude of the input and output, regardless of how linear that is.

The big issue in the design of "linear" power amplifier is the trade-off between linearity (linearity of the input-output relationship) and efficiency. The most linear power amplifiers are those in which the amplification device is always conducting current; the instantaneous current through the device is a function of the instantaneous input, to first order. However, in order for this to be true, the DC bias of the input signal must be quite high, to allow for the device to still conduct current when the input swing is at its lowest point. As a result, the average power that is consumed, even when no signal is applied, can be quite high, and the efficiency will suffer. The above description describes the group of power amplifier known as Class A power amplifier, in which the amplifying device conducts current for the entire input sinusoid cycle. Figure 3.4 shows the basic



(a). A Basic Class A Implementation

(b). Input Signal for a Class A power amplifier

(c). Output Signal for Class A power amplifier

Figure 3.4 - Class A Power Amplifier Operation

implementation and operating conditions of a Class A power amplifier. The amplifying device (shown here as an FET device) is biased in such a way that it always remains in its amplification region, even under maximum input signal conditions. The input bias voltage is set such that the maximum input swing keeps the input signal above the threshold voltage required to keep the device on (Figure 3.4.b). The output voltage swings around its bias point. In general the bias point is the voltage supply, V_{DD}. In an ideal view, the maximum amplitude of the output swing is just V_{DD}, which can help us determine the peak efficiency of the Class A configuration. If we consider that

$$\hat{I}_o = \text{average current} \tag{3.5}$$

then the peak efficiency is given by

$$\eta = \frac{\frac{1}{2} V_{DD} \hat{I}_o}{V_{DD} \hat{I}_o} = \frac{1}{2} = 50\% \tag{3.6}$$

So the peak efficiency of the Class A power amplifier is 50%. However, this is in general not approachable in a realistic implementation. First, most components (both passive and active) will have some amount of real resistive loss in them, and thus there will be some measure of resistive loss due to the current flowing through these parasitic resistances. Moreover, in practice, getting a peak voltage swing of V_{DD} is usually not possible, as the amplification device will leave its amplification region and enter its resistive region (linear for FET, saturation for bipolar devices). As a result, the peak voltage swing is reduced; these and other real issues limit the attainable efficiencies of fully Class A power amplifier implementations to about 30%.

The next possibility would be to consider a power amplifier with a device that was not conducting current all the time. This is the idea behind the Class B power amplifier, sometimes also referred to as a “push-pull” output stage. In standard

implementations, two amplification devices are used, each of which amplifies the signal for half the sinusoidal period. The easiest way to accomplish this is to bias the devices such that they are on the edge of conduction in the quiescent state, and then to have the voltage excursions in each direction turn the appropriate device on. That is, as the voltage starts to go above its steady-state value, one of the devices will turn on; as the voltage falls below its steady-state value, that first device turns off and the second device turns on. Figure 3.5 depicts a common implementation and the standard waveforms of the Class B power amplifier. Because each of the devices is only conducting for one half of the input cycle (and not consuming power for the other half of the cycle), the efficiency of this implementation is greater than that of the Class A implementation[7]. The theoretical peak efficiency of the Class B power amplifier can be calculated to be:-

