LOW POWER RF LOW NOISE AMPLIFIER AND POWER AMPLIFIER DESIGN

by

YANBO TIAN

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York

2006

UMI Number: 3213137

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the
Graduate Faculty in Engineering in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Professor Norman Scheinberg

Date — Chair of Examining Committee

Professor Mumtaz Kassir

Date — Executive Officer

Professor Xinghao Chen
Co-Chair of Examining Committee

Professor Mark Arend

Professor Jizong Xiao

Dr. John I. Smiths
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

LOW POWER RF LOW NOISE AMPLIFIER AND POWER AMPLIFIER DESIGN

by

Yanbo Tian

Adviser: Professor Norman Scheinberg

Professor Xinghao Chen

Low Power RF design is important for portable wireless communication systems. Low noise amplifiers and power amplifiers are two important components in such systems. The low noise amplifier is at the input to the receiver and sets the overall system sensitivity, and the power amplifier is at the end of the transmitter chain and provides the power to communicate between base station and mobile station

An optimization for two-stage low noise amplifier is presented. This approach allows getting minimal power dissipation under a power and linearity constraint.

A technique is presented to improve the efficiency of linear Class “A” power amplifiers by using dynamical current biasing with gain compensation. This technique is applied to a 2.4 GHz power amplifier. Simulations show that the dynamical current biasing results in a significant improvement in efficiency and that the gain compensation maintains the linearity as compared to a conventional implementation. With two tone input, the efficiency improvement is larger than 50% for amplifier driven 6-dB below saturation.

Acknowledgements

I will take this opportunity to show my most sincere gratitude and appreciation to my two advisors, Professor Norman Scheinberg and Professor Xinghao Chen, for their great guidance and supporting.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Wireless Applications

Wireless technology is widely used in our daily life, especially in cell phones. The first generation (1G) mobile communication systems—Advanced Mobile Phone Service (AMPS) in North America, and Nordic Mobile Telephone in Europe—are based on analog transmission technology. They employ frequency-division multiple access with analog frequency modulation (FM) and frequency-division duplexing. Later, modern communication system incorporates digital modulation.

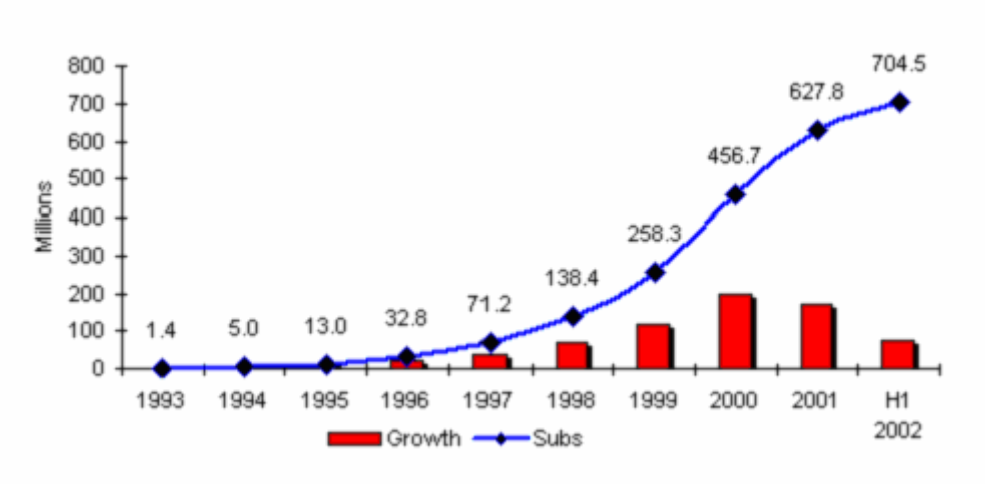


Figure 1.1 Global GSM Users, 1992–2002

Source: EMC World Cellular Database

In the 1990s, the second generation (2G) mobile network—Global System for Mobile communications (GSM), initially used in Europe—was deployed in 110 countries. The 2G systems employ digital modulation technology to improve capacities. As shown in Figure

1.1, the user base has grown exponentially [1].

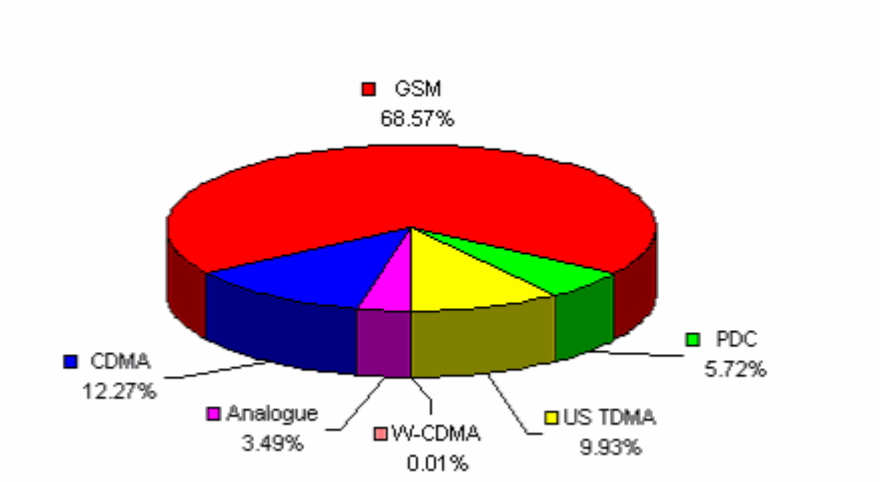


Figure 1.2 Different Technologies' Share of the Cellular Market in June 2002

Source: EMC World Cellular Database

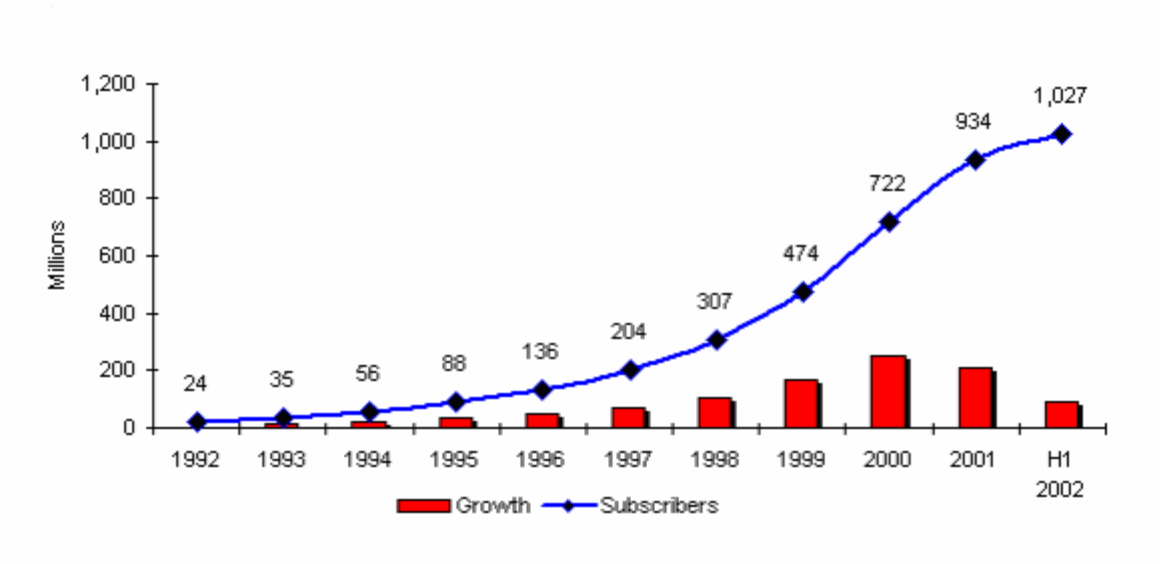


Figure 1.3 World Cellular Subscribers, 1992–2002

Source: GSM World

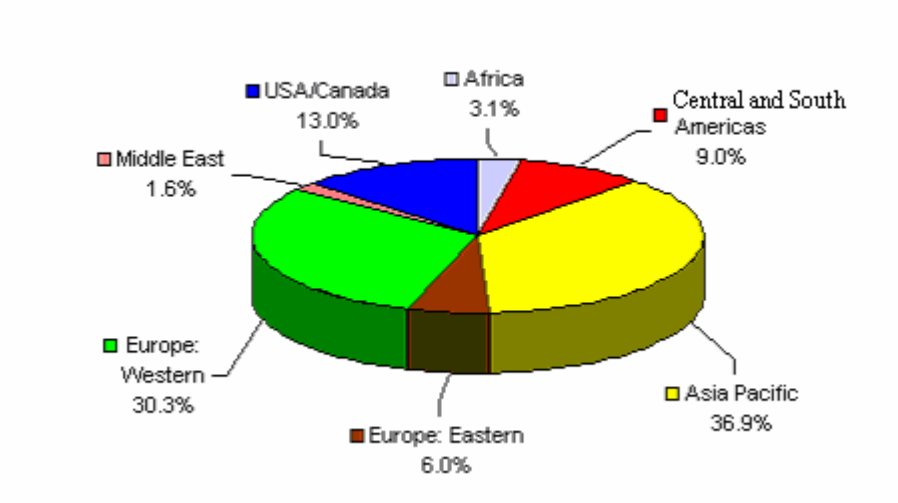


Figure 1.4 Globe Cellular Shares by Region

Source: EMC World Cellular Database

Because most of the word adopted the same GSM standards, subscribers can roam the GSM system nearly anywhere in the world. However, in the United States, two standards coexisted in 1990s—time division multiplex access (TDMA, standard IS-54) and code division multiple access (CDMA, standard IS-95)—and GSM was not widely deployed. Figure 1.2 shows different technologies' share of the world cellular market. Globally, the number of subscribers grew exponentially into 2002. Figures 1.3 and 1.4 show the number of subscribers worldwide and each region's share of the global cellular market, respectively; about 70% of wireless and cellular users are in Europe and Asia-Pacific.

The third generation (3G) technology, which provides better quality of voice, high data rate, and multimedia services, has begun to deploy worldwide. W-CDMA and CDMA2000 are two 3G standards. The number of CDMA2000 subscribers, which has 99% of the 3G market, approached 50.6 million in May 2003. The CDMA Development

Group forecasted that CDMA2000 will continue to lead in the near future. Figure 1.5 shows the growth of the CDMA2000 user base [2]. The first commercial 3G CDMA200 network was launched in Korea in October 2000. The total number of (2G and 3G) CDMA users reached 154 million in 2003 with an annual growth above 28%; 80% of those users are in Asia-Pacific and North America [2]. The reason the CDMA technologies were adopted as the 3G standard is that CDMA increases capacity by using coherence detection with synchronized transmitter and receiver clocks. The disadvantages of this system are that its power control is complex and it is implemented with less-efficient linear power amplifiers.

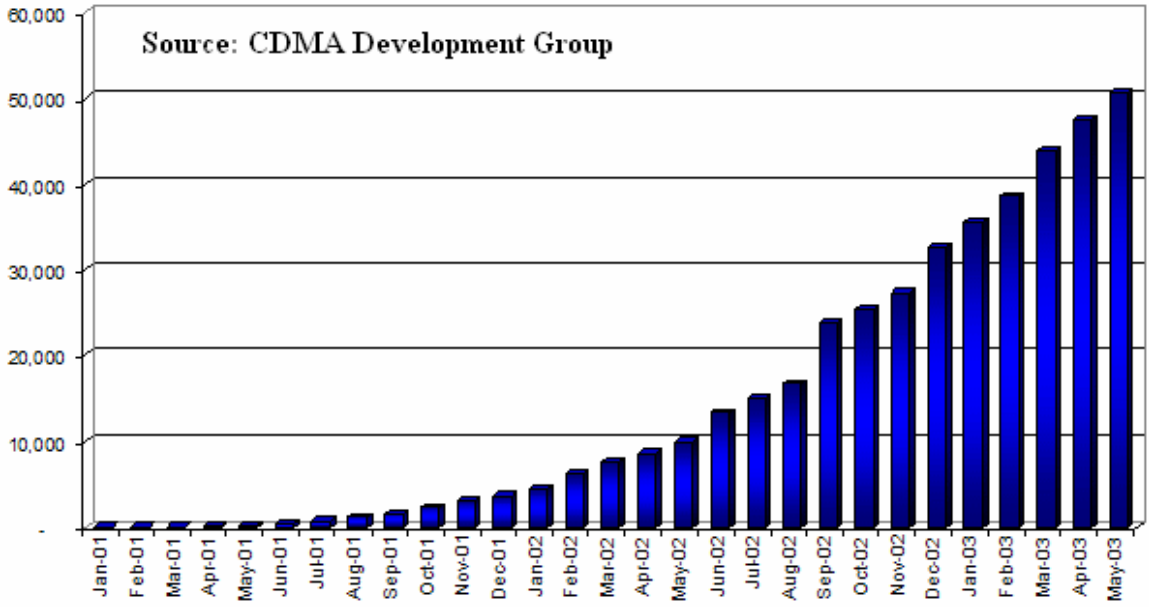


Figure 1.5 Growth in CDMA2000 Users

Source: CDMA Development Group

RF circuits are essential to cellular phones, and their functions are to provide

communications between a base station (the access point) and a mobile station (the phone). Each mobile phone must include both a transmitter and a receiver. They embed low-frequency digital data in a high-frequency carrier, sending the data through the antenna via the transmitter and decoding the digital data in the receiver. Although cellular technology is a major area for RF research, in reality, it is applicable in many other areas such as Wi-Fi (wireless LAN, WLAN), Bluetooth, Global Positioning System (GPS), etc.

1.2 Low Power RF Circuit Design

Although radio frequency (RF) is technically the electromagnetic spectrum from about 9 KHz up to thousands of gigahertz, most of today's new applications (cellular phones at 900 MHz and 1800 MHz, GPS at 1.6 GHz and Wi-Fi at 2.4 GHz and 5 GHz) use frequencies from several hundred megahertz to tens of gigahertz. RF technology is hardly new, tracing its roots to the beginning of the twentieth century. During the last hundred years, scientists and engineers have invented and improved numerous RF circuits in wireless systems for applications such as telegraph, radio and TV. All these applications have become parts of our normal life.

Yet before the 1960s, all RF circuit components were discrete, which made them big, heavy and expensive. The quality varied greatly between products due to process variations. Integrated-circuit (IC) technology, which emerged in the 1960s, has made huge advancements; today there are billions of transistors in a single chip. However, for RF circuits, IC applications are very limited. This is because conventional digital IC technologies (before the 1990s) do not provide a means to implement inductors, which are

widely used in RF circuits.

To take advantage of the benefits of IC technology and at the same time to meet strict power consumption and reliability requirements (among others), RF products must use large numbers of off-chip components, making them expensive and inconvenient. Fast-growing cellular technology calls for smaller, cheaper RF implementations. Integrating more components, including on-chip inductors, is a promising way to achieve these objectives. Radio-frequency integrated circuit (RFIC) technology is now the focus of industry and academic research.

Table 1.1 Some Popular Cell Phones and Their Features

Model	Size (mm)	Talk time (m)	Standby time (h)	Weight
Motorola V70[3]	94 x 38 x 18	135-215	70-145	83 g
Motorola V60[3]	87 x 45 x 24	150-240	150-242	111 g
LG VX4400[4]	88 x 48 x 27	110-180	110	113 g
Sony Ericsson T80i[5]	100 x 48 x 20	< 720	< 390	84 g
Nokia 3650[6]	130 x 57 x 26	120-240	150-250	130 g

With so many portable devices entering our daily life, consumers want them to be cheaper, smaller and easier to use. Table 1.1 lists some popular cell phones and their features. The average talk time is about 2 to 4 hours, adequate for conversations but too short for such applications as web surfing. 3G network deployments have enabled many data-oriented and multimedia applications, which typically take more time. Improvements in digital IC technology have enabled many new features in mobile devices, such as short messages and

the web camera for taking and transferring pictures. Meanwhile, IC technology scaling has enabled digital circuits to consume less power, leaving front-end RF circuits to consume a large portion of the total system power. Thus, low power RF circuit technology is in great demand.

The largest power consumer in a cell phone is the power amplifiers (PA), which can dissipate as much as several watts at peak output. This is because, in wireless communication, signal power fades proportionally to the square of the distance in open space. In a harsh environment such as a city, the fading will be even worse, $Power \propto 1/(distance)^{\alpha}$, with α between 2 and 4, which indicates that the required power is much greater than in wired communication systems.

To lower the overall power consumption and extend battery life, careful considerations must be made from the system design and architecture selection to the circuits and IC technology. Figure 1.6 shows the considerations for low-power RF circuit design. At the top, the system must incorporate power control algorithms to put RF units into sleep mode when these circuits are not in use and wake them up from time to time to determine if they are needed. On the circuit and technology level, low-power circuit topologies and high-quality passive components (such as on-chip inductors) can reduce power consumption substantially. Currently, with IC technology advances, the quality of both active and passive devices has improved. CMOS, originally used in digital applications, has recently shown promise in RF applications. However, there are still many challenges in low-power RF design.

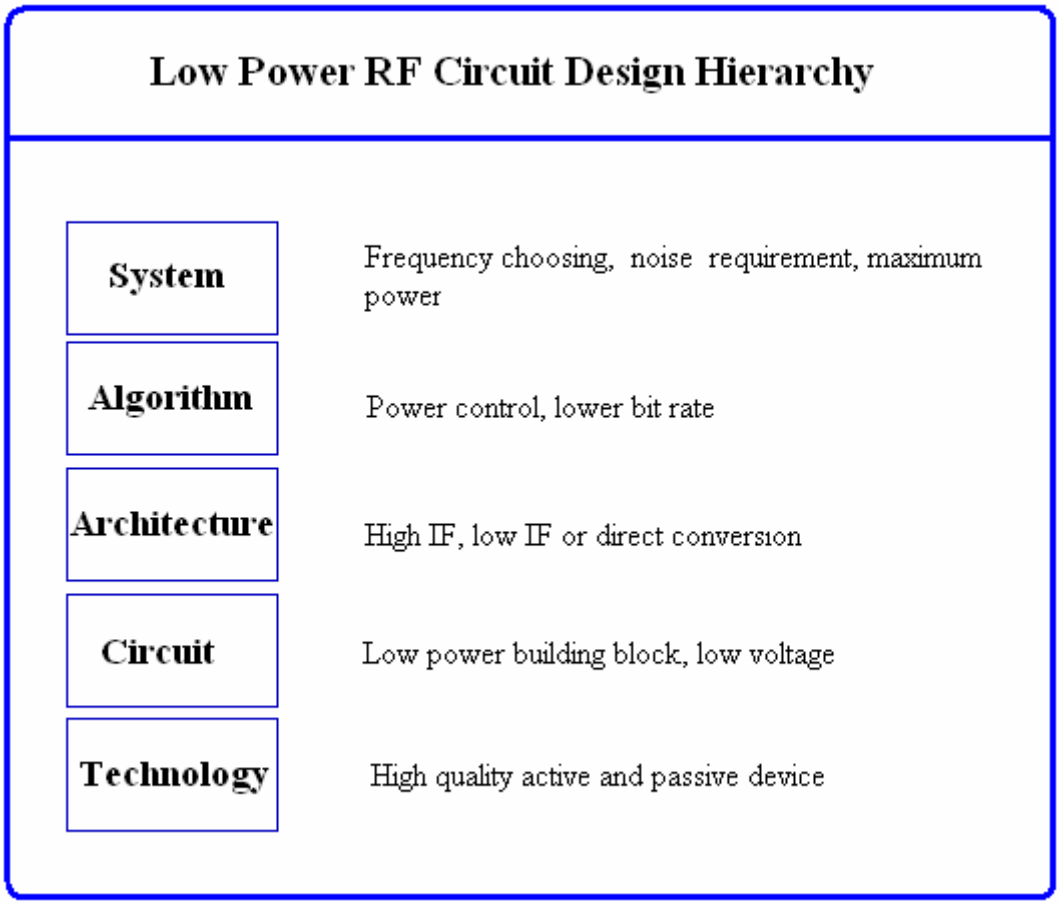


Figure 1.6 Low-Power RF Circuit Design Hierarchy

1.3 RF Transceiver Architecture

All communication-system RF parts include at least one transmitter and one receiver, sometimes called a transceiver. A generic RF transceiver system is shown in Figure 1.7. In the receiver (Rx) path, a signal coming from the antenna first passes a band-pass filter to attenuate out-of-band signals. The signal then goes through a low-noise amplifier (LNA), which must provide enough gain and at the same time generate a minimum of noise. After the LNA, the signal goes to a down-converter circuit, which shifts the frequency from RF to the baseband. After the baseband amplifier, the signal is fed to the analog-to-digital

converter (ADC). Finally the digital signal is handled by a powerful digital signal processor (DSP). In the transmitter (Tx) path, a digital signal is first fed to a digital-to-analog converter. It then passes through a pulse-shaping block, which can alleviate intersymbol interference (ISI) due to narrow channel filters. Then an up converter modulates the baseband signal and shifts the frequency from baseband to the RF range. The PA amplifies the RF signal and sends it to the antenna. A good RF designer, therefore, must be familiar with analog circuits and know some digital design as well.

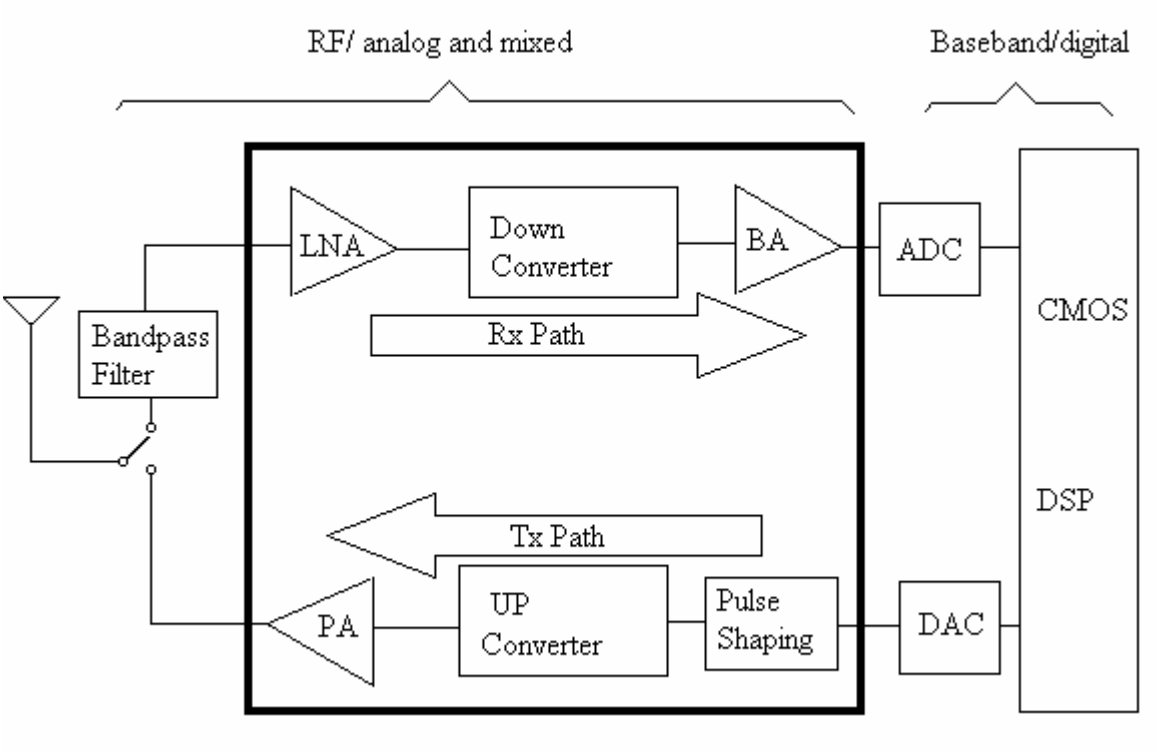


Figure 1.7 Typical Transceiver System

Note: the bold box indicates the integrated RF front end

There are three types of architecture to shift signals from RF to baseband, heterodyne, low–intermediate frequencies (Low IF) and homodyne (zero-IF or direct conversion).

Each has its advantages and problems. Heterodyne is the most mature architecture for down converters. In this architecture, the signal band is translated to one or several IFs before being shifted to baseband. This method relaxes the quality-factor (Q) requirement of each stage's channel-select filters (used to suppress adjacent channel signals and image signals). Image frequency is the frequency that is shifted to the same IF as the signal after the down-converter, see the second part. By using high-quality discrete passive components, high-speed and good-quality signals can be achieved with lower power consumption. The main problem is the low degree of circuit integration because high-quality inductors are not available in modern IC components. Heterodyne architecture requires a large number of off-chip components. This approach is not only costly due to the cost of the discrete filters and high pin count for receiver chips, but it also increases power consumption because those filters have to be driven by a 50-Ω matched signal source (which is the de facto impedance standard).

The homodyne architecture (zero-IF) shifts signals directly from RF to DC. This topology can achieve a much higher degree of integration; there is no image filter needed because the mirror signal is the signal itself. By removing inter-stage filters, homodyne circuits are simpler, less expensive and more efficient, which is a premium for very low-power designs. However, homodyne suffers from such problems as DC offset, I/Q mismatch, crosstalk by RF and local oscillator, even-order distortion, local oscillator leakage, and flicker noise ($\propto 1/f$). When only discrete active and passive devices are available, process variations make it difficult to solve such problems. Even with modern

IC technology, mismatch is still a problem, though it can be held within tolerable levels. Consequently, zero-IF receivers have been limited to rather low-performance applications such as pagers and industrial, scientific and medical applications in which the coding can be scrambled so that high-pass filters can be inserted to avoid the DC offset problem. There are several research reports [7, 8] on zero-IF.

On the other hand, the trade-off of noise and low-power low-cost make the Low-IF approaches more attractive. There is usually only one low-frequency IF stage in Low-IF architecture, typically half of channel width. The corner frequency of flicker noise is below this IF, and there is no the DC offset problem. The image frequency of a Low-IF system is located within band and its power is predictable, which will relax the requirements of the image reject filter. Unfortunately, the Low-IF architecture still needs an image filter. Additionally, if the circuit's ability to handle a large signal is poor, the frequency halfway between a local oscillator and the RF frequency may generate substantial noise [9].