$$\eta = \frac{\Pi}{4} = 0.78 = 78\% \tag{3.7}$$

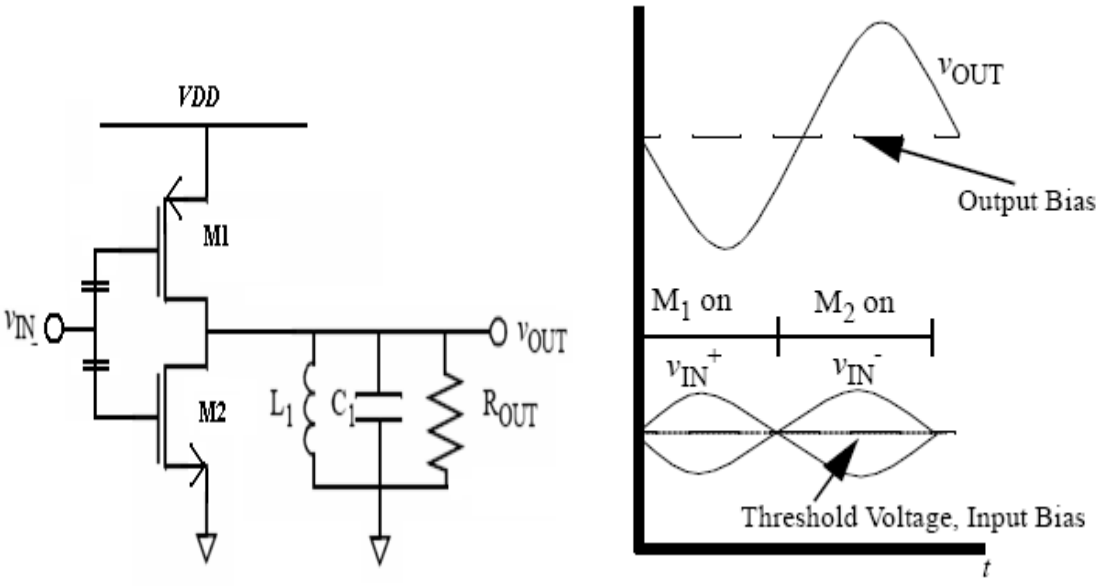


Figure 3.5 - Class B Power Amplifier Implementation

In practice, efficiencies of the Class B implementation can reach 50% or slightly more. However, the linearity of the power amplifier is degraded in this implementation. Not only are there issues of matching between the two devices (if the gain through the

devices is not exactly the same, the output will not be a smooth sinusoid), but if the two devices are not biased exactly at their threshold voltages, the issue of crossover distortion arises. Crossover distortion occurs when both devices are in their off-states when the input signal crosses zero. So while the efficiency of the stage has increased, the linearity has decreased.

One method of achieving somewhat better linearity is to compromise between the Class A and Class B implementations. This is known as the Class AB implementation. The power amplifier is now biased such that it is on for more than half the cycle (but not the entire cycle). In this case, the issue of a "dead" period when the crossover point is reached is avoided, because there is a portion when both devices in a push-pull implementation are on.

3.3.2 "Nonlinear" or Switch-mode Power Amplifier

The group of nonlinear power amplifier is also known by a more descriptive name: switch-mode power amplifier. These are the power amplifier in which the device is meant to act as a switch. For RF power amplifier, the two classes of switched mode power amplifier which have received the most attention are Class D power amplifier and Class E power amplifier. If the process of signal amplification is thought of as a power conversion process, then switched mode techniques used in power conversion systems, such as DC-DC converters or regulators, can be used in these power amplifier applications. That is the heart of the Class D and Class E architectures.

The Class D power amplifier is the first of these switched-mode classes of power amplifier, and has recently been implemented in a CMOS implementation. In the Class D power amplifier, current from the supply is steered between the device, when the switch is closed, and the load, when the switch is open. The Class D architecture is similar to what is used in a bridge DC-DC converter. In that style of DC-DC converter, devices acting as switches change the polarity of the input voltage onto the load, and the resulting

output is averaged to create a output voltage that is some fraction of the input voltage, depending on the duty cycle of the switching. This push-pull action can also be used in the design of an RF power amplifier, where the load is connected to switches which switch the voltage across the load.

If the switching is done at the output carrier frequency, the narrowband nature of the transmitted signal allows the use of the RF filters to pass only the fundamental frequency component. Again, because of the ability to filter out unwanted components of the output signal, this type of amplification can be done with only one device, in which case the power from the supply is either sunk in the device or the load. Because of the use of a series L-C circuit tuned to the output frequency, the current in the device will be a sinusoid for the period that it conducts current. If two devices are used, each will carry a half-sinusoidal current waveform (each will be on for half of each cycle). If the implementation of the switch is assumed to be ideal, i.e. no on-resistance and the output of the power stage can be 100%, as no power will be consumed in the transistor. However, due to non-zero on-resistance, the maximum attainable efficiencies can be dramatically reduced, especially in CMOS implementations; one published result of a Class D power amplifier in CMOS shows a drain efficiency of 62%.

The Class E power amplifier, while like the Class D power amplifier, uses the idea of soft switching in order to further reduce any power consumed by the device in the switched-mode power amplifier. This class of power amplifier has also been recently implemented in a CMOS implementation. Essentially, the Class E power amplifier tries to force the voltage on the output node to zero at the instant that the switch is closed, so that there is ideally no time at that transition when both the output voltage and current are non-zero. Similarly, the Class E power amplifier is designed to force the current flowing in the switch to be zero at the instant that the switch opens. Again, this is to ensure that there is no period of time around that transition when both the current and voltage are non-zero. Also, in order to account for timing errors in the switching instants, the slope of the output voltage waveform should also be zero at the instant of that the switch closes.