A detailed architecture is shown in Figure 1.8 along with different technologies to build the blocks. Note that there are two down-converters in the receiving path, which can relax the *Q* requirement of the image reject filter. In this architecture, CMOS dominates the digital processing circuits. With technology scaling, CMOS's cut-off frequency, f_T, keeps increasing. Thus CMOS technology has begun to show up in RF areas. Much research has been done to build the CMOS RF components. Almost all building blocks, including low-noise amplifiers, mixers, voltage-controlled oscillators (VCO) and PAs, are reported using CMOS technology [7, 8, 10–18]. Currently, these CMOS designs are limited to 5

GHz, and SiGe and GaAs are major technologies for higher-frequency RF applications. The major commercial RF components still use Si BJT or BiCMOS technology, which provides both BJT and CMOS devices on a single chip. For applications whose working frequencies are greater than 3 GHz, SiGe technology gives better performance. GaAs heterojunction bipolar transistor (HBT) technology, which provides higher linearity and lower power-consumption characteristics, dominates the PA section. That is by far the hardest part to integrate onto a single chip.

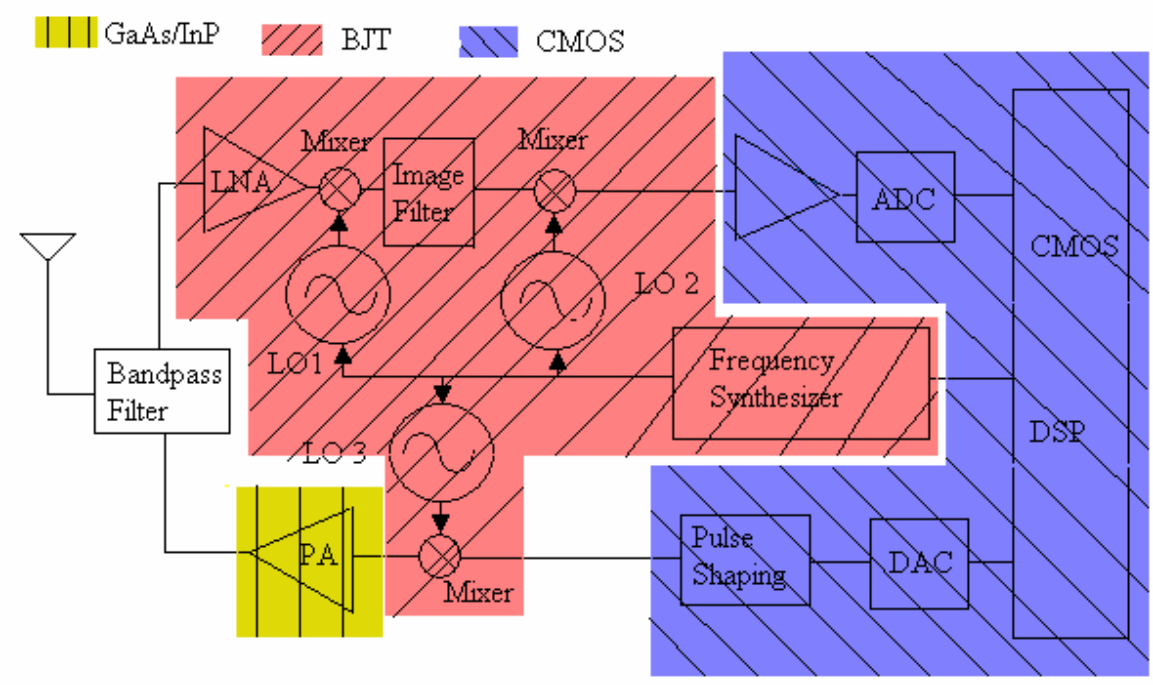


Figure 1.8 RF Front-End Architecture and Technology to

Implement Different Building Blocks

References

[1] http://www.gsmworld.com

[2] http://www.cdg.org/worldwide/cdma_world_subscriber.asp.

[3] http://commerce.motorola.com/consumer/QWhtml/phone_comp.html.

[4] http://www.lginfocomm.com/products/product_detail.php?p=26.

[5] http://www.sonyericsson.com/us/spg.jsp?page=start.

[6] http://www.nokia.com/phones/.

[7] S. Mahdavi and A. Abidi, "Fully integrated 2.2-mW CMOS front end for a 900-MHz wireless receiver," *IEEE J. Solid-State Circuits,* vol. 37, pp. 662-669, May 2002.

[8] J. Ryynanen, K. Kivekas, J. Jussila, L. Sumanen, A. Parssinen, and K. A. I. Halonen, "A singal-chip multimode receiver for GSM900,DCS1800, PCS1900, and WCDMA," *IEEE J. of Solid-State Circuits,* vol. 38, pp. 594-602, Apr. 2003.

[9] D. Manstretta, M. Brandolini, and F. Svelto, "Second-order intermodulation mechanisms in CMOS downconverters," *IEEE J. Solid-State Circuits*, vol. 38, pp. 394-406, Mar. 2003.

[10] D. J. Cassan and J. R. Long, "A 1-V transformer-feedback low-noise amplifier for 5-GHz wireless LAN in 0.18-um CMOS," *IEEE J. Solid-State Circuits*, vol. 38, pp. 427-435, Mar. 2003.

[11] C. J. Debono, F. Malberti, and J. Micallef, "On the design of low-voltage, low power CMOS analog multipliers for RF application," *IEEE Tans. Very Large Scale*

Integration (VLSI) Systems, vol. 10, pp. 168-174, Apr. 2002.

[12] Q. Li *et al*., "CMOS RF low-noise amplifier design for wireless communication," *Proc. 43rd IEEE Midwest Symp. on Circuits and System*, Aug. 2000.

[13] M. Lin, Y.-m. Li, and H.-y. Chen, "A 0.18-um CMOS LNAs with sub-2 dB NF and low power for bluetooth applicatios," *Proc. 4th Int.l Conference on ASIC*, 2001.

[14] R.-C. Liu, C.-R. Lee, H. Wang, and C.-K. Wang, "A 5.8-GHz two-stage high-linearity low-voltage low noise amplier in a 0.35-um CMOS technology," *Symp. Radio Frequency Intergated Circuits (RFIC)*, 2002.

[15] M. S. J. Steyaert, B. D. Muer, P. Leroux, M. Borremans, and K. Mertens, "Low-voltage low-power CMOS-RF transceiver design," *IEEE Trans. Microwave Theory and Techniques*, vol. 50, pp. 281-287, Jan. 2002.

[16] F. Svelto, S. Deantoni, G. Montagna, and R. Catello, "Implementation of a CMOS LNA plus mixer for GPS applications with no external components," *IEEE Trans. Very Large Scale Integration (VLSI) System*, vol. 9, pp. 100-104, Feb. 2001.

[17] C.-Y. Wu and H.-S. Kao, "A 2-V low-power CMOS direct-conversion quadrature modulator with integrated quadrature voltage-controlled oscillator and RF amplifier for GHz RF transmitter applications," *IEEE Trans. Circuits and Systems*, vol. 49, pp. 123-134, Feb. 2002.

[18] K.-H. Cheng, L.-J. Tzou, W.-B. Yang, and S.-S. Sheu, "A CMOS low power voltage controlled oscillator with split-path controller," *The 8th IEEE Int. Conf. Electronic, Circuits and Systems (ICECS)*, 2001.

Chapter 2 RF LNA and PA Design

.

2.1 Low-Noise Amplifier Design

2.1.1 Noise Figure and Linearity Definition

The LNA is an important part of the whole RF system; it must have enough gain, provide sufficient linearity and add a minimum of noise. As the first active device in the receiving path, it usually provides a 50-Ω matching input impedance.

Noise factor (*NF*), which is the parameter to measure added noise when a signal passes through a network, is defined as:

Definition 1: $NF = SNR_{in}/SNR_{out.}$ (*SNR* stands for signal to noise ratio)

or alternative $NF = \dfrac{\text{Total output noise}}{\text{Total ouput noise due to the source}}$

Noise figure is the logarithms of *NF*, defined as

Definition 2**:** Noise Figure = $10log_{10}(NF)$.

Generally speaking, the less noisy a system, the smaller the noise figure; the noise figure is zero if the system is noiseless. If *m* stages are cascaded, the overall *NF* is give by the Friis equation [1],

$$NF_{tot} = 1 + (NF_1 - 1) + \frac{NF_2 - 1}{A_{p1}} + \cdots + \frac{NF_m - 1}{A_{p1} \cdots A_{p(m-1)}}, \tag{2.1}$$

where NF_k is the noise factor of the *k*th stage and A_k is the available power gain of that stage.

Definition 3: $A_k = \dfrac{\text{Output Power at } k\text{th stage(matched condition)}}{\text{Input Power at } k\text{th stage(matched condition)}}$

The Friis equation shows clearly that the noise contributed by each stage decreases as the gain in the preceding stage increases. Thus noise factors of early stages are very critical and attenuation is equivalent to amplifying the noise.

In addition to minimizing noise contribution, the LNA must provide high linearity, the ability to handle large signals. The input inception point (IIP_3) or the 1-dB compress point ($P_{1\text{-}dB}$) are used to characterize the linearity.

Definition 4: IIP_3 = input power when the third-order intermodulation distortion (IMD3) matches the signal power.

Definition 5: $P_{1\text{-dB}}$ = input power when power gain decreases by 1 dB.

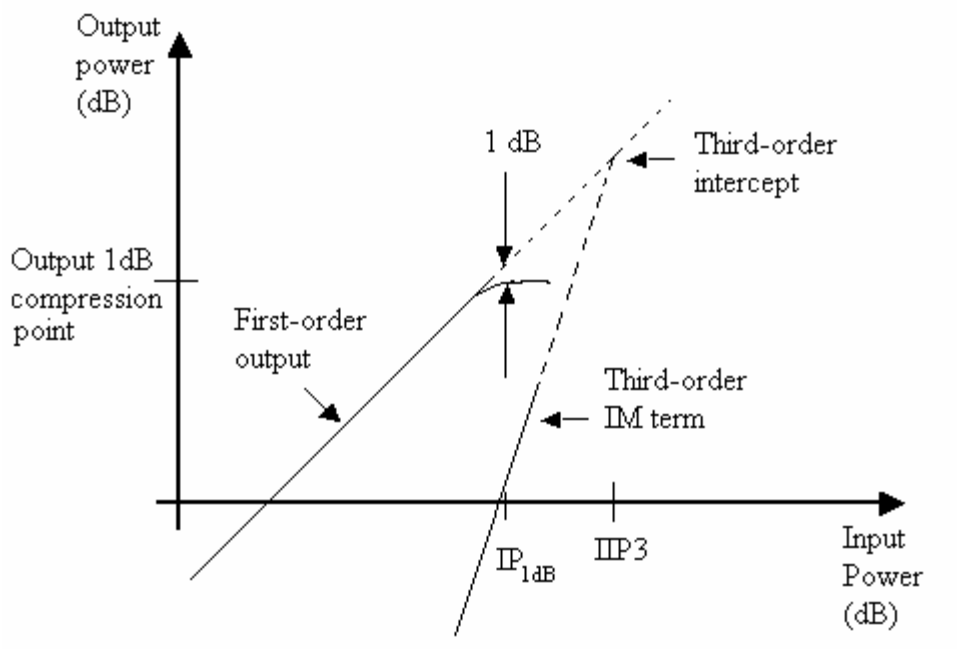


Figure 2.1 Illustrations of IIP_3 and $P_{1\text{-dB}}$

When input and output signals are of polynomial relations as defined by Equation (2.2),

$$y(x) = c_1 x + c_2 x^2 + c_3 x^3 + \cdots, \tag{2.2}$$

IIP_3 is calculated by

$$IIP_3 = \frac{2 \cdot |c_1|}{3 \cdot |c_3| \cdot R_s}. \tag{2.3}$$

Spur-free dynamic range (SFDR) is a parameter to define receiver dynamic range in terms of two undesired interferers and the receiver noise floor. The SFDR is the difference in dB between the receiver noise floor and the level of each of two equal-amplitude, out-of-band interfering tones that produce an in-band spurious product equal in power to the noise floor. To get a larger SFDR, usually a single-stage LNA circuit needs a larger bias current. Hence, the larger dissipation power [2]. LNAs can also be configured with two stages to reduce the total power consumption.

2.1.2 Noise Models of Resistors and Transistors

To estimate *NF*, we need to know the noise models of the devices used in the LNA. Passive resistors and active devices generate noise, while capacitors and inductors do not. Figure 2.2 shows the models of a resistor (a), a BJT transistor (b) and a MOSFET transistor (c).

k is Boltzmann constant and *f* is the frequency. In Figure 2.2 (b), *q* is the electron charge, I_B and I_C are DC base and collector currents, K_b is a constant depending on technology, and r_b is the base resistance. In Figure 2.2 (c), g_{d0} is zero-biasing transconductance, which is defined as

$$g_{d0} = \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{DS}=0}. \tag{2.4}$$

g_{d0} is roughly equal to g_m which may depend on source-drain voltage. K_m is a constant.

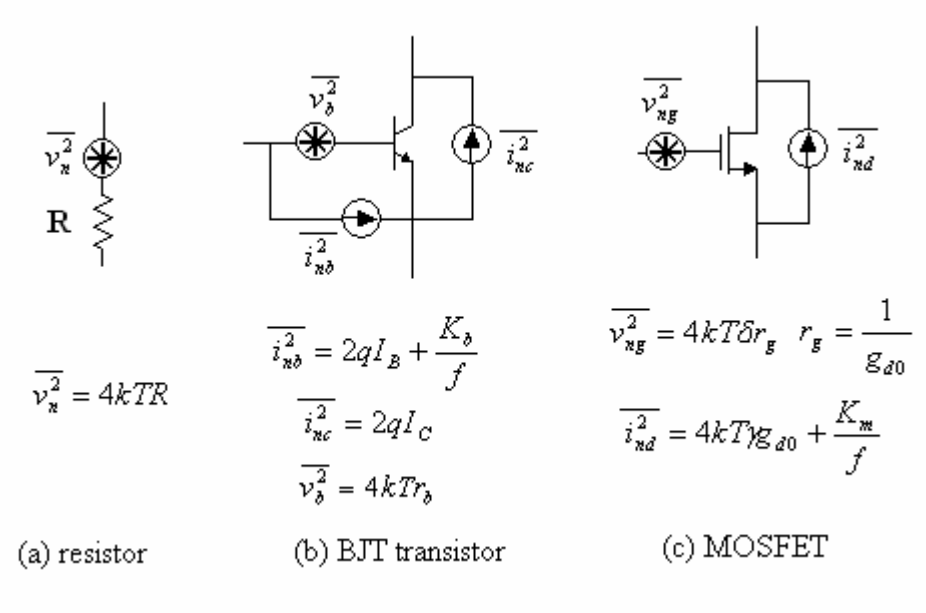


Figure 2.2 Noise Models of a Resistor, a BJT Transistor and a MOSFET

All three devices generate thermal noise. The noise from the resistor is proportional to its resistance. The thermal noise of the BJT comes from the base resistor: $\overline{v_b^2} = 4kTr_b$, Figure 2.2 (b). The thermal noise of the MOSFET is associated with the channel, $\overline{i_d^2} = 4kT\gamma g_{d0}$. The BJT also displays shot noise, which is proportional to the currents flowing though the junction ($2qI_b$, $2qI_c$). The transistors produce flicker noise, too, which is inversely proportional to frequency. Flicker noise may cause trouble when an up or down converter shifts frequency directly from DC to RF or vice versa. The dominant noise source for the BJT is shot noise; for the MOSFET it is the thermal noise.

2.1.3 Low-Noise Amplifier Architectures

There are several architectures for building a LNA. Figure 2.3 gives five architectures and all provide 50-Ω input matching.

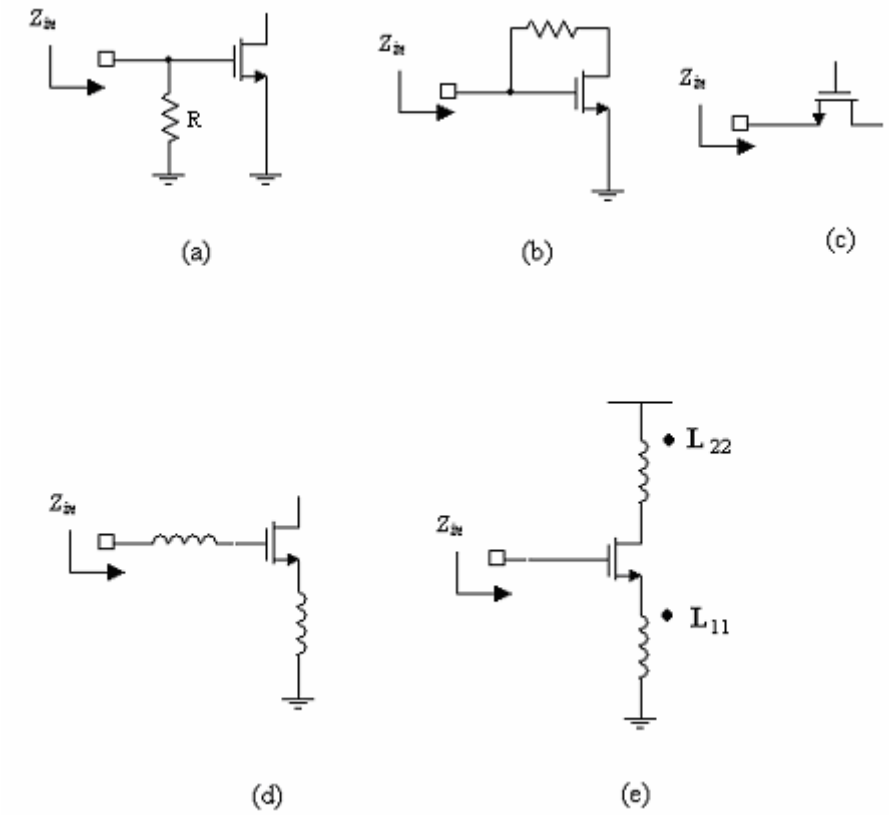


Figure 2.3 LNA Architectures (a) Resistive Termination (b) Shunt-Series Feedback (c) Common Gate (d) Inductive Degeneration (e) Transformer Feedback

(a) Resistive Termination: Input matching is achieved by adding a resistor R as shown in Figure 2.3 (a); the resistor provides the source impedance. This architecture produces more noise with matched input.

$$NF = \frac{4KTR_s(1/2g_mZ_l)^2 + 4KTR_{in}(1/2g_mZ_l)^2 + 4KT\gamma g_{d0}Z_l^2}{4KTR_s(1/2g_mZ_l)^2} = 2 + \frac{2\gamma g_{d0}}{R_s g_m^2} \qquad (2.5)$$

The lower limitation of Equation (2.2) is 3 dB. In reality, the lower limit of *NF* is much greater. According to Shaeffer [3], a *terminationless* amplifier with a 6-dB noise figure would likely possess an 11.5-dB noise figure with the addition of the termination resistor. The reason is that this resistor not only produces its own share of noise, but also attenuates the input signal. The Friis equation shows that if an amplification stage attenuates the signal instead of amplifying it, the *NF* after that stage is amplified. The combined noise contribution makes this topology unattractive.

(b) Shunt-Series Feedback: This architecture employs the shunt-series resistor, as shown in Figure 2.3 (b), to provide load matching. However, it suffers from high power consumption with the same noise figure as other architectures [3]. This is due to its broadband design, which is unnecessary in narrowband applications.

(c) Common Gate: In common gate amplifiers, shown in Figure 2.3 (c), the input impedance is $1/g_m$, which should be equal to the source resistor, R_s = 50 Ω. Properly choosing the bias current, we can make the input impedance equal to the source impedance. We assume that the capacitors are negligible, the load, Z_L, is noiseless, and r_0 is much larger than Z_L. From MOSFET small-signal model (Figure 2.4), the noise factor is given by Equation (2.6).

$$NF = \frac{\overline{v_{ns}^2}(\frac{g_m Z_l}{2})^2 / Z_l + \frac{\overline{i_d^2}}{4} Z_l}{\overline{v_{ns}^2}(\frac{g_m Z_l}{2})^2 / Z_l} = \frac{\overline{v_{ns}^2} g_m^2 + \overline{i_d^2}}{\overline{v_{ns}^2} g_m^2} = 1 + \frac{4KT\gamma g_{d0}}{4KTR_s g_m^2} = 1 + \frac{\gamma g_{d0}}{g_m} = 1 + \frac{\gamma}{\alpha}, \tag{2.6}$$

$$\text{where } \alpha = \frac{g_m}{g_{d0}}.$$

Long-channel devices have the inequality $\frac{2}{3} < \gamma < 1$. For short-channel devices, γ is much larger, 2/3 in the saturation region [4]. So, letting $\gamma_{L\text{-}C}$ = 2/3 and α = 1, we find the lower limit of $NF_{L\text{-}C}$ for CMOS: 1 + 2/3 = 5/3 = 2.2 dB

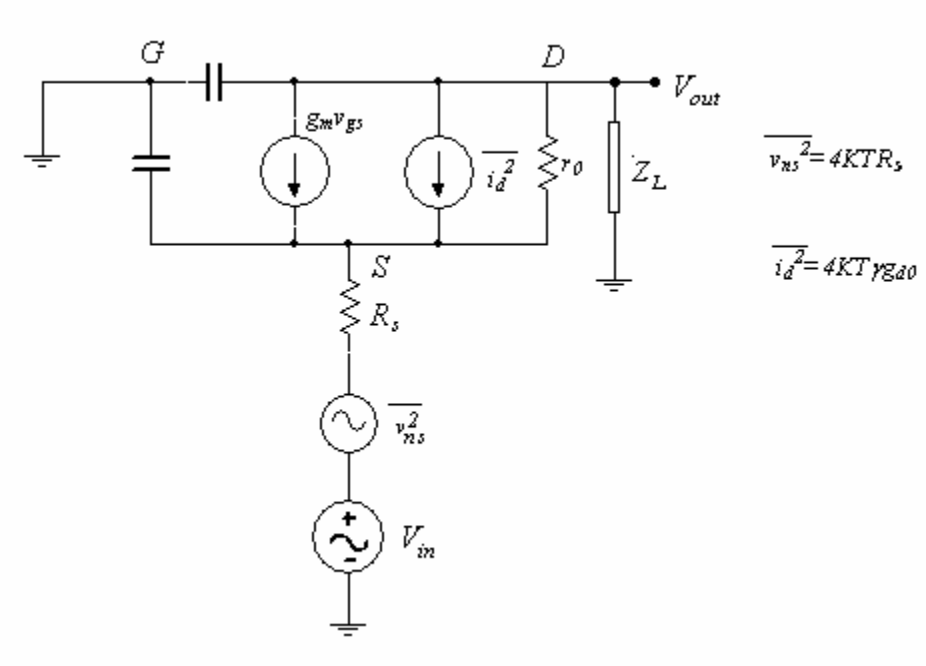


Figure 2.4 Small-Signal Circuit Model of a Common-Gate LNA

For BJT, the shot noise dominates and $g_m = I_{bias}/V_T$, $\overline{i_d^2} = 2qI = 2qg_m V_T = 2KTg_m$. Following the above procedure, we get NF_{BJT} = 3/2 = 1.76 dB.

The advantage of the common-gate architecture is that it is easier to match, has greater linearity and provides great reverse isolation for homodyne architecture, which is

especially useful in a homodyne system where RF signals are directly shifted to baseband. However, this architecture suffers from an inherently high noise figure. For applications needing a sub-2-dB NF LNA, this architecture would not be a good choice.

(d) Inductive Degeneration: This is the most popular architecture used in LNA designs. It can provide a sub-2-dB *NF*. The reason it can achieve that low *NF* is that the noise contributed by the transistor is on the output. The topology is shown in Figure 2.3 (d). L_g and the parasitic capacitor C_{gs} form a simple L-section matching network to provide the 50-Ω input impedance. Figure 2.5 shows a typical LNA circuit with channel noise. The cascoded topology minimizes the Miller effect. Thus, C_{gd} is negligible.

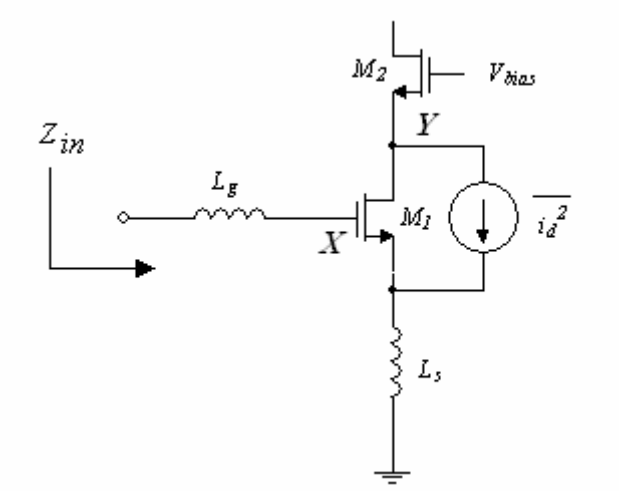


Figure 2.5 Common-Source Cascaded LNA Stage

Noise from the second transistor (M2) is minor and is ignored. M2 serves to alleviate the Miller effect of C_{gd} (*X* and *Y*). Figure 2.6 is the small-signal model of the device in Figure 2.5, the input impedance is given by Equation (2.7).

$$Z_{in} = sL_g + g_m \frac{L_s}{C_{gs}} + sL_s + \frac{1}{sC_{gs}} = j\omega(L_g + L_s - \frac{1}{\omega C_{gs}}) + \omega_T L_s \tag{2.7}$$

At the resonance frequency, $\omega = 1/\sqrt{(L_g + L_s)C_{gs}}$, the input impedance is expressed as

$$Z_{in} \approx \omega_T L_s \text{ at } \omega = 1/\sqrt{(L_g + L_s)C_{gs}} . \tag{2.8}$$

The 50-Ω input impedance can be achieved by choosing the proper L_s. This inductor can use a bondwire, which has high Q. L_g can be made with a low-Q on-chip spiral inductor or a high-Q off-chip inductor.

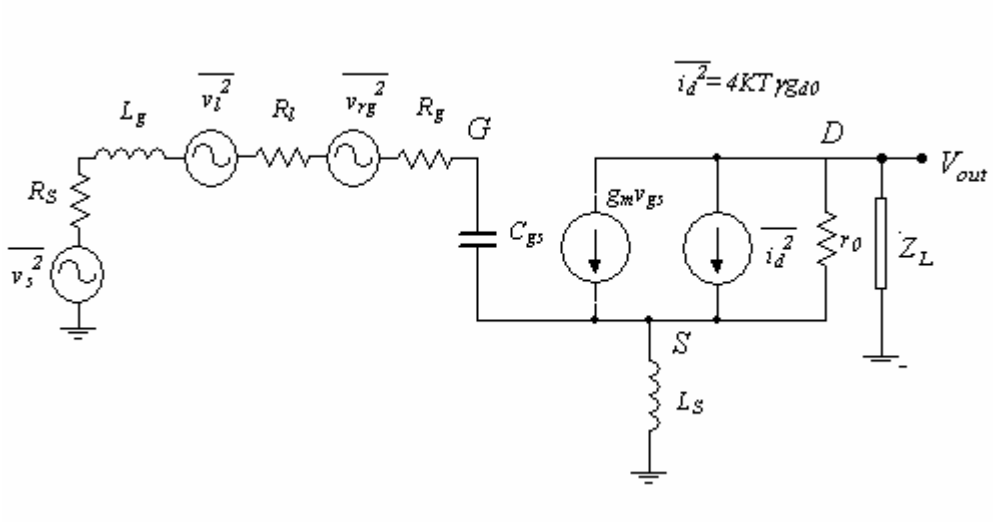


Figure 2.6 Small-Signal Equivalent Model of an Inductive Degeneration LNA

Assuming $r_0 >> Z_L$, L_s is small and its Q is high, so that the series resistor from L_s is neglected. Only thermal noise from channel is considered. The gain from input to output is

$$Gain = A = \frac{g_m Z_l}{sC_{gs}(R_s + sL_g + R_l + R_g) + 1 + s^2 L_s C_{gs} + sL_s g_m} \tag{2.9}$$

At resonance frequency—that is, $\omega_0 = 1/\sqrt{(L_g + L_s)C_{gs}}$ —Equation (2.8) is simplified:

$$|Gain| = \frac{g_m Z_l}{\omega_0 C_{gs}(R_s + R_l + R_g + \frac{L_s g_m}{C_{gs}})} = \frac{\omega_T Z_l}{\omega_0 (R_s + R_l + R_g + \omega_T L_s)} \approx \frac{\omega_T}{2\omega_0 R_s} Z_l , \quad (2.10)$$

where R_l and R_g are small compared to R_s [3].

The output noise comes from $\overline{i_d^2}$ and is described as

$$\overline{i_{out}^2} = \frac{\overline{i_d^2}(R_s + R_l + R_g)^2}{(R_s + R_l + R_g + \omega_T L_s)^2} \approx \frac{\overline{i_d^2}}{(1 + \frac{\omega_T L_s}{R_s})^2} \approx \frac{1}{4}\overline{i_d^2} , \quad (2.11)$$

and the total *NF* is

$$NF = \frac{\text{Total output noise}}{\text{Total output noise due to the source}} = \frac{\overline{v_s^2}A^2 + \overline{v_l^2}A^2 + \overline{v_g^2}A^2 + \overline{i_{out}^2}Z_l^2}{\overline{v_s^2}A^2}$$

$$= 1 + \frac{R_l + R_g}{Rs} + \frac{KT\gamma g_{d0}}{4KTR_s(\frac{\omega_T}{2\omega_0 R_s})^2} = 1 + \frac{R_l + R_g}{Rs} + \gamma g_{d0} R_s (\frac{\omega_0}{\omega_T})^2 , \quad (2.12)$$

with small R_l and R_g, the last term dominates. According to Equation (2.12), by decreasing g_{d0} without modifying ω_T, we can simultaneously improve the noise figure and reduce the power dissipation. This can be achieved by scaling down the width of the device while keeping ω_T constant (which only depends on the biasing voltage) and leaving the channel length unchanged.

The analysis above does not consider the gate noise, nor the noise due to the parasitic resistance of L_g. Choosing a high-Q inductor not only lowers R_l, and thus *NF*, but also reduces the power loss. On the other hand, Q should not be selected too high, which will detune the circuits and make them sensitive to the impedance of band filter or antenna.

Considering the gate noise for the long-channel case, Shaeffer and Lee [3] gave a detail analysis of the inductive degeneration architecture. They showed that the minimal *NF* is given by the following.

$$NF_{min} = 1 + 1.33\ (\omega/\omega_T)\ \textit{fixed}\ G_m\ (\textit{total transconductance})\ \textit{optimization} \tag{2.13}$$

$$NF_{min} = 1 + 1.62\ (\omega/\omega_T)\ \textit{fixed}\ P_D\ (\textit{total power consumption})\ \textit{optimization} \tag{2.14}$$

Lee [5] also gave a rule-of-thumb guidance to choose the optimal LNA MOSFET transistor width in the power-constrained condition.

$$W_{op,P} = \frac{3}{2}\frac{1}{\omega L C_{ox} R_s Q_{sP}} \approx \frac{1}{3\omega L C_{ox} R_s}\ L \text{ is gate length,} \tag{2.15}$$

where $Q_{sP} = \dfrac{1}{\omega C_{gs} R_s}$ is defined as the input's Q. (Assume Q_{sP} = 4.5 here)

The second transistor in the cascode load adds some noise, too; the analysis follows the same procedure. The overall noise figure is given as

$$Gain = A = \frac{\dfrac{1}{sC_{gs1}} g_{m1} Z_l}{R_s + sL_g + R_l + R_g + \dfrac{1}{sC_{gs1}} + sL_s + \dfrac{L_s g_{m1}}{C_{gs}}} \frac{g_{m2}}{g_{m2} + sC_{gs2}} \tag{2.16}$$

$$\begin{aligned} NF_{tot} &= 1 + \frac{R_l + R_g}{Rs} + \gamma_1 g_{d01} R_s \left(\frac{\omega_0}{\omega_{T1}}\right)^2 + \frac{\dfrac{1}{1 + \left(\omega_0 / \omega_{T2}\right)^2} 4KT\gamma_2 g_{d02}}{\dfrac{KT}{R_s}\left(\dfrac{\omega_{T1}}{\omega_0}\right)^2} \\ &= 1 + \frac{R_l + R_g}{Rs} + \gamma_1 g_{d01} R_s \left(\frac{\omega_0}{\omega_{T1}}\right)^2 \left(1 + \frac{4}{1 + \left(\omega_0 / \omega_{T2}\right)^2} \frac{\gamma_2 g_{d02}}{\gamma_1 g_{d01}}\right) \end{aligned} \tag{2.17}$$

Typically, with $\omega_0 << \omega_T$ and $\gamma_1 \approx \gamma_2$, the total noise figure, Equation (2.17) can be simplified to

$$NF_{tot} = 1 + \frac{R_l + R_g}{Rs} + \gamma_1 g_{d01} R_s (\frac{\omega_0}{\omega_{T1}})^2 (1 + \frac{4g_{d02}}{g_{d01}}) . \quad (2.18)$$

If $W_2 << W_1$ (that is usually the case) and $4g_{d02}/ g_{d01} << 1$, Equation (2.15) returns back to Equation (2.11).

The disadvantage of inductive degeneration is its narrowband architecture; it is thus unsuitable for broadband applications.

(e) Transformer Feedback [6]: This configuration is shown in Figure 2.3 (e) and can also be viewed as inductive degeneration. In ultra-low-voltage designs (less than one volt), where inductive degeneration's transistor stack is not workable, the transformer feedback architecture can work. In this architecture, negative feedback from the transformer neutralizes the Miller effect from C_{gd} by providing some output signal to flow back and cancel the input going into the output through C_{gd}, thus linearity is improved. This architecture not only provides a way to design an LNA with a power supply under one volt, but also increases the linearity [6].

The input impedance matching analysis is the same as for inductive degeneration. However, here an on-chip transformer replaces the inductor L_s. Having only one active component lowers *NF*, but it needs an on-chip transformer, which typically occupies a larger area. (Long [7] provided a detailed analysis and design procedures for a monolithic on-chip transformer.) Additionally, poor Q and an inaccurate approximation limit this architecture's utility.

2.1.4 Reported Low Noise Amplifier Specification

In some cases, one stage may not be enough to meet both low *NF* and large IIP_3—that is, large SFDR. Two-stage LNAs are the choice for such requirement.

Though originally dominated by GaAs and SiGe technology, technology advancements have made the CMOS LNA possible, which is cheaper to make with high integration and can be integrated with digital circuits. The power consumption of a typical LNA is between 10 and 40 mW, usually below 20 mW, with noise figures below 4 dB. On the other hand, GaAs technology provides high linearity and low *NF* (Table 2.1), but is more expensive.

Table 2.1 Some Recent Reports of Low-Noise Amplifiers

Reference	Tech.	*NF* (dB)	Gain (dB)	P_{1dB} (dBm)	*f* (GHz)	P (mW)	Stage
Aoki ’02 [8]	InGaP	2.32	22.5	9.8	5.25	28.3	2
Chui ’02 [9]	CMOS	2.17	11	-8.3	5.2	10	1
Liu ’02 [10]	CMOS	3.2	7.2	-3.7	5.8	20	2
Debono ’01 [11]	CMOS	4	10.5	-24 (IIP_3)	1.8	40	2
Chi ’03 [12]	CMOS	4	12	-13	2.4	7.94	2
Kaukovuori ’03 [13]	SiGe	2.3	10.2	-26	2	-	2
Cassan ’03 [6]	CMOS	0.9	14.2	0.9 (IIP_3)	5.75	16	1
Nguyen ’02 [14]	CMOS	1.7	18.9	-15.6	5.2	22.3	2

2.2 Power Amplifier

2.2.1 Power Amplifier Classification

The PA is the transceiver's most power-consuming block. Designing PAs, especially for high-linearity and low-voltage operations, remains difficult and still defies an elegant solution. In practice, designing a PA is still essentially a trail-and-error process. By far, the PA is also the most difficult block to integrate into a single-chip solution, mainly due to the requirement that the PA handle large current while maintaining high linearity. GaAs or InGaP technologies, which provide high linearity and low noise, still dominate most, if not all, commercial cellular products.

Amplitude-insensitive modulation can use more-efficient nonlinear PAs, but for the amplitude sensitive modulation, linear PAs are mandated due to issues with distortion and spectral regrowth. (When variable-envelope signals pass a nonlinear PA, the resulting spectrum becomes broader than the original. This effect is called spectral regrowth and can be quantified by the power of the relative adjacent channel.)

PAs are categorized by classes. Class A, AB, B, C, D, E and F are widely used depending on the application. A generic PA topology is shown in Figure 2.7.

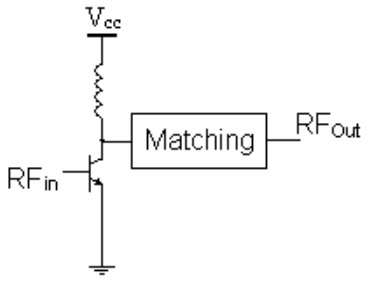


Figure 2.7 A Generic PA Topology

In Class A, AB, B and C PAs, the transistor acts as a current source. In Class D, E and F PAs, the transistor acts as a switch. Class A, AB and B PAs possess high linearity and are called linear PAs; the others are called nonlinear PAs. Nonlinear PAs are more efficient and widely used in constant envelope applications where the amplitude contains no information.

Class A, AB, B and C PAs are distinguished by their operation points. Figure 2.8 shows the I–V curve of a typical MOSFET (a BJT is similar). The drain voltage is $V_{DS} = V_{DD} - IR_L$., and this line is called the load line (with arrow in Figure 2.8). The black dots indicate the transistor's operation points. Class A PAs enjoy highest linearity because the load line is totally within the active region. The Class AB PA load line lies partially within the active region, and that of the Class B is only half in the active region. For the Class C PAs, the transistor working less than 50% in the active region. Comparing linearity, A > AB > B > C, but for efficiency, the order is reversed, A < AB < B < C.

In Class D PAs, the transistors are driven to deep saturation (BJT) or the triode region (MOSFET), and the output waveform is very square. The efficiency of Class D PAs can reach almost 100%. However, this PA can only work with a constant-envelope signal, such as FM. The major applications of Class D PAs are in radio-station PA, and the output power is typically several hundred watts.

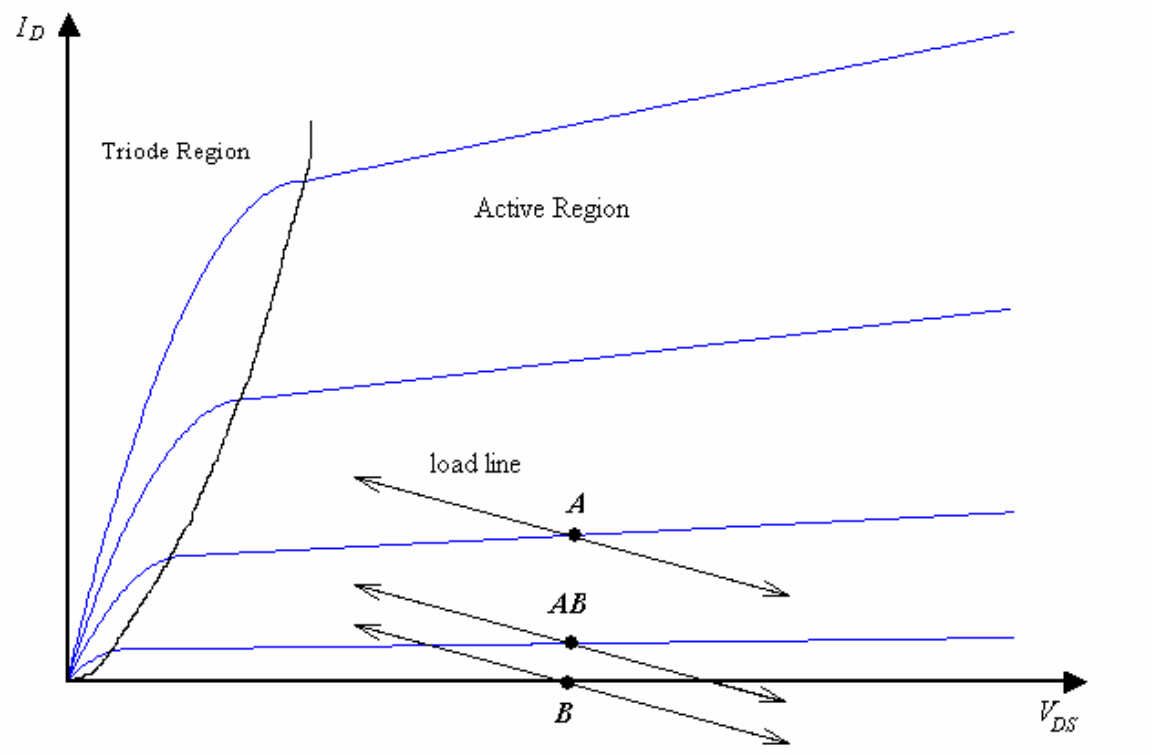


Figure 2.8 Bias Points of Different PAs, Class A, AB and B

Class E and F PAs are relatively new. In these PAs, the transistors are no longer seen as a current source; rather they are seen as switch. Figure 2.9 presents a simple switching PA circuit.

The transistor *M* will turn on and off ideally. High efficiency can be reached under the following conditions: (a) If the switch is on, the voltage at the drain is small; (b) if the transistor is off, the voltage at the drain is high; and (c) the transient time between on and off is minimized. If all three conditions are met, the efficiency will be almost 100%. So ideally, the input waveform should be rectangular to meet (a) and (c). This is true in Class D PAs, but condition (b) is not easily reached.

The implementation of Class E PAs involves proper load design. It uses a high-order reactive network that provides enough degrees of freedom to shape the switch voltage to have both zero value and zero slope when the switch on, thus reducing the switch losses. The problems of Class E PAs are that (a) they do nothing for the turn-off transition, (b) their ability to handle large power is limited and (c) they, thus, require the use of rather oversized devices to deliver a given amount of power to a load compared to other classes of PAs.

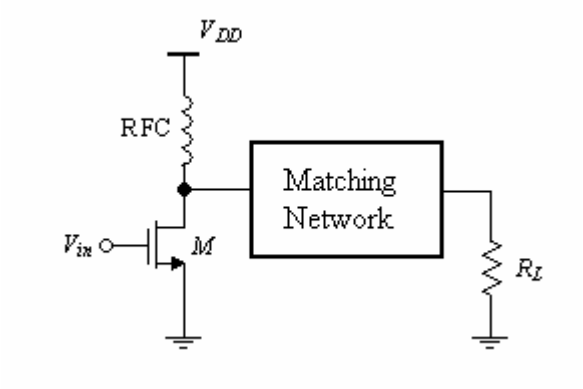


Figure 2.9 A Simple Switching PA Circuit

Class F PAs use harmonic terminals that show load impedance at the first-order harmonic, almost zero for higher-order even harmonics and open for higher-order odd harmonics. A square wave with 50% duty cycle has only odd components, so ideally a square-wave voltage results at the drain.

Figure 2.10 gives examples of different PAs. The inductor labeled RFC is a large inductor with an impedance so large that it is open for RF signals. The RFC doubles the transistor-drain peak to $2V_{DD}$, thus relaxing the requirement for a high voltage power

supply.

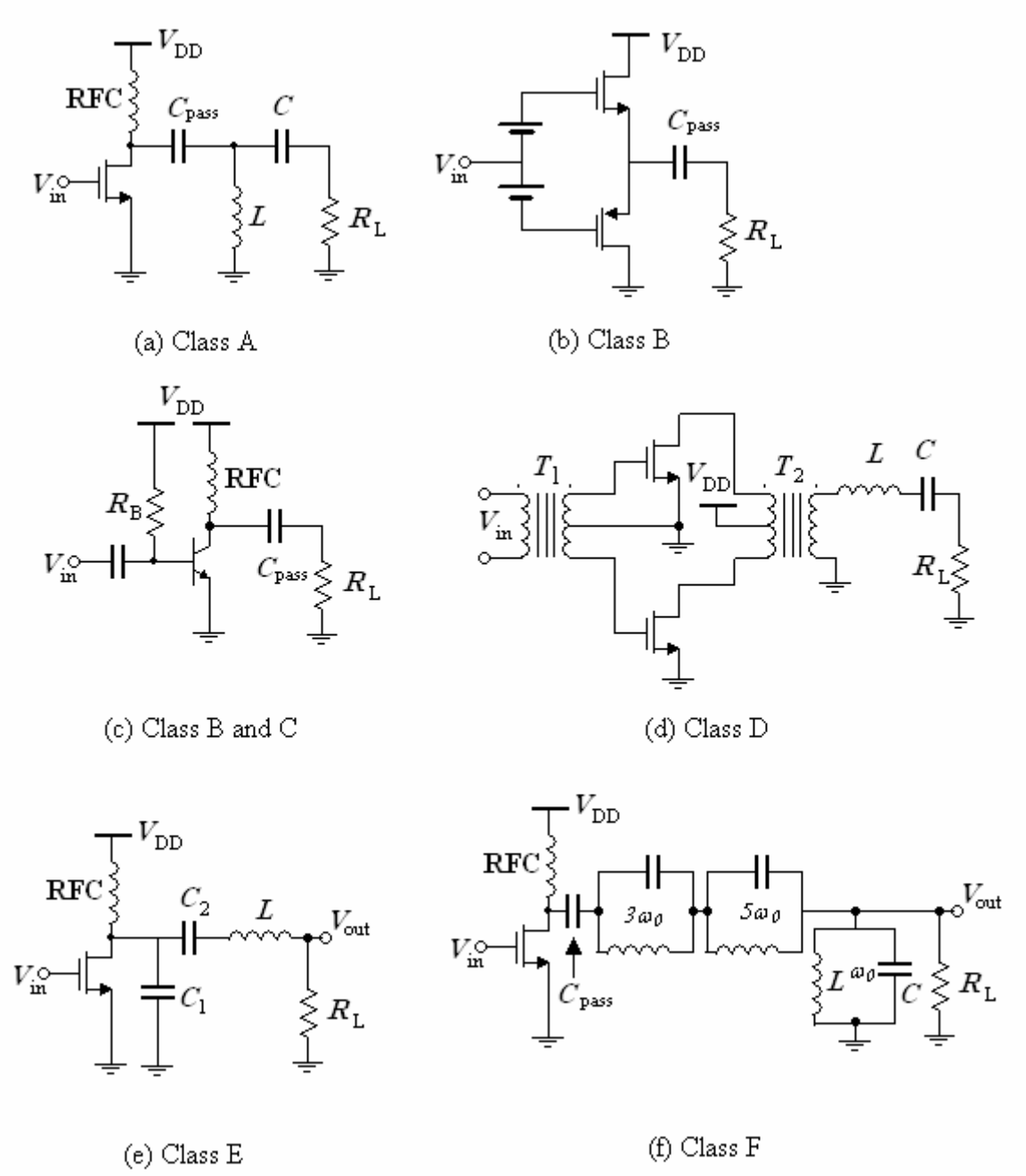


Figure 2.10 Different PAs

Two important parameters for the PA are the efficiency and linearity. The efficiency

characterizes the how much power is wasted and the linearity states how precisely the PA tracks the input waveform. A PA usually provides input and output matching if it cannot be integrated into one chip. (For example, GaAs cannot be integrated on silicon). The matching can be on chip or off chip, matching networks of one or more stages may be used.

There are two methods to define efficiency, drain efficiency, η, and power-added efficiency (PAE). They are defined as

$$\eta = \frac{P_{out}}{P_{ps}} \qquad PAE = \frac{P_{out} - P_{in}}{P_{ps}} \quad P_{ps} \text{ is the power from the power supply} \tag{2.19}$$

When the gain is large, the two definitions are roughly the same.

The maximum efficiency of a Class A PA is 50%, when the output reaches the maximum value.

$$\eta_{CLASSA} = \frac{P_{out}}{P_{in}} = \frac{\frac{1}{2}\frac{V_o^2}{R}}{V_{dd} I} = \frac{1}{2}\frac{V_{dd} I}{V_{dd} I} = 0.5 = 50\% \tag{2.20}$$

The maximum efficiency of a Class B PA is $\pi/4 \approx 78.5\%$ from Equation (2.21). The efficiency of the Class AB PA lies between those of the Class A and B PAs, and the efficiency of the Class C PA is greater than that of the Class B PA.

$$\eta_{CLASSB} = \frac{P_o}{P_{ps}} = \frac{\frac{1}{2}\frac{V_{DD}^2}{R}}{\frac{2}{T}\int_0^{T/2} V_{DD} I dt} = \frac{\frac{1}{2} V_{DD} I_{peak}}{\frac{2}{T}\int_0^{T/2} V_{DD} I_{peak} \sin(\omega t) dt}$$
$$= \frac{\frac{1}{2}}{\frac{2}{T\omega}(-\cos\omega t\big|_0^{T/2})} = \frac{1/2}{\frac{2}{2\pi}\cdot 2} = \frac{\pi}{4} \qquad (2.21)$$

The efficiencies of Class D, E and F PAs are 100% in ideal conditions. Though the efficiencies of these PAs are high, they show poor linearity and can only be used to amplify constant-envelope signals. Table 2.2 summarizes the different class PAs.

Table 2.2 Traditional PA Classification

Class	Mode	Conduction Angle Φ	Output Power	Maximum Efficiency	Gain	Linearity
A	Current Source	100%	fair	50%	large	good
AB		B < Φ < A	fair	A < η < B	large	fair
B		50%	fair	78.5%	fare	fair
C		< 50%	small	100%	small	poor
D	Switch	50%	large	100%	small	poor
E		50%	large	100%	small	poor
F		50%	large	100%	small	poor

The Class A PA enjoys highest linearity, which makes it the only choice in many applications. Class B is also a good candidate because it has a higher efficiency while maintaining reasonable linearity, especially, in a low-power design. Class B is almost the unique choice for low-power applications requiring linear amplifiers.

2.2.2 General Design Approaches of Different Power Amplifiers

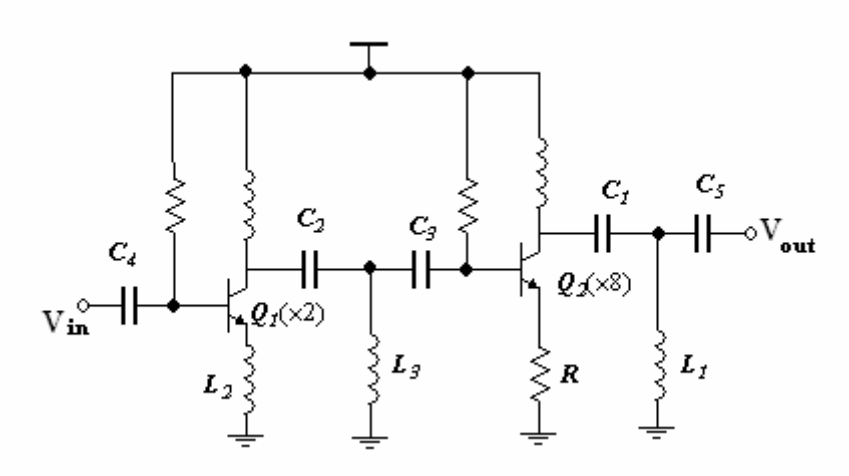


Figure 2.11. A Class A Power Amplifier

Class A, AB, B and C PA designs are similar: The first step is to figure out how much current is required in the last stages and make the proper impedance transformation. Then, according to the required gain, two or three stages may be needed to meet the goal and overall linearity. Then, depending on the PA class, the biasing network is designed to make the circuits work in the corresponding region. Using more than four stages is rare for stability reasons. To maximize efficiency, PA design precludes the use of any technique that diminishes the supply headroom. A typical two-stage Class A PA is given in Figure 2.11.

R is a small resistor to prevent thermal runaway due to a negative thermal coefficient. L_1 and C_1 convert the output impedance from 50 Ω to some smaller number. L_3, C_2 and C_3 form the matching network to connect Stage 1 and Stage 2. C_4 and C_5 are bypass capacitors to block DC currents. There are eight identical transistor cells in parallel to

meet the current requirement. Detailed consideration and modeling of the parasitic inductors and capacitors during simulation and careful layout are important; in most cases, iterations are unavoidable.

The design of a Class E PA involves determining the value of *L*, C_1 and C_2 in Figure 2.10 (e). According to Sokal [15], inductor and capacitors values are given by.

$$L = \frac{QR}{\omega} \tag{2.22}$$

$$C_1 = \frac{1}{\omega R(\pi^2/4-1)(\pi/2)} \approx \frac{1}{\omega(R\cdot 5.447)} \tag{2.23}$$

$$C_2 \approx C_1(\frac{5.447}{Q})(1+\frac{1.42}{Q-2.08}) \tag{2.24}$$

Designing a Class F PA is straightforward. Shown in Figure 2.10 (f), just choose L and C to make the tank resonate at the appropriate harmonic frequency.

Because PAs deal with large signals, the maximum power transfer theorem is useless though it still has its role in designing the input circuits. Load pull is usually the way to determine the optimal output impedance. Still, there are trade-offs between linearity, efficiency, gain and even noise.

Many applications need linear PAs, such as QPSK, π/4 - QPSK and the quickly developing Wi-Fi technology. Of the three Wi-Fi standards—802.11a, 802.11b and 802.11g (adopted in 2003)—802.11a and 802.11g use orthogonal frequency-division multiplexing (OFDM) to increase the bit rate. This is a kind of multiple-carrier modulation, and it has a high peak-to-average ratio. This modulation requires very linear

PAs. Currently, Class A and Class AB are used at the output stages. The maximum efficiency of linear Class A PAs is around 30–40%, 40–50% for Class AB PAs.

Table 2.3 Recent Reports on PA Design

Ref.	Tech.	Freq.	Gain	PAE	IMD3	P_{out}	Class
Xie'03[16]	GaN, HEMT	3 G	13 dB	34%	32 dBc/P_o = 26 dBm	36 dBm	B
Sowlati'03[17]	CMOS	2.4 G	31 dB	45%	- (GFSK)	23.5	-
Yen'03[18]	CMOS	2.45 G	11.2 dB	28%	20 dBm/$P_{o\text{-}1dB}$	20 dBm	AB
Shirvani'02[19]	CMOS	1.4 G	-	49%	-	22.6 dBm	F
Berkhout'03[20]	HV-DMOS SOI	1 K	-	> 90%	-	200 W	D
Fortes'01[21]	CMOS	1.9 G	-	42%	-	23 dBm	F
Mertens'02[22]	CMOS	700 M	-	62%	-	30 dBm	E

2.2.3 Ways to Improve Efficiency of Linear Power Amplifiers

The efficiency of linear PAs is poor. However, as mobile communication technology advances, it calls for high-efficiency linear PAs. There are some techniques to improve the average efficiency of such PAs because they work at low efficiency when the input signal is small. The major methods are the Doherty amplifier, the Chireix outphasing amplifier, envelope elimination and restoration and dynamic biasing. They all try to lower the average power consumed by the power supply. Other methods are feedforward and digital baseband predistortion. Such methods depend on the characteristics of PAs and are therefore only suitable for complete system design.

(a) The Doherty Amplifier

The Doherty amplifier [23], first proposed in 1936, is primarily an efficiency enhancement, a power-conservation technique. The basic configuration of Doherty PA is shown in Figure 2.12.

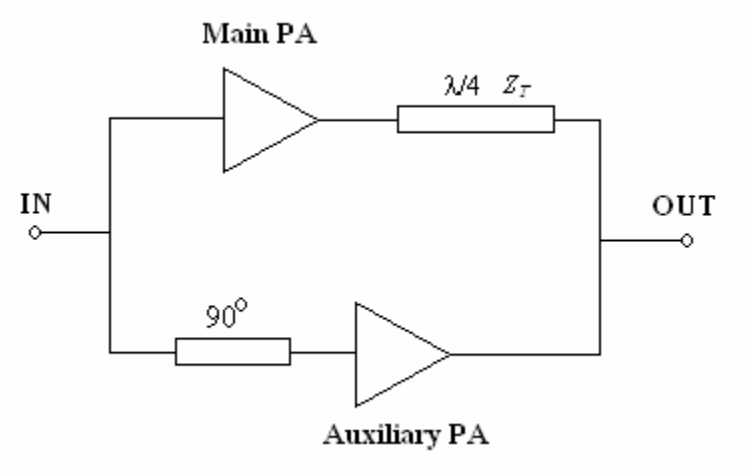


Figure 2.12 A Doherty PA

The λ/4 transmission line (λ is the wavelength of the RF signal) with the characteristic impedance $Z_T = R_{opt}$ (optimal load) ≈ V_{dc}/I_{dc} acts as impedance transformer. For the optimum operation of the Doherty configuration, the resistive load needs to be half of the characteristic impedance of the quarter-wave transmission line [24]. The 90° transmission line in the auxiliary PA path compensates for the phase shift in main PA path. Note, this load is before output matching, so it does not need to be 50 Ω. The Doherty PA works as follows: When signal is small, only the main PA is on, the main PA can biased as Class A, Class AB or Class B. Because the load is twice the optimal load, the main PA becomes saturated when the current reaches half of the maximum value, improving the efficiency. When the signal becomes larger than P_{on}= P_{max} -6 dB (up 6 dB range before saturation),

the main PA saturates and suitable biasing turns on the auxiliary PA. The auxiliary PA current increases the equivalent load impedance. When both the main and auxiliary PA saturate, the load impedance seen from the main PA is the optimal load R_{opt}. Figure 2.13 gives the ideal efficiency curves of the Doherty and Class B PA vs. power back off (main PA biasing as a Class B PA).

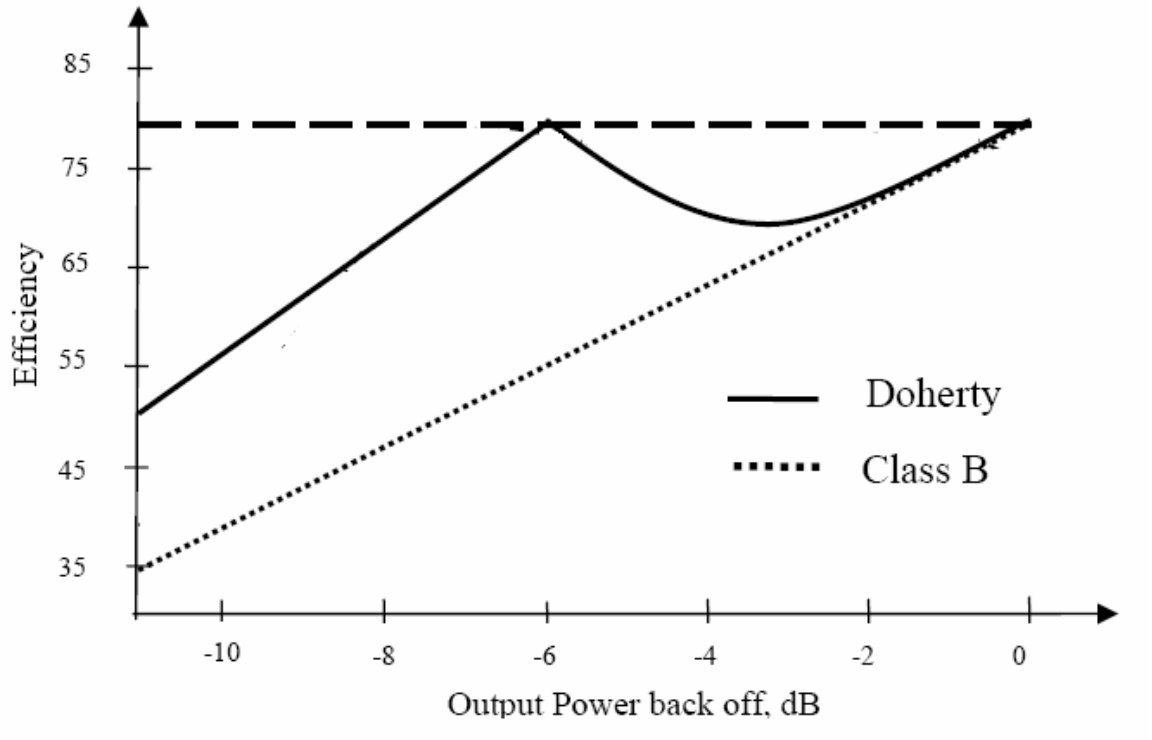


Figure 2.13 Ideal Efficiency Curves of Doherty, Class B and Power Back Off

The advantage of the Doherty PA is that it maintains maximum voltage and efficiency in the up 6-dB regime. This means a given level of linearity at a given level of mean RF power can be achieved using the same device but at a substantially higher efficiency than a conventional configuration. The disadvantage is that the quarter-wave transmission line is too large to fit into the die. Some recent reports are shown in [25, 26]

(b) The Chireix Outphasing Amplifier [27]

This technique is also called linear amplification using nonlinear components LINC [28]. A given amplitude-modulation signal can be separated into two fixed-magnitude phase-modulation signals. The derivation is shown below.

$$\begin{aligned} &\text{if} \quad S_{in} = A(t)\cos(\omega t) \\ &\text{then} \quad S_1 = \cos\{\omega t + \cos^{-1}[A(t)]\} \quad \text{and } S_2 = \cos\{\omega t - \cos^{-1}[A(t)]\} \\ &\text{so that} \quad S_1 + S_2 = S_{in} \end{aligned} \tag{2.25}$$

Because each component of S_1 and S_2 has a fixed amplitude, a nonlinear PA can be used. The system block diagram is shown on Figure 2.14.

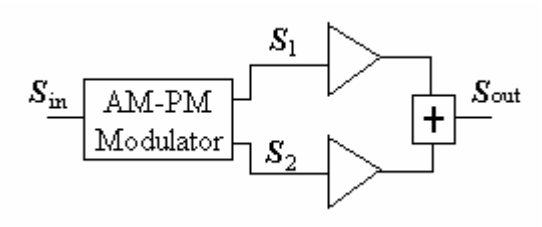


Figure 2.14 Block Diagram of an Outphasing Amplifier

The key component is the phase modulator, which converts the amplitude modulation (AM) into two phase-modulated signals with opposite senses. The combiner recover back the AM signal. The difficulty is how to separate them. Hakala [29] reported such technique design in 2005.

(c) Envelope Elimination and Restoration

EER—together with envelope tracking, polar modulation and polar loop modulation—tries to vary the power-supply voltage at the envelope frequency in order to achieve the highest possible efficiency. The block diagram [30] is given in Figure 2.15

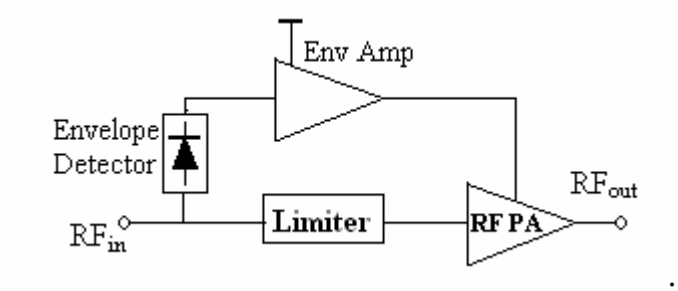


Figure 2.15 EER Block Diagram

The RF signal is separated into phase and amplitude parts by the limiter and the envelope detector. Note that the bandwidth of envelope is the baseband bandwidth and thus much smaller than that of the RF signal. The phase portion of the signal can be amplified by the nonlinear Class E or F PAs. The envelope PA typically involves a high-speed DC–DC converter to modulate the power supply of the nonlinear PA to restore the RF envelope. For a working frequency as high as several MHz, a high-efficiency DC-DC converter is not easy to design.

(d) Dynamic Biasing

The dynamic biasing includes dynamic current biasing (DCB) and dynamic voltage biasing (DVB). The idea is very simple, bias the PA at a lower current or voltage when the input is small. Figure 2.16 gives the concept [31]. When the input signal is smaller, as determined by the envelope detector or a system setting, the PA adjusts the biasing points (voltage, current or both) to reduce the average power consumption. Several reports have used such methods [32-36].

DVB is convenient provided different power supplies are available. Otherwise, a high-speed high-efficiency DC–DC converter is required. DCB is easier to implement, but

gain expansion is a problem.

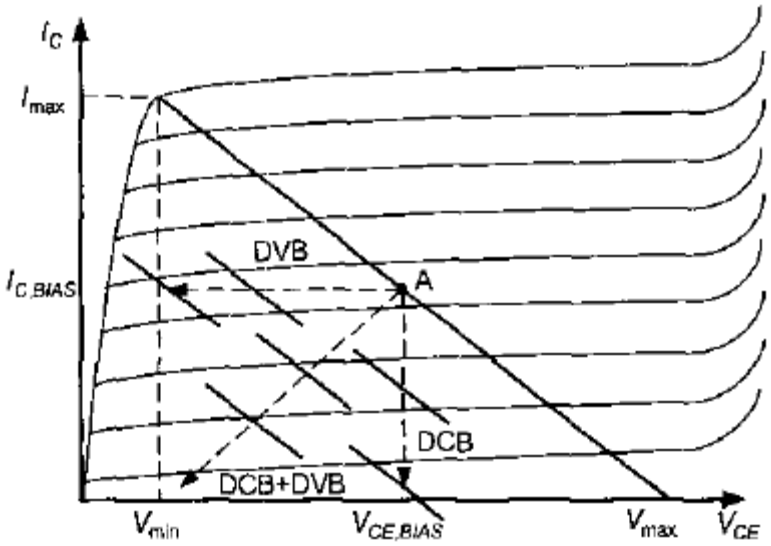


Figure 2.16 DCB and DVB

There are still some other ways to improve the PA's linearity, such as predistortion, feedforward, Cartesian feedback and polar loop [37]. Since efficiency can be a trade-off with linearity, it is important to evaluate the design before using linearization techniques. Overall efficiency is improved if linearity is held constant. However, those methods need to know the PA's characteristics in advance, or to build sophisticated control loops into the baseband circuits; and are not suitable for stand-alone PA designs, particularly for handset which has severe cost constraint.

References

[1] H. T. Friis, "Noise figure of radio receiver," *Proc. IRE*, vol. 32, pp. 419-422, July 1944.

[2] A. A. Abidi, G. J. Pottie, and W. J. Kaiser, "Power-conscious design of wireless circuits and systems," *Proc. IEEE*, vol. 88, pp. 1528-1545, Oct. 2000.

[3] D. K. Shaeffer and T. H. Lee, "A 1.5V, 1.5-GHz CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745-759, May 1997.

[4] B. Wang, J. R. Hellums, and C. G. Sodini, "MOSFET thermal noise modeling for analog integared circuits," *IEEE J. Solid-State Circuits*, vol. 29, pp. 833-835, July 1994.

[5] T. H. Lee, *The Design of CMOS Radio Frequency Integrate Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[6] D. J. Cassan and J. R. Long, "A 1-V transformer-feedback low-noise amplifier for 5-Ghz wireless LAN in 0.18-um CMOS," *IEEE J. Solid-State Circuits*, vol. 38, pp. 427-435, Mar. 2003.

[7] J. R. Long, "Monolithic transformers for silicon RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1368-1382, Sept. 2000.

[8] Y. Aoki, N. Hayama, M. Fujii, and H. Hida, "A 23/3-dB dual-gain low-noise amplifier for 5-GHz-band wireless applications," *IEEE GaAs Digest*, 2002.

[9] H.-W. Chui and S.-S. Lu, "A 2.17 dB NF, 5 GHz band monolithic CMOS LNA with 10 mW DC power consumption," *Symp. on VLSI Circuitd digest of*

Technical papers, 2002.

[10] R.-C. Liu, C.-R. Lee, H. Wang, and C.-K. Wang, "A 5.8-GHz two-stage high-linearity low-voltage low noise amplier in a 0.35-um CMOS technology," *Symp. Radio Frequency Intergated Circuits (RFIC)*, 2002.

[11] C. J. Debono, F. Maloberti, and J. Micallef, "A 1.8 GHz CMOS low-noise amplifier," *8th Intenal. conf. on Electronics, Circuits and Systems*, 2001.

[12] B. Chi and B. Shi, "A 1.8V 2.4 GHZ CMOS on-chip impedance matching low noise amplifier for WLAN applications," *Proc. of the 2003 Int. Symp. on Circuits and Systems*, 2003.

[13] J. Kaukovuori, M. Hotti, J. Ryynanen, J. Jussila, and K. Halonen, "A linearized 2-GHz SiGe low noise amplifier for direct conversion receiver," *Proc. of the 2003 Int. symp. on Circuits and Systems*, 2003.

[14] T.-K. Nguyen and S.-G. Lee, "A 5.2 GHz, 1.7 dB NF low noise amplifier for wireless LAN based on 0.18 /spl mu/m CMOS technology," *Proc. of 2002 3rd Int. Conf. on Microwave and Millimeter Wave Technology*, 2002.

[15] N. O. Sokal and A. D. Sokal, "Class E, a new class of high-efficiency tuned single-ended power amplifiers," *IEEE J. Solid-State Circuits*, vol. 10, pp. 168-176, June 1975.

[16] S. Xie, V. Paidi, R. Coffie, S. Keller, S. Heikman, B. Moran, A. Chini, S. P. Denbaars, U. Mishra, S. Long, and M. J. W. Rodwell, "High-linearity Class B power amplifiers in GaN hemt technology," *IEEE Microwave and Wireless*

Components Lett., vol. 13, pp. 284-286, July 2003.

[17] T. Sowlati and D. M. W. Leenaerts, "A 2.4-GHz 0.18-um CMOS self-biased cascode power amplifier," *IEEE J. Solid-State Circuits*, vol. 38, pp. 1318-1324, Aug. 2003.

[18] C.-C. Yen and H.-R. Chuang, "A 0.25-um 20-dBm 2.4-GHz CMOS power amplifier with an integrated diode linearizer," *IEEE Microwave and Wireless Components Lett.*, pp. 45-47, Feb. 2003.

[19] A. Shirvani, D. K. Su, and B. A. Wooley, "A CMOS RF power amplifier with parallel amplification for efficient power control," *IEEE J. Solid-State Circuits*, vol. 37, pp. 684-693, June 2002.

[20] M. Berkhout, "An integrated 200-W Class-D audio amplifier," *IEEE J. Solid-Sate Circuits*, vol. 38, pp. 1198-1206, July 2003.

[21] F. Fortes and M. J. d. Rosario, "A second harmonic Class-F power amplifier in standard CMOS technology," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, pp. 1216-1220, June 2001.

[22] K. L. R. Mertens and M. S. J. Steyaert, "A 700-MHz 1-W fully differential CMOS Class-E power amplifier," *IEEE J. Solid-State Circuits*, vol. 37, pp. 137-141, Feb. 2002.

[23] W. H. Doherty, "A new high efficiency power amplifier for modulated waves," *Proc. IRE*, vol. 24, pp. 1163-1182, Sept. 1936.

[24] V. Viswanathan, "Efficeincy enhancement of base station power amplifiers using

Doherty technique," in *M.S. Thesis, Dept. of Electrical Engineering*: Virginia Polytechnic Institute and State University, Feb., 2004.

[25] Y. Yang, J. Cha, B. Shin, and B. Kim, "A microwave Doherty amplifier employing envelope tracking technique for high efficiency and linearity," *IEEE Microwave and Wireless Components Lett.*, vol. 13, pp. 370-372, Sept. 2003.

[26] N. Srirattana, A. Raghavan, D. Heo, P. E. Allen, and J. Laskar, "Analysis and design of a high-effiency multistage Doherty power amplifier for wireless communications," *IEEE Trans. Microwave Theory and Techniques,* vol. 53, pp. 852-860, Mar. 2005.

[27] H. Chireix, "High power outphasing modulation," *Proc. IRE,* vol. 23, pp. 1370-1392, Nov. 1995.

[28] D. C. Cox and R. P. Leck, "A VHF implementation of a LINC amplifier," *IEEE Trans. Commun.*, vol. COM-24, pp. 1018-1022, Sept. 1976.

[29] I. Hakala, D. K. Choi, L. Gharavi, N. Kajakine, J. Koskela, and R. Kaunisto, "A 2.14-GHz Chireix outphasing transmitter," *IEEE Trans. on Microwave Theory and Techniques,* vol. 53, pp. 2129-2138, June 2005.

[30] J. Chen, K. U-yen, and J. S. Kenney, "An envelope elimination and restoration power amplifier using a CMOS dynamic power supply circuit," *2004 MTT-S International*, Jun. 2004.

[31] L. Larson, P. Asbeck, and D. Kimball, "Next generation power amplifiers for wireless communications - squeezing more bits out of fewer joules," *Radio*

Frequency integrated Circuits (RFIC) Symposium, June 2005

[32] N. Srirattana, P. Sen, H. M. Park, C. H. Lee, P. E. Allen, and J. Laskar, "Linear RF CMOS power amplifier with improved efficiency and linearity in wide power levels," *Radio Frequency integrated Circuits (RFIC) Symposium, 2005*, June 2005

[33] W.-C. Hua, H.-H. Lai, P.-T. Lin, C. W. Liu, T.-Y. Yang, and G.-K. Ma, "High-linearity and temperature-insensitive 2.4 GHz SiGe power amplifier with dynamic-bias control," *Radio Frequency integrated Circuits (RFIC) Symposium*, June 2005

[34] Y. W. Kim, K. Han, S. Hong, and J. Shin, "A 45% PAE/18mA quiescent current CDMA PAM with a dynamic bias control circuit," *2004 IEEE Radio Frequency Integrated Circuits Symp.*, 2004.

[35] Y. Yang, "High efficiency CDMA power amplifier with dynamic current control circuits," *IEEE Compound Semicondector Integrated Circuit, digest*, 2004.

[36] N. Schlumpf, M. Declercq, and C. Dehollain, "A fast modulator for dynamic supply linear RF power amplifier," *IEEE J. Solid-State Circuits*, vol. 39, pp. 1015-1025, July 2004.

[37] S. C. Cripps, *RF power amplifier for wireless communications*. Boston: Artech House, 1999.

Chapter 3 Optimize the Two-Stage Low-Noise Amplifier

3.1 Two-Stage LNA Topology and Noise Sources

The LNA is the first active device in a typical RF receiving system. It must provide input matching, usually at 50 Ω, add the least mount of noise, have reasonable gain and give reasonable linearity. Much work has been done on the single-stage LNA to get the minimal noise figure [1, 2]. All the work so far concentrates on getting minimal *NF* under gain constriction or power constriction. However, in some applications, one stage may not meet the requirements of both low noise figure and high linearity. A two-stage LNA is necessary in this situation. Currently, there are no papers dealing with the requirements of both low noise and high linearity (i.e., dynamic range). This work addresses this issue. The objective is, under certain linearity and power restrictions, to optimize circuits to minimize *NF*.

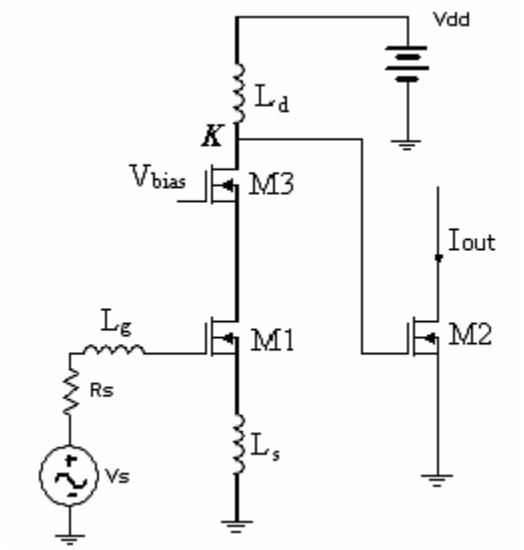


Figure 3.1 A Typical Two-Stage LNA

A high-frequency amplifier usually employs a cascode topology to minimize the Miller effect. We are only concerned with narrowband signals, for which inductive degeneration is usually the best choice for power savings and superior *NF*. Figure 3.1 shows a typical two-stage LNA topology. Input matching is generated by L_s, L_g and the parasitic C_{gs}.

The thermal noise is the dominate noise for a MOSFET. The small-signal MOSFET noise model is shown in Figure 3.2 [2, 3]. In additional to the traditional channel noise, there is gate noise from two sources: the finite conductivity of poly and metal, modeled by R_{geltd}, and the induced gate noise, modeled as R_{gch}. Induced gate noise is due to the fact that the fluctuating channel potential couples capacitively into the gate terminal, leading to a noisy gate current. Because gate noise and drain noise comes from the same source, they are partially correlated. In Figure 3.2, such noise is separated into the uncorrelated part and the correlated part.

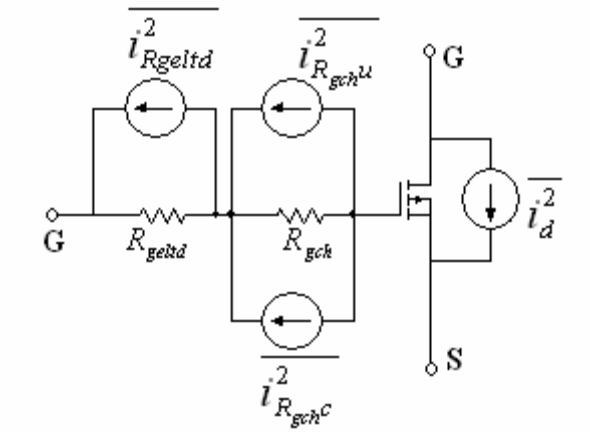


Figure 3.2 Noise Model of a CMOS Transistor

Total gate resistance is the summation of R_{geltd} and R_{gch}, given by

$$R_g = R_{geltd} + R_{gch}. \tag{3.1}$$

The gate noise from R_{geltd} is given by

$$\overline{i_{R_{geltd}}^2} = 4kT\frac{1}{R_{geltd}}$$
$$R_{geltd} = R_{eltd}\left(\alpha\frac{W}{L} + \beta\right) \tag{3.2}$$

where α models the gate electrode's distributed effect and β models the external gate resistance.

The noise from R_{gch} can be calculated by

$$R_{gch} = \frac{1}{\delta G_{ch}} \tag{3.3}$$

δ is a parameter for the distributed effect, approximately 12 to 15. G_{ch}, is given by

$$G_{ch} = G_{st} + G_{ed}. \tag{3.4}$$

G_{st} models the static channel conductance; $G_{st} = I_d/V_{dsat}$ in the saturation region and V_{dsat} is a parameter given by

$$V_{dsat} = \frac{V_{od}}{1 + \frac{V_{od}}{L\varepsilon_{sat}}} \text{ with } V_{od} = V_{gs} - V_t\,. \tag{3.5}$$

ε_{sat} is the velocity-saturation field strength and V_t is the threshold voltage.

G_{ed} models the excess-diffusion conductance and can be calculated by

$$G_{ed} = \eta\frac{kT}{q}\mu_{eff}C_{ox}\frac{W}{L} \tag{3.6}$$

where η is a fit constant roughly equal to 1.

Considering the short-channel effect, the CMOS I–V relationship can be approximately expressed as

$$I_d = \frac{1}{2}\mu_{eff}C_{ox}\frac{W}{L}\frac{V_{od}^2}{1 + \frac{V_{od}}{L\varepsilon_{sat}}} \tag{3.7}$$

So that G_{ch} is given by

$$G_{ch} = \frac{I_d}{V_{dsat}} + \eta \frac{kT}{q} \mu_{eff} C_{ox} \frac{W}{L} = \mu_{eff} C_{ox} \frac{W}{L} (\frac{V_{od}}{2} + \eta \frac{kT}{q}) . \tag{3.8}$$

If the transistor is biased in strong inversion, since $\eta \approx 1$ and $kT/q \approx 26$ mV at room temperature, $V_{od} >> \eta kT/q$, G_{ch} is roughly equal to

$$G_{ch} = \frac{I_d}{V_{dsat}} + \eta \frac{kT}{q} \mu_{eff} C_{ox} \frac{W}{L} = \mu_{eff} C_{ox} \frac{W}{L} \frac{V_{od}}{2} \tag{3.9}$$

To make the calculation simple, the small-signal CMOS model, shown in Figure 3.3 (a), can be converted to the equivalent circuit in Figure 3.3 (b), provided the following requirement is satisfied.

$$R_{gch} << 1/\omega_0 C_{gs} \text{ i.e. } \frac{\delta G_{ch}}{\omega_0 C_{gs}} >> 1 \tag{3.10}$$

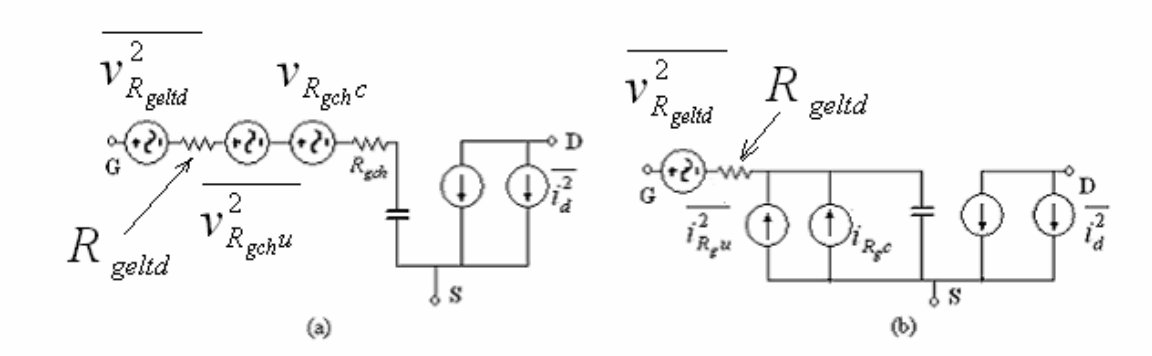


Figure 3.3 (a) CMOS Noise Model, (b) Equivalent Circuit

Since $G_{ch} \approx g_m$ and $\delta g_m/\omega_0 C_{gs} = \delta\omega_T/\omega_0 >> 1$, Equation (3.10) is satisfied in this analysis. The noise source is summarized in Equation (3.11), where c is the correlation factor.

$$\overline{v_{R_{geltd}}^2} = 4kTR_{geltd} \qquad \overline{i_d^2} = 4kT\gamma G_{ch} \quad (\gamma \sim 1.5)$$

$$\overline{v_{R_{gch}u}^2} = \frac{4kT\,(1-|c|^2)}{\delta G_{ch}} \quad v_{R_{gch}c} = \sqrt{\frac{4kT}{\delta G_{ch}}} \cdot c \tag{3.11}$$

$$\overline{i_{R_g u}^2} = \frac{4kT \cdot \omega^2 C_{gs}^2 \cdot (1-|c|^2)}{\delta G_{ch}} \qquad i_{R_g c} = \sqrt{\frac{4kT}{\delta G_{ch}}} \cdot \omega C_{gs} \cdot c$$

To put the transistor model into the circuit configuration, the small-signal two-stage LNA model is shown in Figure 3.4. R_g is the parasitic series resistance of L_g, which is part of the input-matching network. $\overline{v_l^2} = 4kTR_g$ is the noise source due to R_g. The same is true for source impedance (R_s. and v_s here). The effect of *M3* in Figure 3.1 is small and is omitted in this analysis. The inductor L_d resonates with all capacitances at the node *K*, and R_d is the equivalent parallel resistor, which is roughly equal to $Q_d \omega L_d$ provided Q_d is large enough, Q_d is the inductor's *Q*. Most capacitance at node *K* comes from *M2*'s gate-source capacitor, the rest comes from *M3*'s drain. Usually, *M3* is much narrower than the gate of *M2* and its contribution is neglected in the following calculation. All noise sources come from R_g, *M1*, R_L and *M2*. The following gives the *NF* calculation due to those noise sources.

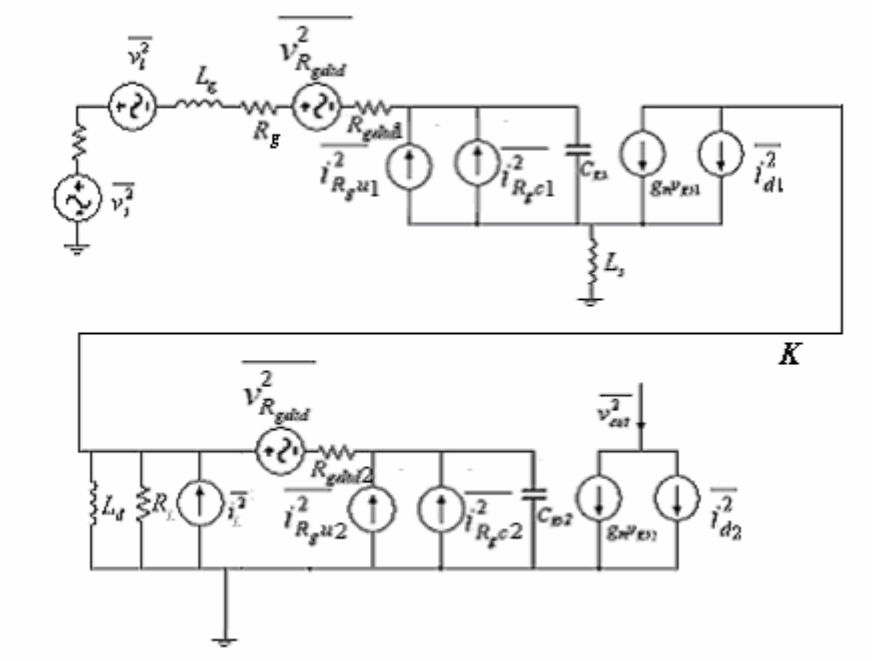


Figure 3.4 Small-Signal Model for a Two-Stage LNA

3.2 Noise Figure Calculation

First, the LNA must provide 50-Ω input matching. The input impedance is equal to

$$R_s = Z_{in} = s(L_s + L_g) + \frac{1}{sC_{gs1}} + \frac{g_m}{C_{gs1}} L_s \approx \omega_T L_s$$
$$\text{at resonance frequency, i.e. } \omega_0{}^2 = \frac{1}{C_{gs1}(L_g + L_s)} \tag{3.12}$$

The short-channel CMOS I–V relationship is given by Equation (3.7), so the transistor's transconductance is calculated by differentiating this expression:

$$g_m = \frac{\partial I_d}{\partial V_{od}} = \frac{1}{2}\mu_{eff} C_{ox} \frac{W}{L} \rho(L\varepsilon_{sat}) \frac{2+\rho}{(1+\rho)^2} = G_{ch} \frac{2+\rho}{(1+\rho)^2} \qquad \rho = \frac{V_{od}}{L\varepsilon_{sat}}. \tag{3.13}$$

Since R_{geltd} scales with the number of fingers, careful layout will minimize the effects of this resistor and thus is neglected in following analysis.

Assuming $R_{geltd} << 1/\omega_0 C_{gs}$, the first-stage, second-stage and total transconductance of the LNA are given by

$$\begin{aligned} G_{m1} &= \frac{g_{m1}}{\omega_0 C_{gs1}(R_s + \omega_T L_s)} = \frac{\omega_T}{2\omega_0 R_s} \\ G_{m2} &= g_{m2} \\ G_{m-total} &= G_{m1} \cdot R_L \cdot G_{m2} \end{aligned} \tag{3.14}$$

The output noise power density due to the 50-Ω source is

$$S_{a,src}(\omega_0) = S_{src}(\omega_0) G_{m-total}^2 = \frac{kT\omega_{T1}^2}{\omega_0^2 R_s} \cdot R_L^2 \cdot g_{m2}^2 . \tag{3.15}$$

The output noise from R_g is similar:

$$S_{a,R_g}(\omega_0) = S_{src}(\omega_0) G_{m-total}^2 = \frac{kT\omega_{T1}^2 R_g}{\omega_0^2 R_s^2} \cdot R_L^2 \cdot g_{m2}^2 . \tag{3.16}$$

Gate noise is partially correlated; it is easier to calculate this noise by separating it into two parts; one is totally uncorrelated and the other is fully correlated. The noise due to the uncorrelated part can be calculated directly from the power level and is given by

$$\begin{aligned} S_{a,R_{gl}u}(\omega_0) &= \frac{g_{m1}^2 (R_s^2 + \frac{1}{\omega_0^2 C_{gs1}^2})}{4\omega_0^2 C_{gs1}^2 R_s^2} \cdot (R_L g_{m2})^2 \cdot 4kTR_{gch1}\omega_0^2 C_{gs1}^2 (1 - |c|^2) \\ &= kTR_{gch1}(1 - |c|^2) g_{m1}^2 (1 + \frac{1}{\omega_0^2 C_{gs1}^2 R_s^2}) \cdot (R_L g_{m2})^2 \end{aligned} \tag{3.17}$$

The noise from the correlated part must be calculated carefully with respect of the direction. c is negative in the direction shown in the Figure 3.4. The amplitudes of the two noise sources are first added and then made square to get the total output noise. The correlated part of the noise is given by Equation (3.18). Here both the noise from the gate and the noise from the channel are included.

$$S_{a,R_{gl}c,i_d}(\omega_0) = kT\{g_{m1}^2 R_{gch1}(1+\frac{1}{\omega_0^2 G_{gs1}^2 R_s^2})|c|^2 + \gamma G_{ch1} + 2g_{m1}\sqrt{R_{gch1}\gamma G_{gch1}}\cdot\left[\frac{1}{R_s\omega_0 C_{gs}}\mathrm{Re}(c) - \mathrm{Im}(c)\right]\}\cdot(R_L g_{m2})^2 \quad (3.18)$$

Adding (3.17) and (3.18), the output noise from *M1* is

$$S_{a,M1}(\omega_0) = kT\{g_{m1}^2\frac{1}{\delta G_{ch1}}(1+\frac{1}{\omega_0^2 G_{gs1}^2 R_s^2}) + \gamma G_{ch1} + 2g_{m1}\sqrt{\frac{\gamma}{\delta}}\cdot\left[\frac{1}{R_s\omega_0 C_{gs}}\mathrm{Re}(c) - \mathrm{Im}(c)\right]\}\cdot(R_L g_{m2})^2 \quad (3.19)$$

To lower the cost and reduce variation, it is attractive to integrate the inter-stage matching network. However, in the CMOS process, the Q of an on-die spiral inductor is poor and the parasitic resistor of the inductor cannot be neglected. Assuming Q of L_d is Q_d, the equivalent parallel resistor and noise current are

$$R_L = (Q_d + \frac{1}{Q_d})\cdot\omega L_d \qquad \overline{i_L^2} = 4kT/R_L \quad (3.20)$$

The output noise due to R_L is

$$S_{a,R_L} = 4kTR_L g_{m2}^2 . \quad (3.21)$$

The noise from transistor *M2* can be calculated as well:

$$S_{a,M2} = \overline{i_{R_{g2},u}^2}\cdot R_L^2\cdot g_{m2}^2 + \left|i_{R_{g2}c}\cdot R_L\cdot g_{m2} + i_{d2}\right|^2 = \frac{4kT}{\delta G_{ch2}}(\omega_0 C_{gs2} g_{m2} R_L)^2 + 4kT\gamma G_{ch2} + 8kT\sqrt{\frac{\gamma}{\delta}}\omega_0 C_{gs2} g_{m2} R_L\cdot\mathrm{Re}(c) \quad (3.22)$$

Because at node K, the capacitance comes mainly from the gate-source capacitor C_{gs2}, the capacitance from the *M3* junction and the *M2* gate-drain are neglected (the Miller effects

can be minimized by employ a cascode topology as the first stage). At the resonance frequency, we have

$$\omega_0 L_d = \frac{1}{\omega_0 C_{gs2}} \text{ and } R_L = (Q + 1/Q)\omega_0 L_d = \frac{Q + 1/Q}{\omega_0 C_{gs2}}$$

$$g_{m2} R_L = g_{m2}(Q_d + \frac{1}{Q_d}) \cdot \omega_0 L_d = (Q_d + \frac{1}{Q_d}) \frac{g_{m2}}{\omega_0 C_{gs2}} = (Q_d + \frac{1}{Q_d}) \frac{\omega_{T2}}{\omega_0} \tag{3.23}$$

$$\therefore \quad \omega_0 C_{gs2} g_{m2} R_L = (Q_d + \frac{1}{Q_d}) g_{m2} \tag{3.24}$$

And the Equation (3.22) becomes

$$S_{a,M2} = \frac{4kT}{\delta G_{ch2}} g_{m2}^2 (Q_d + \frac{1}{Q_d})^2 + 4kT\gamma G_{ch2} + 8kT\sqrt{\frac{\gamma}{\delta}} g_{m2} (Q_d + \frac{1}{Q_d}) \cdot \mathrm{Re}(c) \tag{3.25}$$

The *NF* is

$$\begin{aligned} NF &= \frac{S_{a,src} + S_{a,R_g} + S_{a,M1} + S_{a,R_L} + S_{a,M2}}{S_{a,src}} = 1 + \frac{R_g}{R_s} \\ &+ R_s \left(\frac{\omega_0}{\omega_{T1}} \right)^2 \left(\gamma G_{ch1} + \frac{g_{m1}^2}{\delta G_{chq}} \left(1 + \frac{1}{(\omega_0 C_{gs1} R_s)^2} \right) + 2 g_{m1} \sqrt{\frac{\gamma}{\delta}} \left(\frac{\mathrm{Re}(c)}{\omega_0 C_{gs1} R_s} - \mathrm{Im}(c) \right) \right) \\ &+ 4 \left(\frac{\omega_0}{\omega_{T1}} \right)^2 \frac{\omega_0 C_{gs2} R_s}{(Q_d + 1/Q_d)} + 4 R_s \left(\frac{\omega_0}{\omega_{T1}} \right)^2 \left(\frac{\gamma G_{ch2}}{g_{m2}^2 R_L^2} + \frac{(\omega_0 C_{gs2})^2}{\delta G_{ch2}} + 2\sqrt{\frac{\gamma}{\delta}} \frac{\omega_0 C_{gs2}}{g_{m2} R_L} \mathrm{Re}(c) \right) \end{aligned} \tag{3.26}$$

According to (3.13),

$$g_m = G_{ch} \frac{2 + \rho}{(1 + \rho)^2}, \tag{3.27}$$

so the noise figure becomes

$$
\begin{aligned}
NF &= \frac{S_{a,src} + S_{a,R_g} + S_{a,M1} + S_{a,R_L} + S_{a,M2}}{S_{a,src}} = 1 + \frac{R_g}{R_s} \\
&+ \gamma R_s g_{m1} \left(\frac{\omega_0}{\omega_{T1}} \right)^2 \left(\begin{array}{l} \dfrac{(1+\rho_1)^2}{2+\rho_1} + \dfrac{2+\rho_1}{\gamma\delta(1+\rho_1)^2} \left(1 + \dfrac{1}{(\omega_0 C_{gs1} R_s)^2} \right) \\ + 2\sqrt{\dfrac{1}{\gamma\delta}} \left(\dfrac{\mathrm{Re}(c)}{\omega_0 C_{gs1} R_s} - \mathrm{Im}(c) \right) \end{array} \right) \\
&+ 4 \left(\frac{\omega_0}{\omega_{T1}} \right)^2 \frac{\omega_0 C_{gs2} R_s}{(Q_d + 1/Q_d)} \\
&+ 4\gamma R_s g_{m2} \left(\frac{\omega_0}{\omega_{T1}} \right)^2 \left(\frac{\omega_0}{\omega_{T2}} \right)^2 \left(\begin{array}{l} \dfrac{(1+\rho_2)^2}{(Q_d + 1/Q_d)^2 (2+\rho_2)} + \dfrac{2+\rho_2}{\gamma\delta(1+\rho_2)^2} \\ + 2\sqrt{\dfrac{1}{\gamma\delta}} \dfrac{\mathrm{Re}(c)}{Q_d + 1/Q_d} \end{array} \right)
\end{aligned}
\tag{3.28}
$$

3.3 Linearity Estimation

LNAs require not only a low *NF*, but high linearity, which is the ability to handle a large input signal. Poor linearity degrades the transceiver system's performance and increases the bit error rate. The linearity can be characterized as the third-order IIP_3, defined in chapter 2.

Any transistor's I–V characteristics can be written as the small-signal expansion near the operation point.

$$I_d = V(V_{od}) = V(V_0 + v) = V_0 + c_1 v + c_2 v^2 + c_3 v^3 + \cdots \tag{3.29}$$

In the strong-biasing condition, the CMOS I–V characteristic is approximated by Equation (3.13), and the values of c_1 and c_3 are given by

$$c_1 = \frac{1}{2}\mu_{eff} C_{ox} \frac{W}{L} \frac{V_0^2}{(1+\rho)} \frac{2L\varepsilon_{sat} + V_0}{V_0(L\varepsilon_{sat} + V_0)}$$
$$c_3 = \frac{1}{2}\mu_{eff} C_{ox} \frac{W}{L} \frac{V_0^2}{(1+\rho)} \frac{(L\varepsilon_{sat})^2}{V_0^2(L\varepsilon_{sat} + V_0)^3} \tag{3.30}$$

V_0 is the DC biasing voltage, v is AC signal. Since the output voltage is drain current multiplied by the load of the impedance, the amplitude of the IIP_3 can be calculated.

$$A_{IIP3}^2 = \left|\frac{4}{3}\frac{c_1}{c_3}\right| = \frac{4}{3}\frac{(2L\varepsilon_{sat} + V_0)(L\varepsilon_{sat} + V_0)^2 V_0}{(L\varepsilon_{sat})^2} = \frac{4}{3}(\rho^2 + 2\rho)(1+\rho)^2(L\varepsilon_{sat})^2 \tag{3.31}$$

This IIP_3 is based on the common-source connection where the input is applied to the gate and the source. In the configuration of Figure 3.4, the first-stage voltage across the gate source is amplified by the input Q, Q_{input},

$$Q_{input} = \frac{1}{\omega_0 C_{gs1} R_s} \tag{3.32}$$

Compared to the input, the second-stage gate-source voltage is amplified by $G_{M1}R_L$

The total IIP_3 is given by

$$\frac{1}{A_{IIP3,tot}^2} = \frac{Q_{input}^2}{A_{IIP3,M1}^2} + \frac{(G_{M1}R_L)^2}{A_{IIP3,M2}^2} = \frac{1/(\omega_0 C_{gs1} R_s)^2}{\frac{4}{3}(L\varepsilon_{sat})^2(1+\rho_1)^2(\rho_1^2 + 2\rho_1)}$$
$$+ \frac{\left(\frac{\omega_{T1}(Q_L + 1/Q_L)}{2\omega_0^2 R_s C_{gs2}}\right)^2}{\frac{4}{3}(L\varepsilon_{sat})^2(1+\rho_2)^2(\rho_2^2 + 2\rho_2)} \tag{3.33}$$

For the short-channel MOSFET, the overlap capacitances are no longer negligible. In fact, the overlap capacitance can be three to four times the intrinsic capacitance [4], $2/3C_{ox}WL$. The following calculation assumes $C_{gs} \approx 3 \cdot C_{intr} = 2C_{ox}WL$. The cutoff frequency is

$$\omega_T \approx \frac{g_m}{C_{gs}} = \frac{\mu_{eff} C_{ox} \frac{W}{L} \rho L \varepsilon_{sat} \frac{1+\rho/2}{(1+\rho)^2}}{2C_{ox}WL} = \frac{\mu_{eff} \rho \varepsilon_{sat}}{2L} \frac{1+\rho/2}{(1+\rho)^2} = \frac{v_{sat}}{2L} \frac{2\rho+\rho^2}{(1+\rho)^2} \tag{3.34}$$

The saturation velocity is

$$v_{sat} = \frac{\mu_{eff} \varepsilon_{sat}}{2}, \tag{3.35}$$

and the total power consumption is

$$\begin{aligned} P_{tot} &= V_{cc}(I_{d1} + I_{d2}) = \frac{1}{2} V_{cc} \mu_{eff} C_{ox} \frac{(L\varepsilon_{sat})^2}{L} \left(\frac{\rho_1^{\,2}}{1+\rho_1} W_1 + \frac{\rho_2^{\,2}}{1+\rho_2} W_2 \right) \\ &= V_{cc} C_{ox} L \varepsilon_{sat} v_{sat} \left(\frac{\rho_1^{\,2}}{1+\rho_1} W_1 + \frac{\rho_2^{\,2}}{1+\rho_2} W_2 \right) \end{aligned} \tag{3.36}$$

3.4 Optimize the Noise Figure

Putting (3.28), (3.33) and (3.36) together, there are three parameters that can be played with, W_1, W_2 and ρ_1. ρ_2 is know since *M2* is biased such that V_{od} is V_{cc} - V_t. V_t is the threshold voltage.

$$NF = NF(W_1, W_2, \rho_1) = NF(P_{tot}, A_{IIP3,tot}, \rho_1) \tag{3.37}$$

Fixing P_{tot} and $A_{IP3,tot}$, we can optimize the *NF* under the power-constrained situation. Since the equation is so complex, numerical calculations will be made in following optimization procession. The transistor model is based on TSMC 0.35 μm technology. Equation (3.38) lists the parameters used in the calculation [1, 5]. Assume the LNA works at 2.4 GHz.

$$
\begin{aligned}
&L_{eff} = 0.25\mu m, \quad t_{ox} = 7.9nm \quad and \quad V_t = 0.553V \\
&\varepsilon_{sat} = 4.7\times10^{6}\,\text{V/m},\ v_{sat} = 10^{5}\,\text{m/s},\ \ \delta = 14,\ \ \gamma = 1.5 \\
&V_{cc} = 3.3\text{V} \quad Q_d = 4 \quad \omega_0 = 1.508\times10^{10}\,/\,\text{s} \quad R_s = 50\Omega
\end{aligned}
\tag{3.38}
$$

So that

$$
\begin{aligned}
&C_{ox} = \frac{\varepsilon_{SiO}\varepsilon_0}{t_{ox}} = 4.6\times10^{-3}\ (\text{pF/um}^2) \qquad g_m = 4.7\times10^{-4} W(\mu\text{m})\rho\frac{2+\rho}{(1+\rho)^2} \\
&C_{gs} = 2C_{ox}WL = 2.3\times10^{-3} W\ (\text{pF}) \qquad L\varepsilon_{sat} = 1.175\,(\text{V})
\end{aligned}
\tag{3.39}
$$

and the $A_{IP3,tot}$ and NF become

$$
A^2_{IIP3,tot} = \frac{1}{\dfrac{1.73\times10^5}{W_1^2(1+\rho_1)^2(2\rho_1+\rho_1^2)} + \dfrac{1.3743\times10^8(2\rho_1+\rho_1^2)^2}{W_2^2(1+\rho_1)^4(1+\rho_2)^2(2\rho_2+\rho_2^2)}}
\tag{3.40}
$$

$$
\begin{aligned}
NF &= 1+\frac{R_g}{R_s} + 2.00402\times10^{-4}\frac{(1+\rho_1)^2}{\rho 1(2+\rho_1)}W1 \\
&\cdot\left(\frac{(1+\rho_1)^2}{2+\rho_1} + \frac{2+\rho_1}{21(1+\rho_1)^2}\left(1+\frac{3.185\times10^5}{W_1^2}\right) + 0.4364\left(\frac{564.366}{W_1}\text{Re}(c) - \text{Im}(c)\right)\right) \\
&+ 9.48\times10^{-6}\frac{(1+\rho_1)^4}{(2\rho_1+\rho_1^2)^2}W_2 + 2.523\times10^{-7}\frac{(1+\rho_1)^4(1+\rho_2)^2}{(2\rho_1+\rho_1^2)^2(2\rho_2+\rho_2^2)}W_2 \\
&\cdot\left(\frac{(1+\rho_2)^2}{2+\rho_2} + 0.86\frac{2+\rho_2}{(1+\rho_2)^2} + 1.855\,\text{Re}(c)\right)
\end{aligned}
\tag{3.41}
$$

$$
P_{tot} = 1.82243\times10^{-3}\left(\frac{{\rho_1}^2}{1+\rho_1}W_1 + \frac{{\rho_2}^2}{1+\rho_2}W_2\right)
\tag{3.42}
$$

Equation (3.41) can be further simplified if the following condition is satisfied.

$$
\begin{aligned}
&9.48\times10^{-6}\frac{(1+\rho_1)^4}{(2\rho_1+\rho_1^2)^2}W_2 >> 2.523\times10^{-7}\frac{(1+\rho_1)^4(1+\rho_2)^2}{(2\rho_1+\rho_1^2)^2(2\rho_2+\rho_2^2)}W_2 \\
&\cdot\left(\frac{(1+\rho_2)^2}{2+\rho_2} + 0.86\frac{2+\rho_2}{(1+\rho_2)^2} + 1.855\,\text{Re}(c)\right)
\end{aligned}
\tag{3.43}
$$

$$\text{i.e.}\quad 37.6 >> \frac{(1+\rho_2)^2}{(2\rho_2+\rho_2^2)}\cdot\left(\frac{(1+\rho_2)^2}{2+\rho_2}+0.86\frac{2+\rho_2}{(1+\rho_2)^2}+1.855\,\mathrm{Re}(c)\right)$$
$$=\left(1-\frac{1}{2\rho_2+\rho_2^2}\right)\cdot\left(\frac{(1+\rho_2)^2}{2+\rho_2}+0.86\frac{2+\rho_2}{(1+\rho_2)^2}+1.855\,\mathrm{Re}(c)\right)$$

ρ_2 is about 2 in this configuration, so that (3.43) is satisfied and the last term of (3.41) is omitted. It is still too complicated to solve W2 and W1 from (3.40) and (3.42), so another assumption is that linearity is mainly limited by the second stage. That is, the first term in the denominator of Equation (3.40) is much smaller than the second term, so that we can get

$$A_{IIP3,tot}^2=\frac{W_2^2(1+\rho_1)^4(1+\rho_2)^2(2\rho_2+\rho_2^2)}{1.3743\times10^8(2\rho_1+\rho_1^2)^2}. \tag{3.44}$$

This assumption requires

$$\frac{1.73\times10^5}{W_1^2(1+\rho_1)^2(2\rho_1+\rho_1^2)}<<\frac{1.3743\times10^8(2\rho_1+\rho_1^2)^2}{W_2^2(1+\rho_1)^4(1+\rho_2)^2(2\rho_2+\rho_2^2)}$$
$$\text{that is}\quad \frac{W_2^2}{W_1^2}\frac{(1+\rho_1)^2(1+\rho_2)^2(2\rho_2+\rho_2^2)}{(2\rho_1+\rho_1^2)^3}<<794.6 \tag{3.45}$$

We will verify this after the numerical results.

From Equation (3.44), the width of *M2* is

$$W2=1.1723\times10^4\,A_{IIP3,tot}\frac{2\rho_1+{\rho_1}^2}{(1+\rho_1)^2(1+\rho_2)\sqrt{2\rho_2+{\rho_2}^2}} \tag{3.46}$$

Plugging this into Equation (3.42), we can get the width of *M1:*

$$W1=\frac{1+\rho_1}{\rho_1^2}\left(548.72P_{tot}-1.1723\times10^4\,A_{IIP3,tot}\frac{{\rho_2}^2(2\rho_1+{\rho_1}^2)}{(1+\rho_1)^2(1+\rho_2)^2\sqrt{2\rho_2+{\rho_2}^2}}\right) \tag{3.47}$$

In principle, if $A_{IIP3,\mathrm{tot}}$ and P_{out} is given, there will be a optimal ρ_1 that gives the minimal noise figure.

In Figure 3.1, V_{gs2} is set to V_{cc}, and $\rho_2 \approx 2.338$, neglecting the effect of R_g,

$$W2 = 1103 A_{IIP3,tot} \frac{2\rho_1 + \rho_1^2}{(1+\rho_1)^2} \tag{3.48}$$

$$W1 = 548.72 \frac{1+\rho_1}{\rho_1^2} \left(P_{tot} - 3.291 A_{IIP3,tot} \frac{(2\rho_1 + \rho_1^2)}{(1+\rho_1)^2} \right) \tag{3.49}$$

$$\begin{aligned} NF = {} & 1 + 0.11 \frac{(1+\rho_1)^3}{\rho_1^3 (2+\rho_1)} \left(P_{tot} - 3.291 A_{IIP3,tot} \frac{(2\rho_1 + \rho_1^2)}{(1+\rho_1)^2} \right) \\ & \cdot \left(\frac{(1+\rho_1)^2}{2+\rho_1} + \frac{2+\rho_1}{21(1+\rho_1)^2} \left(1 + \frac{3.185 \times 10^5}{W_1^2} \right) + 0.4364 \left(\frac{564.366}{W_1} \mathrm{Re}(c) - \mathrm{Im}(c) \right) \right) \\ & + 0.01 A_{IIP3,tot} \frac{(1+\rho_1)^2}{2\rho_1 + \rho_1^2} \end{aligned} \tag{3.50}$$

The correlation factor should be given before further simplification. For a long-channel transistor in strong inversion, the correlation factor is

$$c = -j \cdot 0.395 \tag{3.51}$$

And for short-channel case, there is no determined conclusion. In [6], it is shown that both the real and imaginary parts of c are frequency dependent. The real part is negative and tends to be zero at low frequency, and the imaginary part is positive and tends to increase when reducing the channel length. The results in [7] show real part is zero and imaginary part is slightly smaller than the long-channel value, typically ranging from -0.35 to -0.3 for $0.1 \le f/f_t \le 0.5$. In the [3], c is given by $0.4\angle$-120.

Following is the discussion of different coherence factors. Two P_{IIP3} scenarios are considered, -30 dBm and -20 dBm refer to 50 Ω.

A. $c = -j \cdot 0.395$ for long channel MOSFET

Assume the input interception point is -30 dBm. That is, $A_{IIP3,\mathrm{tot}}$ is 0.01 V.

$$NF = 1 + 0.11\frac{(1+\rho_1)^3}{\rho_1^3(2+\rho_1)}\left(P_{tot} - 0.03291\frac{(2\rho_1+\rho_1^2)}{(1+\rho_1)^2}\right)$$
$$\cdot\left(\frac{(1+\rho_1)^2}{2+\rho_1} + 0.1724\right.$$
$$\left.+\frac{2+\rho_1}{21(1+\rho_1)^2}\left(1+\frac{1.0578}{\left(\frac{1+\rho_1}{\rho_1^2}\left(P_{tot} - 0.03291A_{IIP3,tot}\frac{(2\rho_1+\rho_1^2)}{(1+\rho_1)^2}\right)\right)^2}\right)\right) \quad (3.52)$$
$$+0.0001\frac{(1+\rho_1)^2}{2\rho_1+\rho_1^2}$$

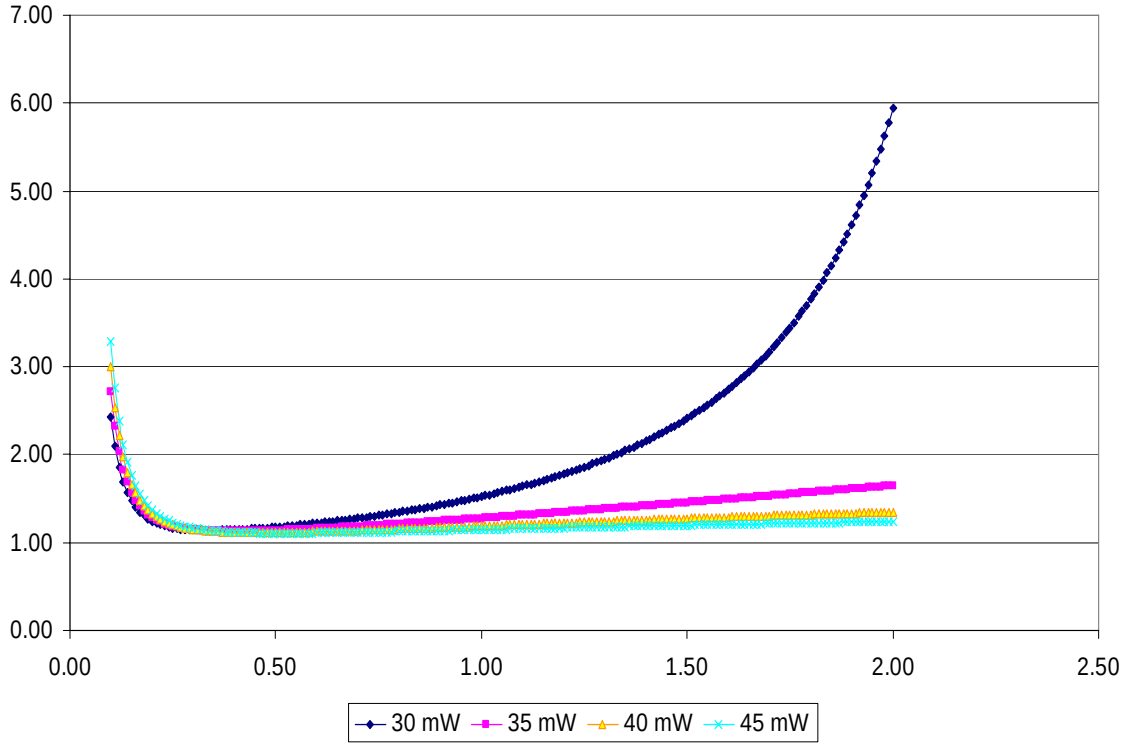


Figure 3.5 NF Under Different Power Restrictions ($P_{IIP3,\text{tot}}$ = -30 dBm)

Figure 3.5 shows *NF* under different power constrictions. Table 3.1 lists the minimal *NF* and corresponding W_1, W_2 and ρ_1. According to the results listed in Table 3.1, Equation

(3.45) is satisfied. Because the gain is above 40 dB, so that the output third-order interception point (OIP_3) is more than 10 dBm.

Table 3.1 Minimal *NF* and Corresponding W_1, W_2 and ρ_1 ($P_{IIP3,\text{tot}}$ = -30 dBm)

P_{tot} (mW)	NF_{min}	ρ_1	W_1 (μm)	W_2 (μm)	Gain (dB)
30	1.14 (0.56 dB)	0.34	98.0	4.9	43.6
35	1.12 (0.49 dB)	0.40	90.7	5.4	44.4
40	1.11 (0.45 dB)	0.46	85.3	5.9	45.1
45	1.11 (0.45 dB)	0.52	81.2	6.3	45.7

If we want higher linearity, to, for example, increase the input interception point by 10 dBm, Figure 3.6 shows *NF* if the interception point is -20 dBm. That is, $A_{IIP3,\text{tot}}$ is 0.0316 V. Table 3.2 gives the correspondent W_1, W_2 and ρ_1. Obviously, the more power consumed, the lower the achievable *NF*.

Table 3.2 Minimal *NF* and Corresponding W_1, W_2 and ρ_1 ($P_{IIP3,\text{tot}}$ = -20 dBm)

P_{tot} (mW)	NF_{min}	ρ_1	W_1 (μm)	W_2 (μm)	Gain (dB)
100	1.10 (0.41 dB)	0.62	82.3	21.6	46.5
105	1.09 (0.37 dB)	0.66	80.9	22.2	46.7
110	1.09 (0.37 dB)	0.71	77.3	22.9	47.0
115	1.09 (0.37 dB)	0.76	74.5	23.6	47.2

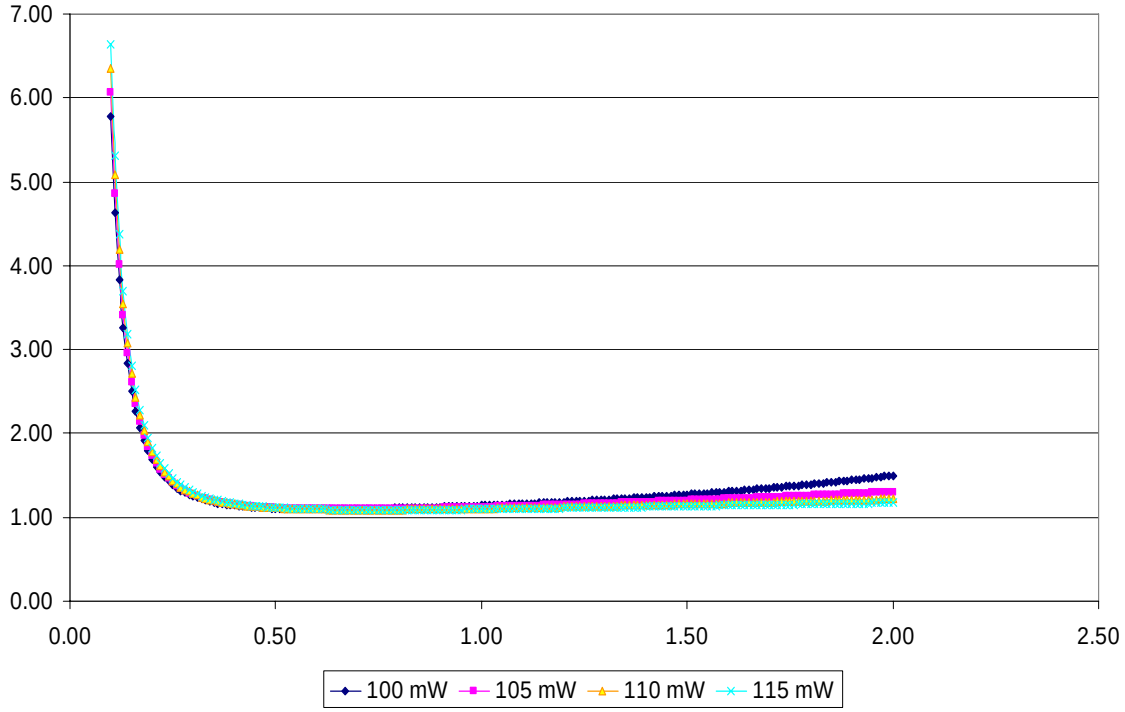


Figure 3.6 NF Under Different Power Restrictions ($P_{IIP3,\text{tot}}$ = -20 dBm)

B. $c = -j \cdot 0.35$ for a short-channel MOSFET

Using the same power constriction and input interception point, *NF* becomes

$$NF = 1 + 0.11\frac{(1+\rho_1)^3}{\rho_1^3(2+\rho_1)}\left(P_{tot} - 3.291A_{IIP3,tot}\frac{(2\rho_1+\rho_1^{\,2})}{(1+\rho_1)^2}\right)$$
$$\cdot\left(\frac{(1+\rho_1)^2}{2+\rho_1} + \frac{2+\rho_1}{21(1+\rho_1)^2}\left(1+\frac{3.185\times10^5}{W_1^2}\right) + 0.15274\right) \qquad (3.53)$$
$$+ 0.01A_{IIP3,tot}\frac{(1+\rho_1)^2}{2\rho_1+\rho_1^2}$$

$A_{IIP3,\,tot}$ = 0.01 V. NF is given by Figure 3.7. Table 3.3 summarizes the corresponding W_1, W_2 and ρ_1. Figure 3.8 and Table 3.4 show the situation when $A_{IIP3,\,tot}$ = 0.03 V

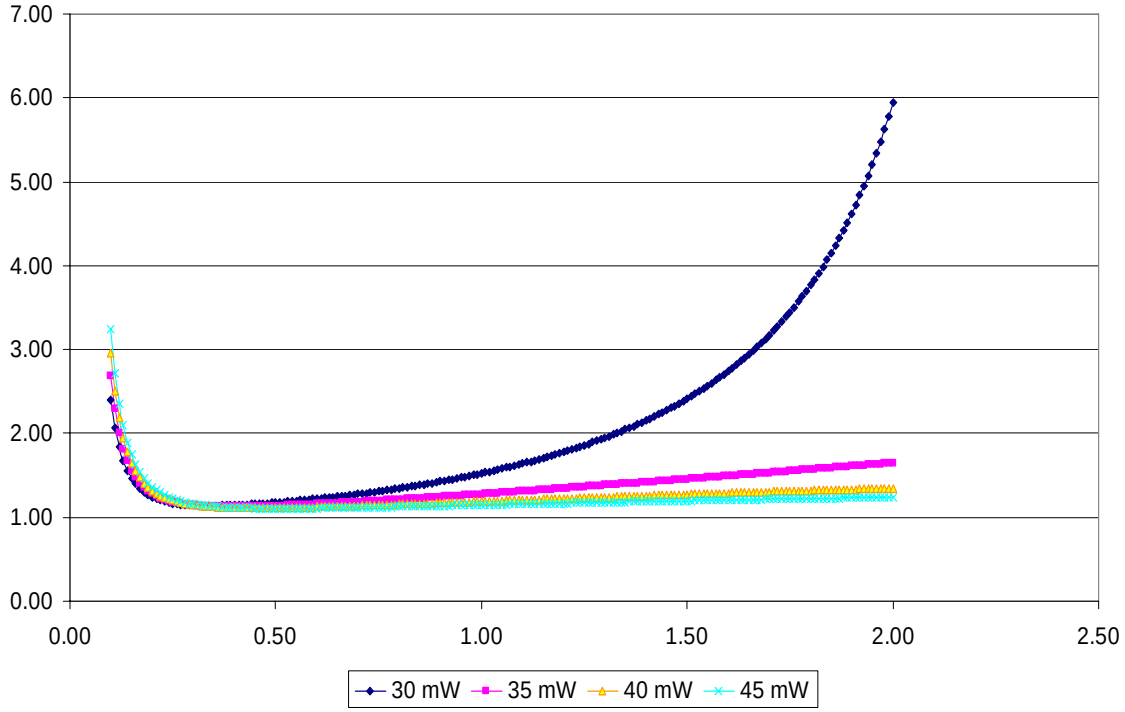


Figure 3.7 NF Under Different Power Restrictions ($P_{IIP3,\text{tot}}$ = -30 dBm)

Table 3.3 Minimal *NF* and Corresponding W_1, W_2 and ρ_1 ($P_{IIP3,\text{tot}}$ = -30 dBm)

P_{tot} (mW)	NF_{min}	ρ_1	W_1 (μm)	W_2 (μm)	Gain (dB)
30	1.14 (0.56 dB)	0.34	98.0	4.9	43.6
35	1.12 (0.49 dB)	0.39	95.9	5.3	44.3
40	1.11 (0.45 dB)	0.45	89.4	5.8	45.0
45	1.11 (0.45 dB)	0.51	84.5	6.2	45.6

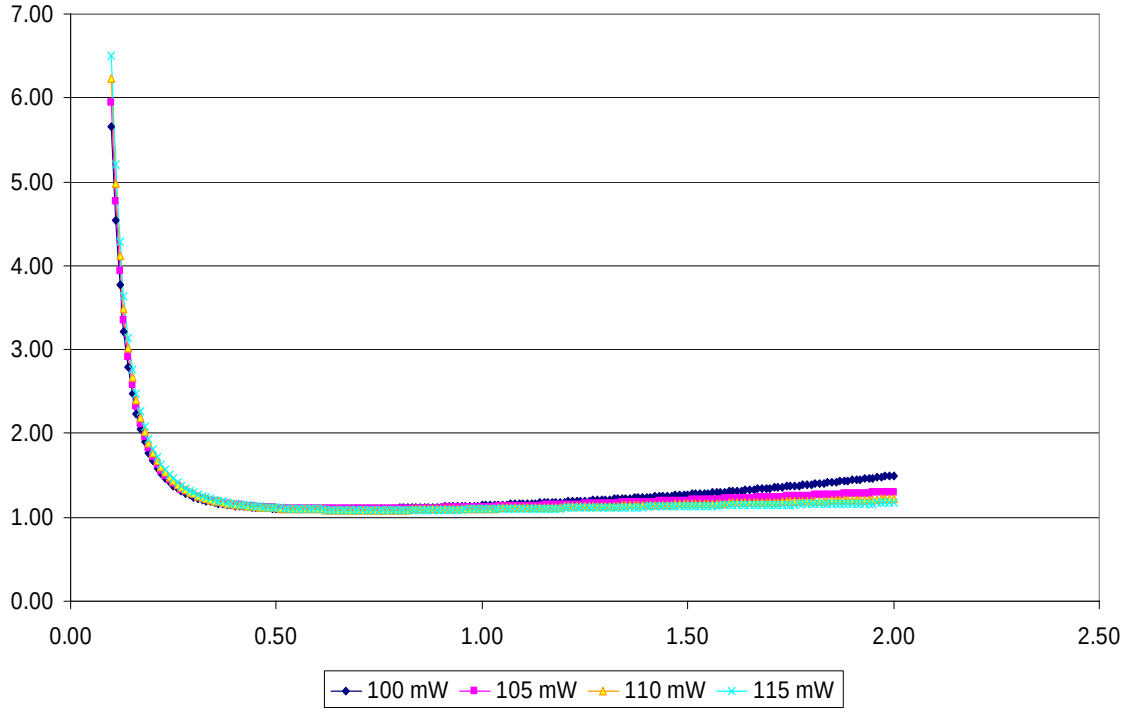


Figure 3.8 NF Under Different Power Restrictions ($P_{IIP3,\text{tot}}$ = -20 dBm)

Table 3.4 Minimal *NF* and Corresponding W_1, W_2 and ρ_1 ($P_{IIP3,\text{tot}}$ = -20 dBm)

P_{tot} (mW)	NF_{min}	ρ_1	W_1 (μm)	W_2 (μm)	Gain (dB)
100	1.10 (0.41 dB)	0.61	85.8	21.4	46.4
105	1.09 (0.37 dB)	0.66	81.0	22.2	46.7
110	1.09 (0.37 dB)	0.71	77.4	22.9	47.0
115	1.09 (0.37 dB)	0.76	74.5	23.6	47.2

C. $c = 0.4\angle -120°$ for short-channel MOSFET

NF is given by (3.51). $A_{IIP3,\ tot}$ = 0.01 V. NF is given by Figure 3.9. Table 3.5 summarizes the corresponding W_1, W_2 and ρ_1. Figure 3.10 and Table 3.6 show the situation when $A_{IIP3,\ tot}$ = 0.03 V.

$$NF = 1 + 0.11\frac{(1+\rho_1)^3}{\rho_1^3(2+\rho_1)}\left(P_{tot} - 3.291A_{IIP3,tot}\frac{(2\rho_1+\rho_1^2)}{(1+\rho_1)^2}\right)$$
$$\cdot\left(\frac{(1+\rho_1)^2}{2+\rho_1} + \frac{2+\rho_1}{21(1+\rho_1)^2}\left(1+\frac{3.185\times10^5}{W_1^2}\right) + 0.4364\left(-\frac{112.8732}{W_1} + 0.34641\right)\right) \quad (3.51)$$
$$+ 0.01A_{IIP3,tot}\frac{(1+\rho_1)^2}{2\rho_1+\rho_1^2}$$

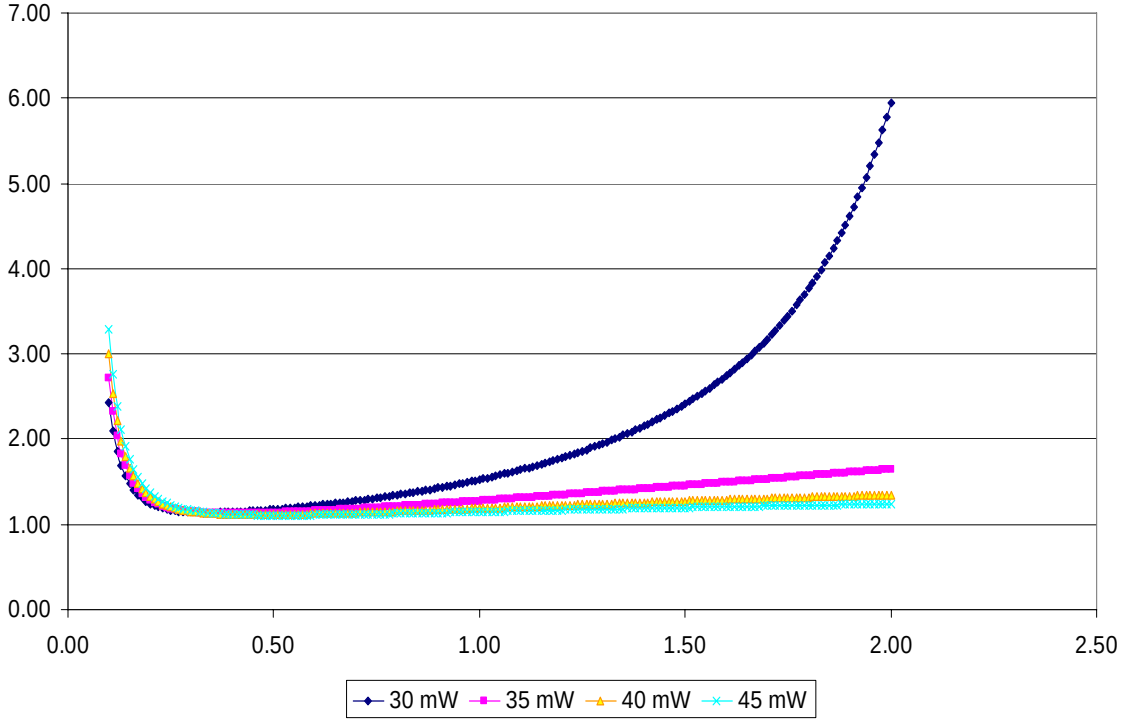


Figure 3.9 NF Under Different Power Restrictions ($P_{IIP3,\text{tot}}$ = -30 dBm)

Table 3.5 Minimal *NF* and Corresponding W_1, W_2 and ρ_1 ($P_{IIP3,\text{tot}}$ = -30 dBm)

P_{tot} (mW)	NF_{min}	ρ_1	W_1 (μm)	W_2 (μm)	Gain (dB)
30	1.14 (0.56 dB)	0.34	98.0	4.9	43.6
35	1.12 (0.49 dB)	0.40	90.7	5.4	44.4
40	1.11 (0.45 dB)	0.46	85.3	5.9	45.1
45	1.11 (0.45 dB)	0.52	81.2	6.3	45.7

Table 3.6 Minimal *NF* and Corresponding W_1, W_2 and ρ_1 ($P_{IIP3,\text{tot}}$ = -20 dBm)

P_{tot} (mW)	NF_{min}	ρ_1	W_1 (μm)	W_2 (μm)	Gain (dB)
100	1.08 (0.33 dB)	0.60	89.3	21.2	46.3
105	1.08 (0.33 dB)	0.65	84.0	22.0	46.7
110	1.08 (0.33 dB)	0.69	82.6	22.7	46.9
115	1.07 (0.29 dB)	0.74	79.1	23.3	47.2

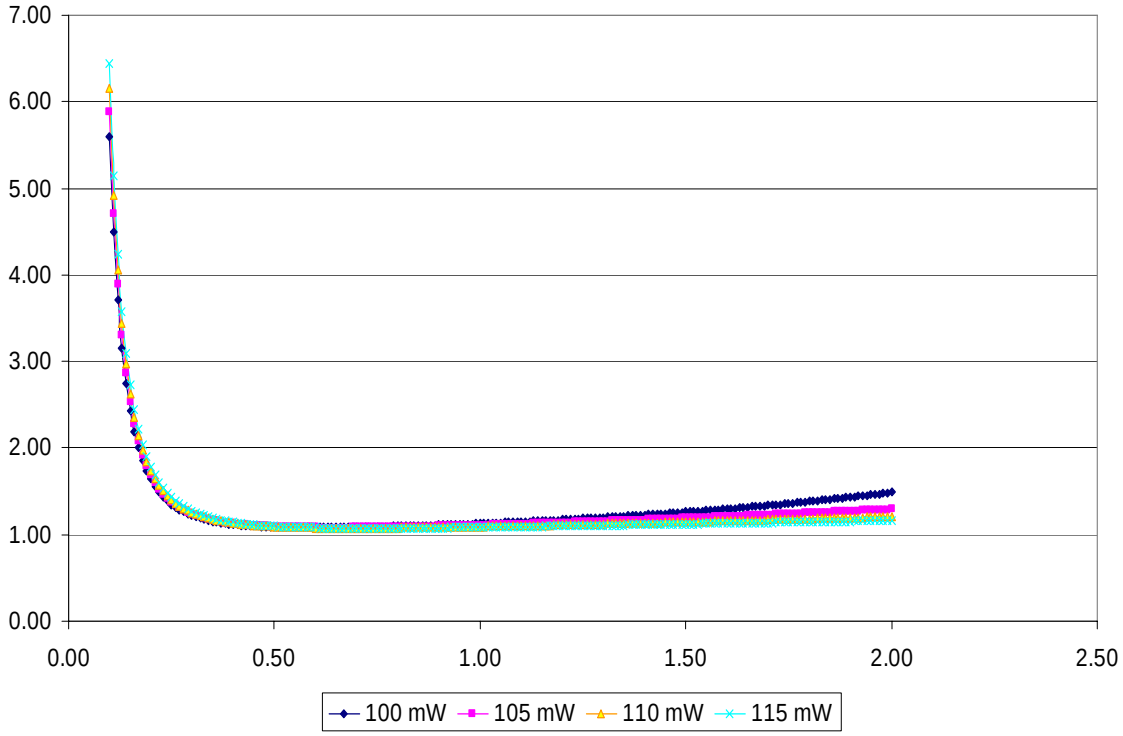


Figure 3.10 NF Under Different Power Restrictions ($P_{IIP3,\text{tot}}$ = -20 dBm)

3.5 Conclusion

From the results shown above, *NF* depends only weakly on the correlation factor. In fact, many models, such as BSIM3, do not include gate noise and its the correlation to the drain noise. The circuit shown in Figure 3.1 does not include the DC biasing network. Current mirror can be used to bias the first stage. The biasing current can be calculated by Equation (3.7) if W_1 and ρ_1 are given.

References

[1] D. K. Shaeffer and T. H. Lee, "A 1.5V, 1.5-GHz CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745-759, May 1997.

[2] P. Andreani and H. Sjoland, "Noise optimization of an inductively degenerated CMOS low noise amplifier," *IEEE Trans. on Circuits and System-II: Analog and Digital Signal Precessing*, vol. 48, Sept. 2001.

[3] J. Ou, X. Jin, P. Gray, and C. Hu, "Recent Developments in BSIM for CMOS RF ac and Noise Modeling," *Advances in Analog Circuit Design*, Nice, France, March 1999.

[4] P. Rossi, F. Svelto, A. Mazzanti, and P. Andreani, "Serendipitous noise reduction in inductively degenerated CMOS RF LNAs," *Proceedings of NORCHIP '03*, Nov. 2003.

[5] https://www.mosis.com.

[6] S. Donati, M. A. Alam, K. S. Krisch, S. Martin, M. R. Pinto, and H. H. Vuong, "Physics-based RF noise modeling of submicron MOSFETs," Dec. 1998.

[7] T. Manku, "RF simulations and physics of the channel noise parameter within MOS transistors," *Proc. IEEE Custom Integrated Circuits Conf.*, May 1999.

Chapter 4 High-Efficiency, High-Linearity PA design

The classification of PAs and several methods to improve linear PA's efficiency were discussed in chapter 2. Of the four methods, DCB is the easiest to realize. One problem with DCB is that the gain drops when the PA biases at low current, decreasing the PA's linearity. One possible solution is to introduce a gain-control circuit to compensate for the gain loss and to improve linearity.

4.1 Dynamic Current Biasing PA with Gain Control

The PA design reported in Chapter 2 works in different modes. The PA can not adjust its biasing point once the mode is selected. This is only suitable for scenarios that require transmitter power to be relatively steady. One way to remove this disadvantage is to implement a power detector to detect the signal level and control the biasing points automatically. The design begins with a conventional Class A amplifier. Figure 4.1 shows a typical Class A PAs working at 2.4 GHz.

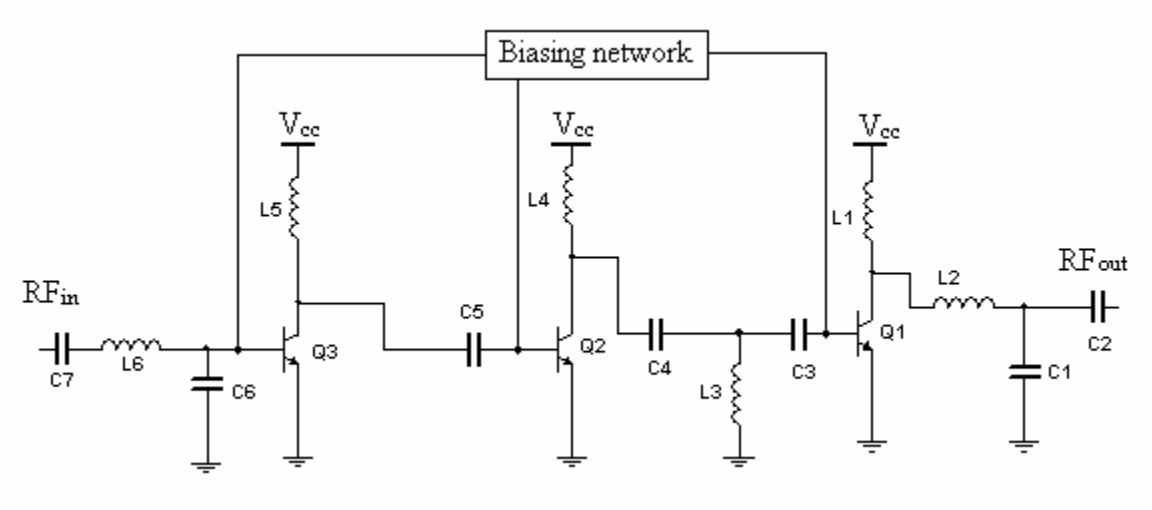


Figure 4.1 A Class A Power Amplifier

C2 and C7 are DC blocking capacitors, C1 and L2 provide output matching, converting the 50-Ω load to the optimal impedance for the maximum output power. C3, C4, C5, C6, L3, L5 and L6 perform interstage and input matching. L1, L4 and L5 are RF chokes. With L1, L4 and L5, the peak voltage of the transistor collectors can be twice that of the power supply, which helps to alleviate the *Q* of the matching network.

In a Class A PA design, the first parameter to be determined is the last-stage biasing current, which is determined by the linearity parameter, such as 1-dB compression point. If the power supply provides 3.3 V and $P_{1\text{-}dB}$ is 30 dBm without loss, the optimal impedance is

$$P_{out} = \frac{V_{cc}^2}{2R_{load}} \tag{4.1}$$

$$R_{load} = \frac{V_{cc}}{I_{cc}} \qquad I_{cc} \text{ is the DC biasing current} \tag{4.2}$$

Therefore, we get

$$R_{load} = \frac{V_{cc}^2}{2} = 5.445\ \Omega \qquad I_{cc} = 606\ \text{mA} \tag{4.3}$$

Considering the knee-down voltage, the load resistor should be smaller, and, correspondingly, the biasing current should be larger.

The RF load is 50 Ω. To get maximum output power, it is necessary to convert the impedance from 50 Ω to 5.445 Ω, which is accomplished by C1 and L2. If applications need more bandwidth, multistage matching may be needed. The major loss of a matching network comes from the low-*Q* on-chip inductance. A lot of commercial PA modules employ off-chip matching, which, usually printing the inductor on the laminate, have much less loss.

Figure 4.2 shows the how matching network works.

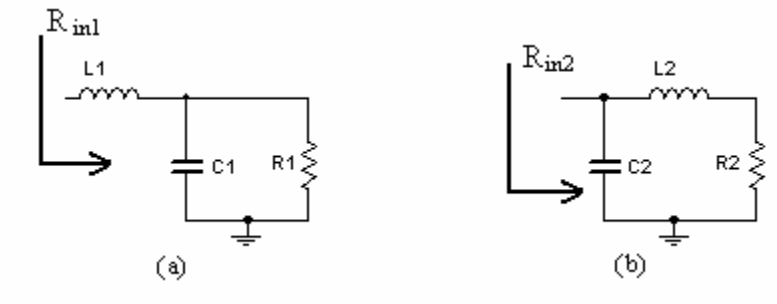


Figure 4.2 Impedance Conversion

The Q plays an important role in the following calculation. The definition of the Q is different for parallel and serial connections. In a parallel connection, such as Figure 4.2 (a),

$$Q_{par} = \frac{R}{X}, \qquad \text{where } X \text{ is the absolute impdence of reactive components.} \tag{4.4}$$

In a serial connection, such as Figure 4.2 (b),

$$Q_{ser} = \frac{X}{R}, \qquad \text{where } X \text{ is the absolute impdence of reactive components.} \tag{4.5}$$

At the working frequency, ω_0, the impedance of the circuit in Figure 4.2 (a) is

$$R_{in1} = j\omega_0 L_1 + \frac{R_1 \cdot \frac{1}{j\omega_0 C_1}}{R_1 + \frac{1}{\omega_0 C_1}} = j\omega_0 L_1 - j\frac{R_1 \cdot \omega_0 R_1 C_1}{1+(\omega_0 R_1 C_1)^2} + \frac{R_1}{1+(\omega_0 R_1 C_1)^2} \tag{4.6}$$

Choosing L_1 can make imaginary part zero, so the impedance becomes

$$R_{in1} = \frac{R_1}{1+(\omega_0 R_1 C_1)^2} = \frac{R_1}{1+Q^2} \qquad Q = \omega_0 R_1 C_1 \tag{4.7}$$

So, the real impedance is decreased by Q^2 if Q is large.

The same procedure can apply to Figure 4.2 (b), the input impedance is

$$R_{in2} = \frac{1}{\frac{1}{j\omega_0 L_2 + R_2} + j\omega_0 C_2} = \frac{1}{\frac{R_2}{R_2 + (\omega_0 L_2)^2} - \frac{j\omega_0 L_2}{R_2 + (\omega_0 L_2)^2} + j\omega_0 C_2} \tag{4.8}$$

Choosing C_2 to remove imaginary part, the real impedance is

$$R_{in2} = \frac{1}{\frac{R_2}{R_2^2 + (\omega_0 L_2)^2}} = R_2(1 + (\frac{\omega_0 L_2}{R_2})^2) = R_2(1 + Q^2) \quad Q = \frac{\omega_0 L_2}{R_2} \tag{4.9}$$

So, the real impedance is increased by Q^2 if Q is large.

In this design, the impedance needs to be transferred from 50 Ω to 5.445 Ω. The matching network in Figure 4.2 (a) should be used, and the values of C1 and L2 in Figure 4.1 are

$$\begin{aligned} C_1 &= 3.8\,\text{pF} \\ L_2 &= 1\ \text{nH} \end{aligned} \tag{4.12}$$

Those values are just crude approximations, the optimal values are determined by fine-tuning to get the best trade-off of linearity and efficiency.

The idea of DCB is to improve efficiency by shutting off part of the PA when the signal is small. This can be realized by controlling either the base or the collector. It is better to control the base biasing for the following reasons.

1) Controlling the collector essentially decreases the effective power supply voltage due to the switch's finite voltage drop. Considering power supply is about 3.3 V, it is important to minimize such phenomena.
2) When the switches are off, the transistor actually lies in the saturation region. For BJT, recovering from saturation to active takes a long time, which will degrade the system linearity during the transition.

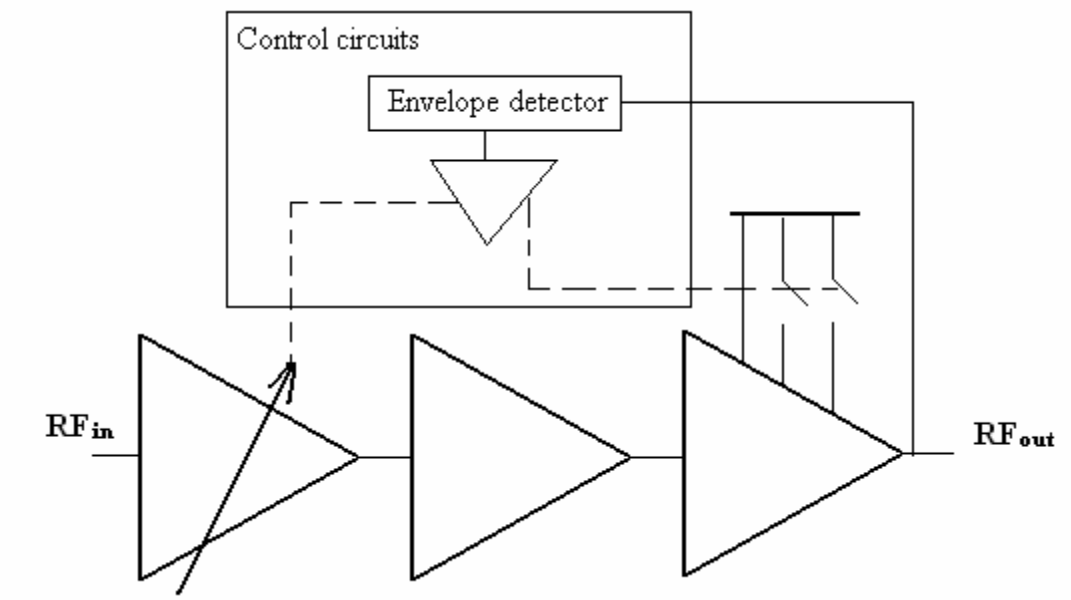


Figure 4.3 Block Diagram of Dynamically Biasing PA

Figure 4.3 presents a block diagram of a dynamically biased PA. The main structure of the PA is roughly the same as Figure 4.1. The last stage is grouped into three blocks whose current can be controlled by control circuits. Depending on the control signal, the current flowing in last stage can be full current, which is the same as the case of the traditional PA. It can also be half or a quarter of the full current. As mentioned, one problem of dynamic biasing is the gain variation. Usually the gain drops when biasing at a low current level, affecting the PA's linearity. To improve linearity, it is better to introduce a gain-compensation scenario. The variable gain stage is added to compensate for the gain loss in Figure 4.3.

The last stage schematic diagram is shown in Figure 4.4. Q3 controls half of the total biasing current of the last stage in Figure 4.3. Q2 controls a quarter and Q1 controls the remaining quarter. If no biasing control is introduced, Q1, Q2 and Q3 work in parallel, and the circuits go back to normal operation. The controlling switches are made with off-chip PMOS. L1 models the emitter's parasitic inductance, which is not negligible in at

high frequencies. Especially for last stage, the collector's impedance is small (< 6 Ω), and neglecting this inductance will cause improper simulation results. L3, L4 and L5 model the bondwire, which is relatively less important due to the large values of R1, R2 and R3. The interstage matching network follows the same procedure, choosing the matching network like that shown in Figure 4.2 (a) or (b) depends on whether the impedance is to be increased or decreased.

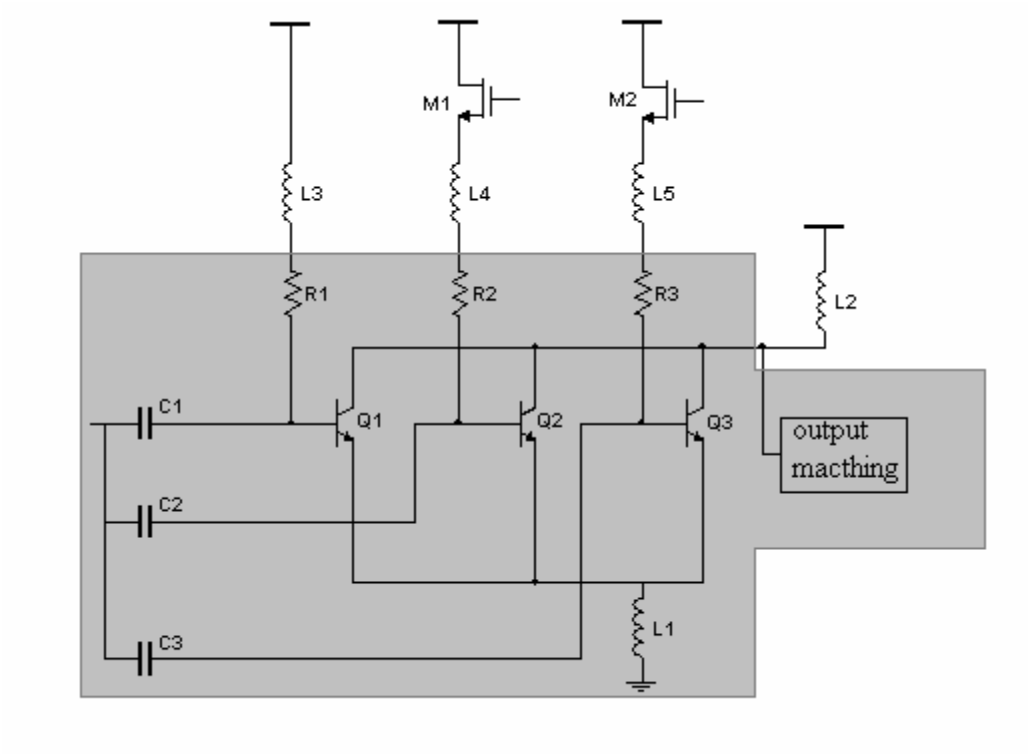


Figure 4.4 Dynamic Current Control Circuits

The gain drop due to DCB is not severe with the current changing by only a small percentage. However, the gain expansion becomes severe and can not be neglected with the current varying by half or three quarters. If only the biasing current is switched, a higher input level gives a greater gain. This will introduce large distortion and degrade the system's linearity. To maintain a flat overall gain and improve linearity, it is better to

introduce a gain compensation circuit. This can be realized by using automatic gain control (AGC) circuits. One of the simplest ways is by controlling the feedback resistance, shown in Figure 4.5.

The first stage is designed in such a way that the feedback resistance is not constant. Two on-chip switches control the equivalent feedback. When the biasing current is low, one or both switches are off: the feedback resistance is greater and so is the gain. This compensates for the gain loss due to the last stage's low biasing current. Off-chip switches are not acceptable here because at a 2.4-GHz working frequency, the bondwire inductance will be too large to neglect (roughly one nanohenry per millimeter). Additionally, such parasitic inductors may cause oscillation. The disadvantage is that the input matching is no longer around 50 Ω, sacrificing some efficiency. Even the input impedance changes, which is a function of the feedback resistor, however, there will be little effect on the overall gain; that is, the gain is insensitive to the input impedance variation. This small efficiency loss can easily be recovered from the huge power saving of the last stage.

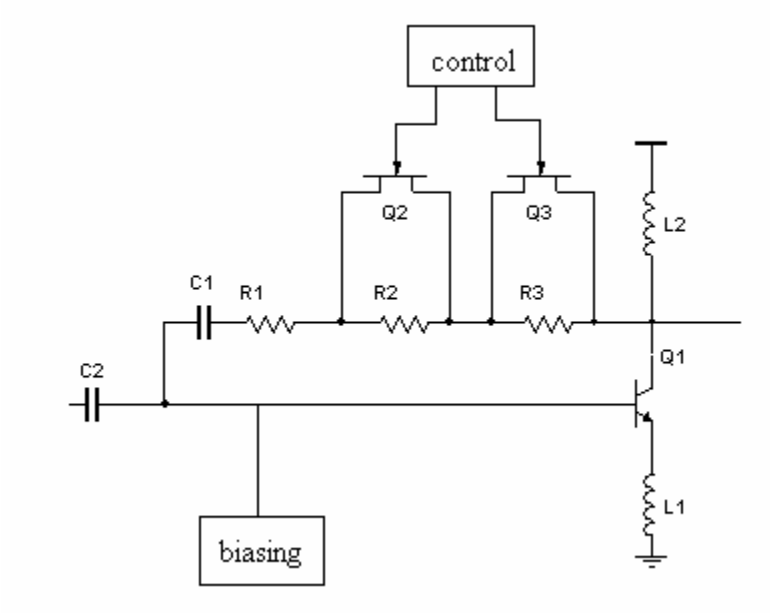


Figure 4.5 Variable Gain Amplifier

4.2 Envelope Detector Design

The envelope detector extracts the baseband amplitude information from the RF source. Functionally, it looks like AM demodulation. A simple envelope detector circuit is shown in Figure 4.6.

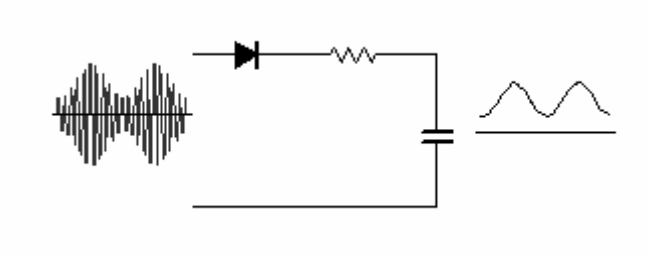


Figure 4.6 A Simple Envelope Detector

The diode only allows the positive signal to pass, so the signal becomes single banded. The low-pass filter is made with a resistor and capacitor. High-frequency RF components are filtered out and the low-frequency envelope information is extracted.

A. Envelope detector core

Unlike field-effect transistors, HBTs process provides good diodes. Figure 4.7 gives the circuit diagram of the envelope detector core.

The detector works as follow. The RF signal is fed to the terminal *In* and the envelope is extracted and sent to the terminal *Out*. C1 is a DC-blocking capacitor, which need not to be very large because the frequency of the RF signal is around 2.4 GHz. The diode-connected Q1 removes the positive parts of the input RF signal. When the RF signal is positive (with refer to the average), the voltage at node *X* will be larger than the quiescent voltage, so Q1 is still on. There is a little more current flowing through Q1 and R2. The voltage at node *X* will be $V_{BE1} + I_{Q1}R_2$, so that voltage at node *X* increases slightly. On the half cycle that the RF signal becomes negative, the voltage at *X* will be smaller than the quiescent voltage, turning Q1 off. Thus, the voltage at node X will track the input. This is equivalent to only passing the negative signal so that the single-banded signal is extracted. The low pass filter consists of R3 and C3, which together get rid of the high-frequency components. Q2 amplifies the extracted envelope signal.

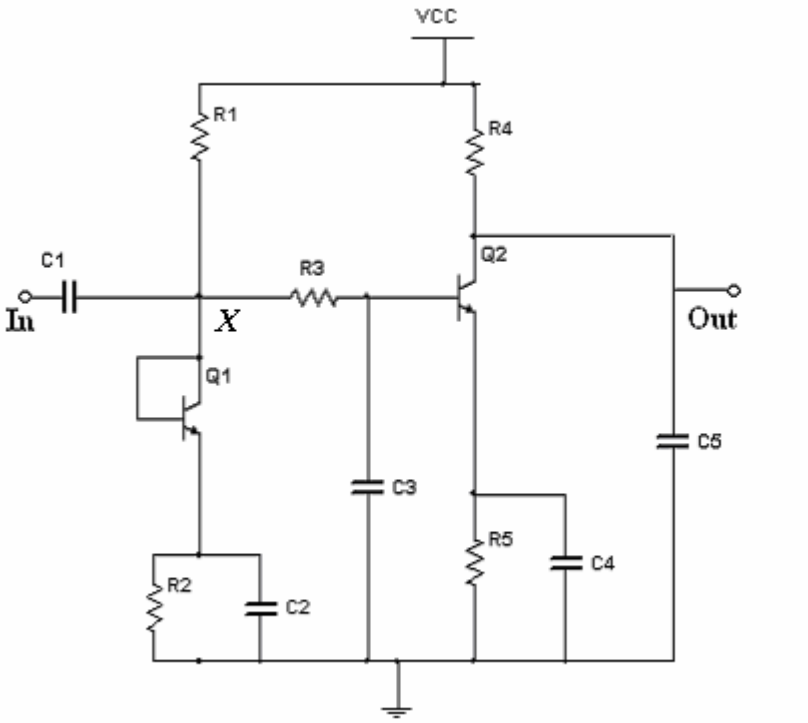


Figure 4.7 Envelope Detector Core

The low-pass filter must provide a time constant large enough to remove the high-frequency RF part, but small enough to leave the low-frequency baseband signal unaffected. In this particular case, the carrier frequency is 2.4 GHz and the baseband bandwidth is less than 20 MHz. so the time constant of low-pass filter should satisfy following.

$$\frac{1}{2.4G} << \tau \equiv R_3 C_3 << \frac{1}{20M} \qquad (4.13)$$
$$0.42ns << \tau << 50ns$$

B. Driver circuits

The signal from the detector core is too weak to drive other circuits. The driver circuit provides the necessary amplification and introduces an external control signal. The driver circuit is given in Figure 4.8. The envelope signal from the detector core is V_{det}. V_{ref} is the

voltage determined by the system algorithm, which determines at which power level the PA begins to switch. In this particular design, we have two switches to be driven, and the input power is set to -22 dBm and -8.5 dBm. This corresponds to a V_{ref} of 1.79 V and 3.0 V respectively. The two-stage differential amplifier amplifies the voltage differences, $\Delta V = V_{det} - V_{ref}$. A large gain will sharpen the rising and falling edges, making the signal more like a square wave. In Figure 4.8, the shaded area is made in CMOS. The CMOS inverters further sharpen the edges and enable the driver to control the nodes. The inverter's rail-to-rail character guarantees the full opening and closing of the switches. Because all transistors of the driver work at low frequency, (2–20 MHz), the parasitic inductance due to the bonding wire is negligible.

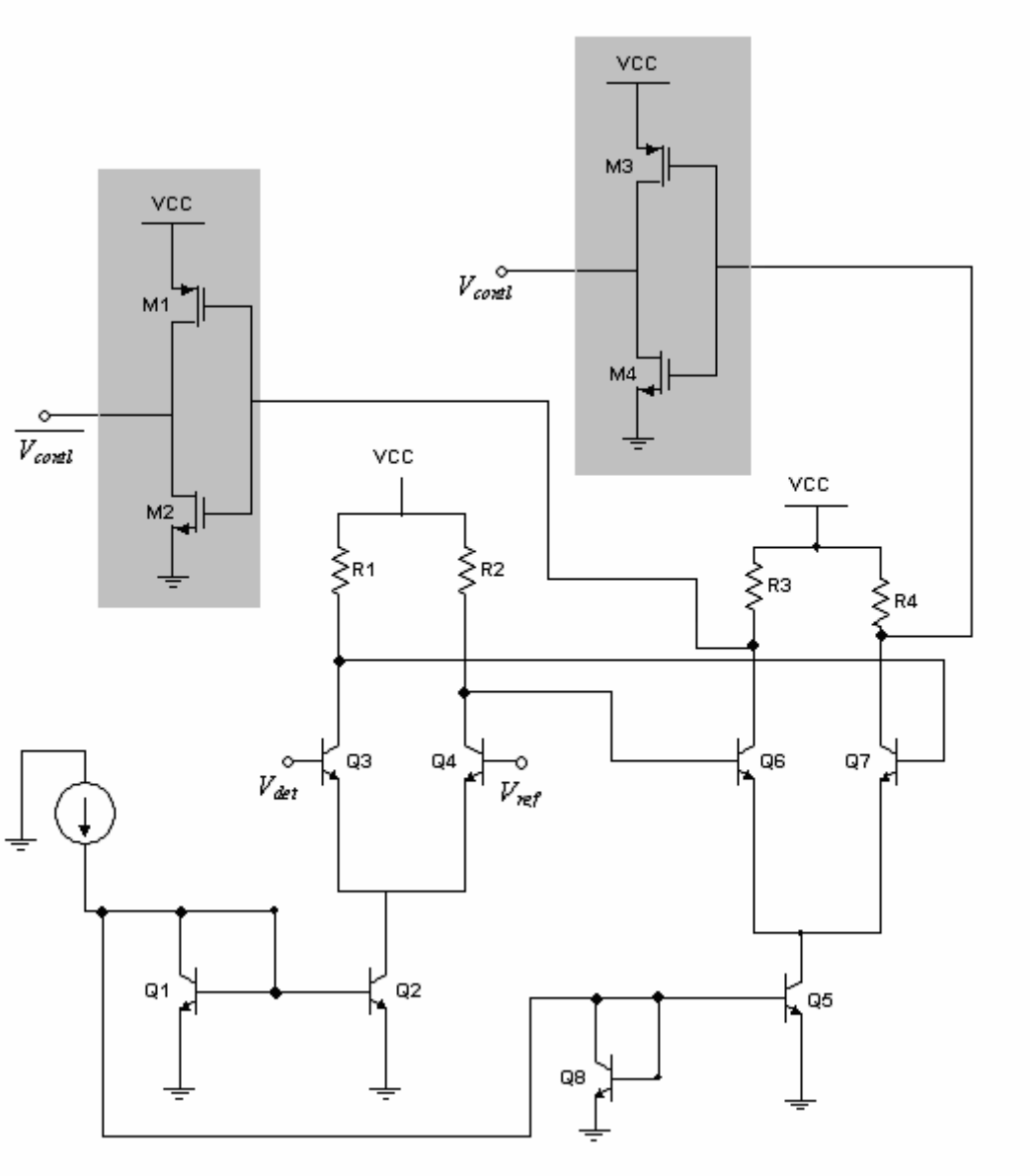


Figure 4.8 Driver Circuit

Chapter 5 Simulation Results and Discussion

This section summarizes the simulation results of the dynamically biased PA shown in chapter 4. The simulator is the Agilent™ Advance Design System (ADS). The circuits' transistor models come from the Anadigics™ HBT and FET model library.

5.1 Harmonic Balance Simulation

Harmonic balance (HB) is a frequency-domain analysis technique for simulating nonlinear circuits and systems and is well-suited to simulate analog RF and microwave circuits. Circuits that are best analyzed using HB simulation include PAs, frequency multipliers, mixers, oscillators, modulators and so on. HB is a steady-state simulation for situations where transient simulation methods are problematic.

The design schematic of the PA is given in chapter 4. There are two switches in the last stage, giving four combinations: full, three quarters, half and quarter current. Figure 5.1 shows the gain versus the input power if only the last stage is switched to save the energy. The green line shows the PA working at full current; the gain is about 30 dB. The red line shows the gain when the biasing current drops to three quarters; the gain decreases by about half a decibel. When the biasing current drops to half of the full current, the PA consumes roughly half the energy compared to the full current mode. However, the gain drops by about one and a half decibels. Further decreasing the current to a quarter drops the gain by four decibels.

A half-decibel gain drop may not cause a problem, but a four-decibel gain variation will definitely degrade the PA's linearity. To increase the linearity, the gain loss must be compensated at low biasing levels. The first stage is designed as a variable-gain amplifier.

The gain curve with compensation is plotted in Figure 5.2. There is no gain compensation for the three-quarters case because the gain loss is within reasonable limits.

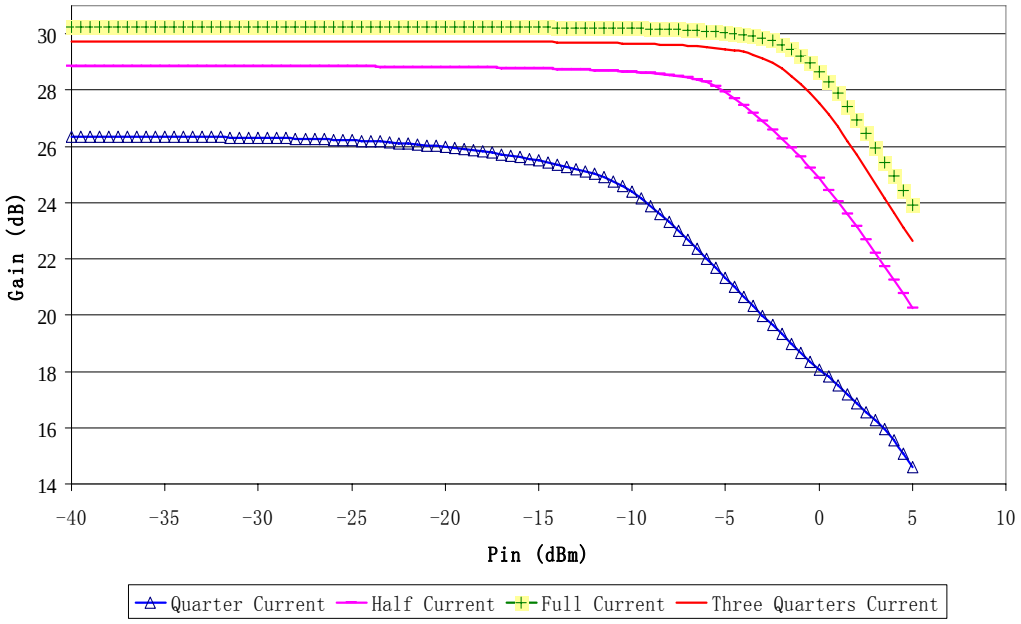


Figure 5.1 Gain Versus Input at Different Biasing Conditions

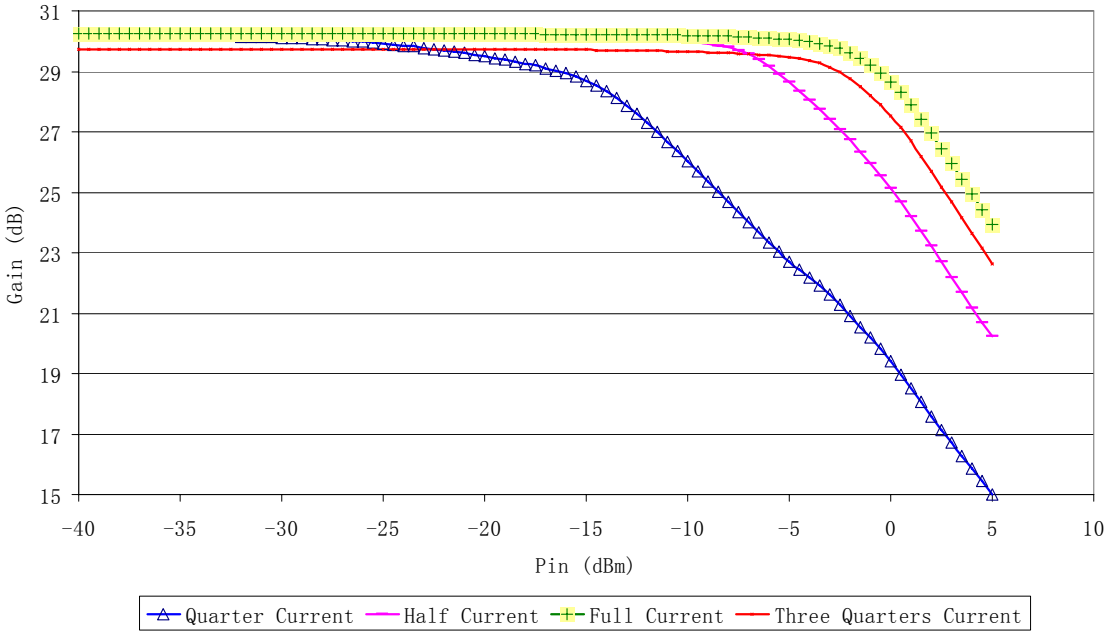


Figure 5.2 Gain Versus Input with Gain Compensation

Switching the biasing point and simultaneously providing gain compensation can meet both high linearity and high efficiency requirements. The gain compensation alleviates distortion due to the gain variation. Figure 5.3 gives the gain curve with the input power, showing that the gain variation is now less than one decibel.

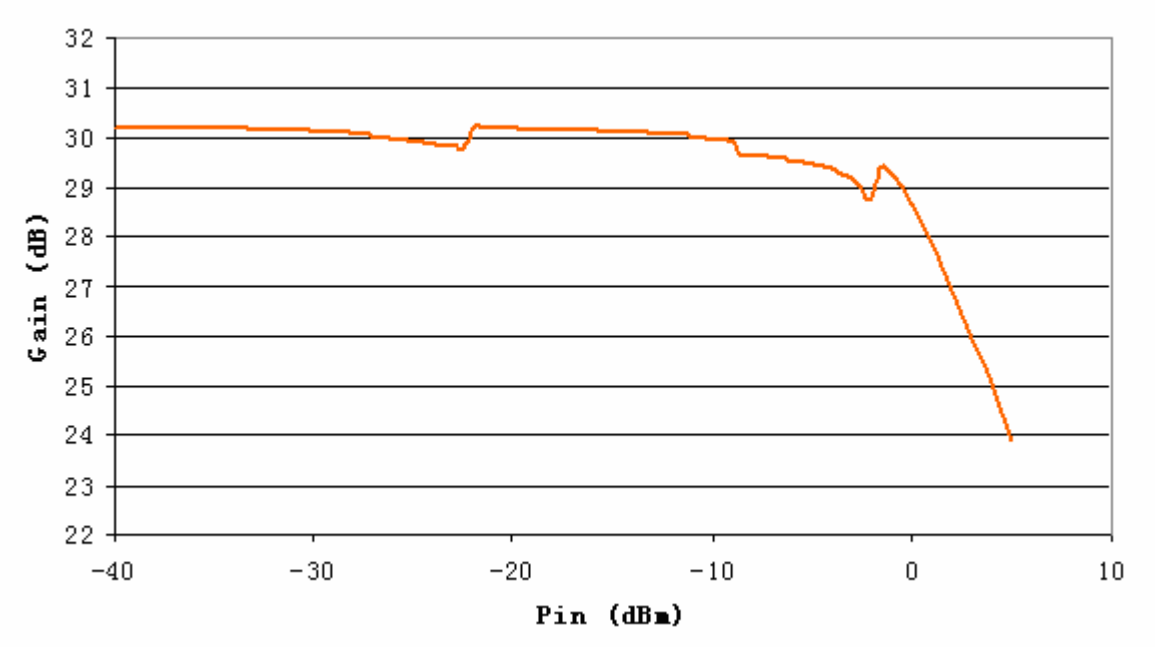


Figure 5.3 Combined Gain Versus Input

$P_{1\text{-}dB}$ and IIP_3 are usually used to characterize the linearity of a PA. However, in the frequency domain, it is more convenient to use the IMD3. It is defined as the ratio of the first harmonic signal to the third-order intermodulation product. The relationship between IMD3 and IIP_3 is

$$IIP_3 = P_{in} + \frac{IMD3}{2} \tag{5.1}$$

Figure 5.4 shows the IMD3 corresponding to Figure 5.3. The instant efficiency is displayed on the same diagram. Because efficiency is proportional to input power, there are three jumps in the efficiency curve due to current switching, shown in red. The blue line shows the case without switching. The pink and dark blue lines compare the IMD3;

as expected, distortion is worse at lower power consumption. Overall however, the IMD3 stays above 40 dBc, an improvement over the situation when the maximum input is greater than -1 dBm.

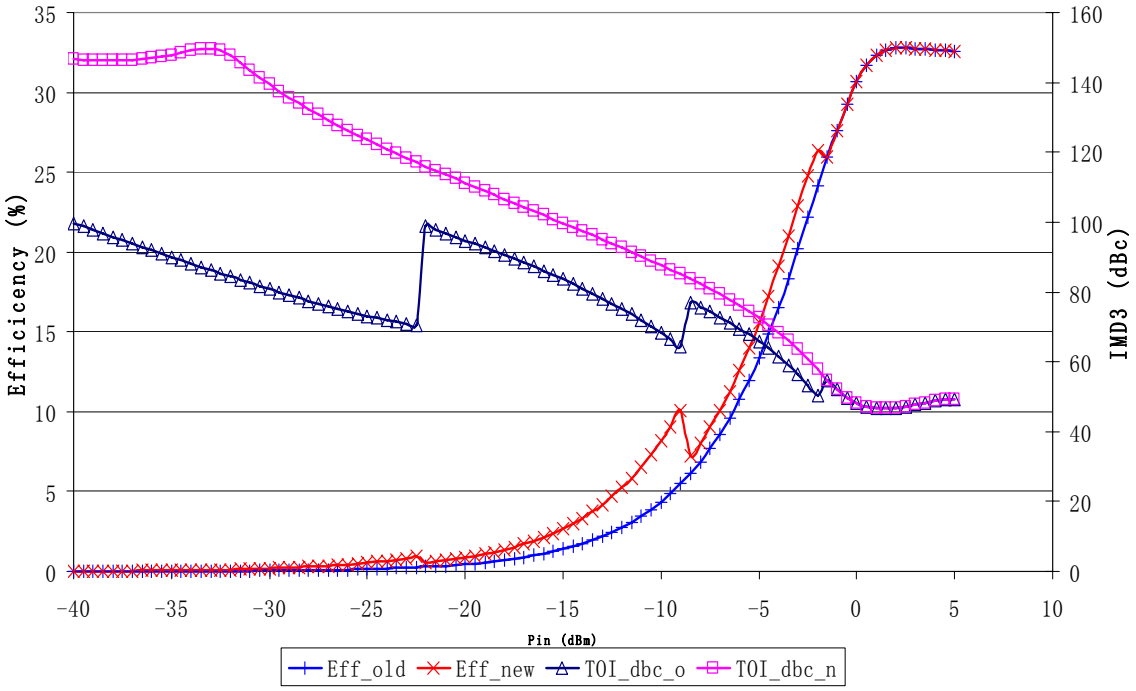


Figure 5.4 IMD3 and Efficiency Versus Input

Figure 5.5 compares the gain with and without gain compensation. Clearly, the gain is much flatter with compensation, especially at low input levels where the gain is improved by almost four decibels.

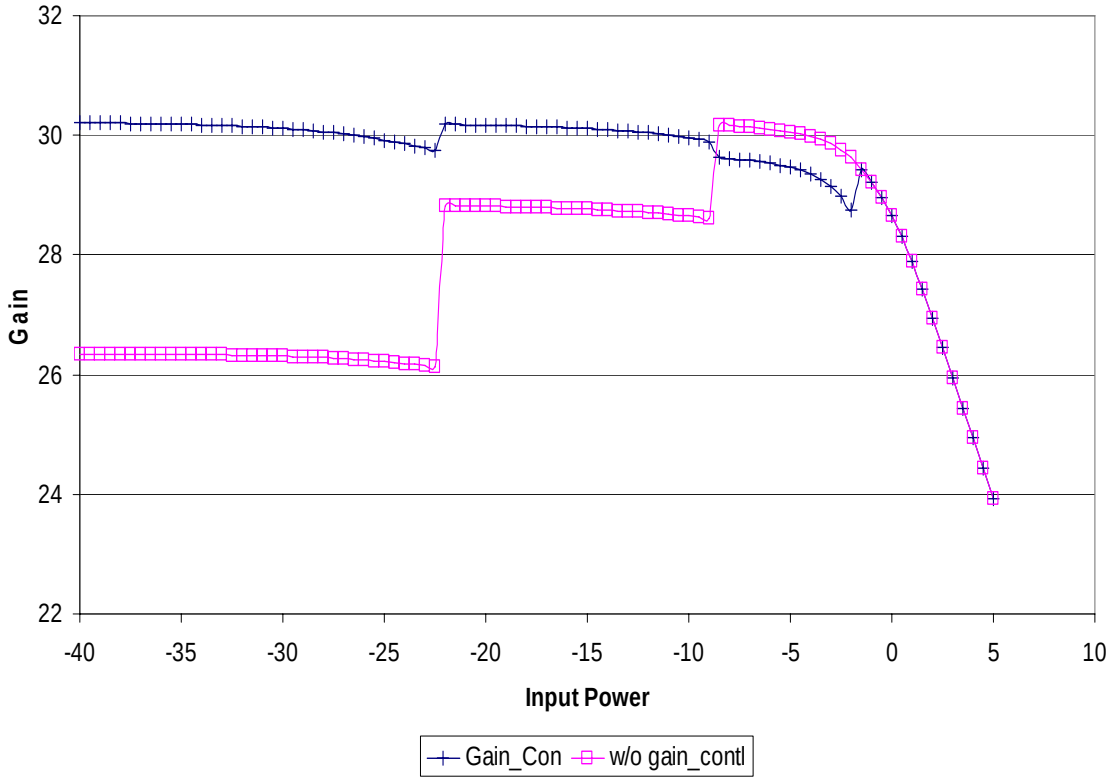


Figure 5.5 Gain Comparison

5.2 Transient Simulation

The analysis in section 5.1 is purely for the steady state; there is no transition time from one biasing point to the other and finite transition effects are not considered. This is the ideal case. The reality is more complex. Transitions from one state to another will cause ripples and spikes in the RF path and degrade system linearity. Thus, time-domain simulation is necessary. According to our control strategy, the characteristics of the source signals directly affect how much power can be saved. The dynamically biased PAs are best suited to applications with large peak-to-average ratios. The larger the ratio, the more power saved by the system. The reason is as follows: To accommodate a large signal, the PA needs a high bias current. However, traditional Class A PAs' biasing

currents do not change, so efficiency is large only at the peak signal. Because the average is small compared to the peak, the average efficiency is also small.

To simplify the comparison, the following simulation uses an AM as signal source; the peak-to-average ratio is two if only two-tone sinusoidal signals are chosen. Both have the same amplitude and are offset by 2 MHz. This bandwidth is usually larger than most applications, including CDMA. If the target application needs more bandwidth (WLAN's bandwidth is 20 MHz), the same procedure is followed and the results will be shown later.

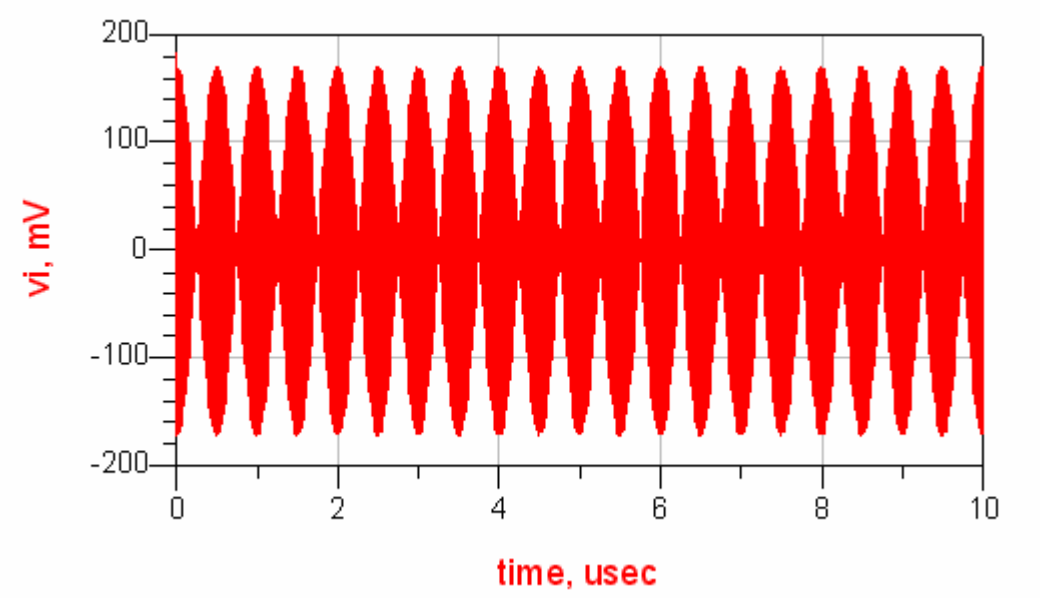


Figure 5.6 The Input Signal Waveform

The simulation is set up as follows. The input is the two-tone sinusoidal waveform,

f_1 = 2.401 GHz, P_1 = -14 dBm f_2 = 2.399 GHz, P_2 = -14 dBm So that the peak power, P_{tot} = -8 dBm	(All refer to 50 Ω)	(5.2)

The peak power is roughly 6 dB lower than the $P_{1\text{-}dB}$, which is usually gives enough linearity in an OFDM system [1]. The current-switching point is selected at P_{in} = -22 dBm and P_{in} = -8 dBm, which divide the PA biasing into quarter-, half- and full-current

regions. In this simulation, the three-quarters–current situation is not considered because it is too close to the 1-dB compression point. Three simulation scenarios are given: The PA working at full current; the PA working with current switching and gain control; and the PA working with current switching and without gain control.

A. PA Working at Full Current

In this biasing condition, we expect the least amount of distortion. Figure 5.7 shows the output waveform at full current.

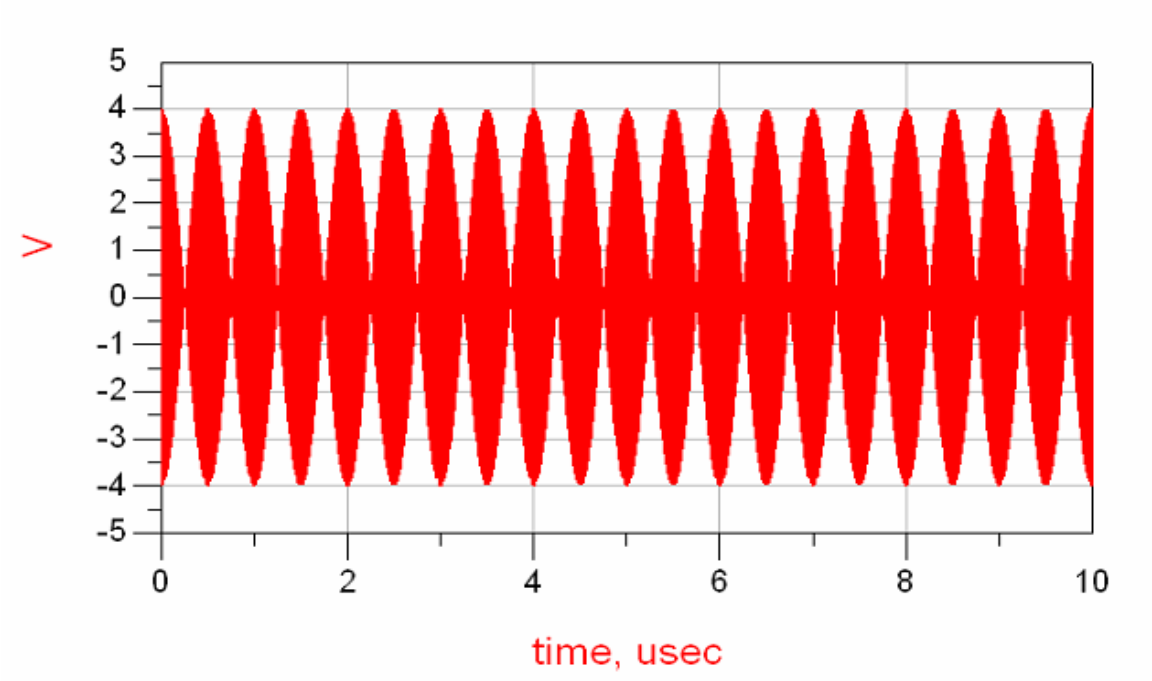


Figure 5.7 Output Waveform at Full Current

The envelope of the output is very smooth and obviously has little distortion. To see it more clearly, the time-domain waveform is transferred into the frequency domain with a Fourier transform; see Figure 5.8. *M8* is the first harmonic component and is the desired output. All other signals are due to distortion. Figure 5.9 presents a close-up of the area near the input frequency. The IMD3 is given by *M9-M10*, which is equal to 51 dBc. This

IMD3 specifies the system's linearity. Figure 5.10 zooms further in on the operation range to show in-band interference.

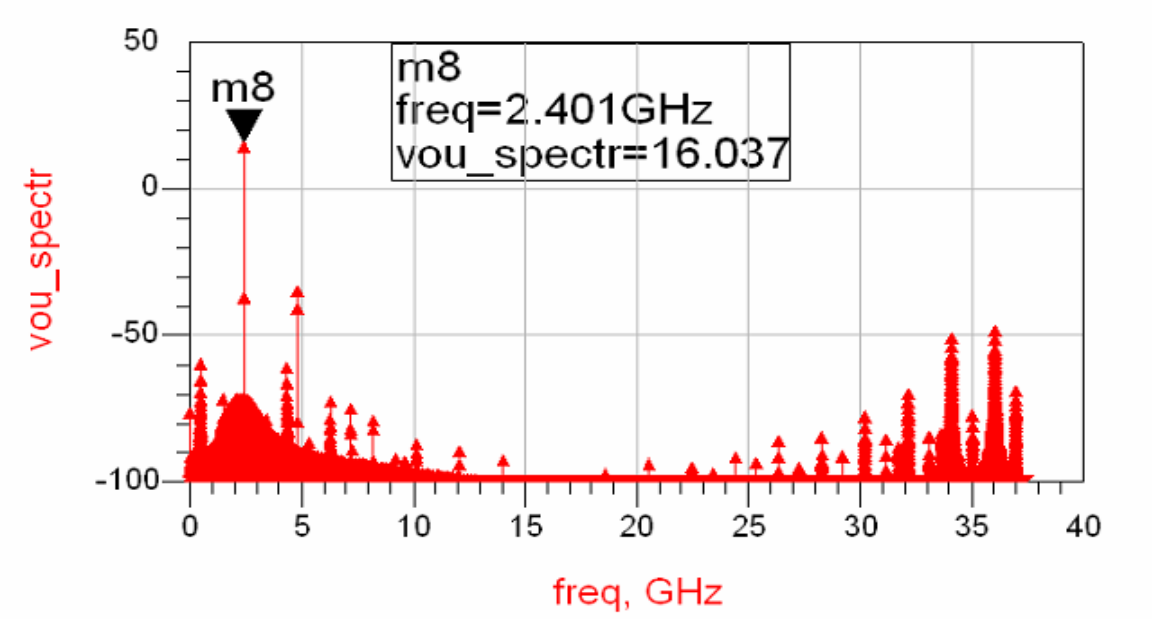


Figure 5.8 Output Spectrum at Full Current

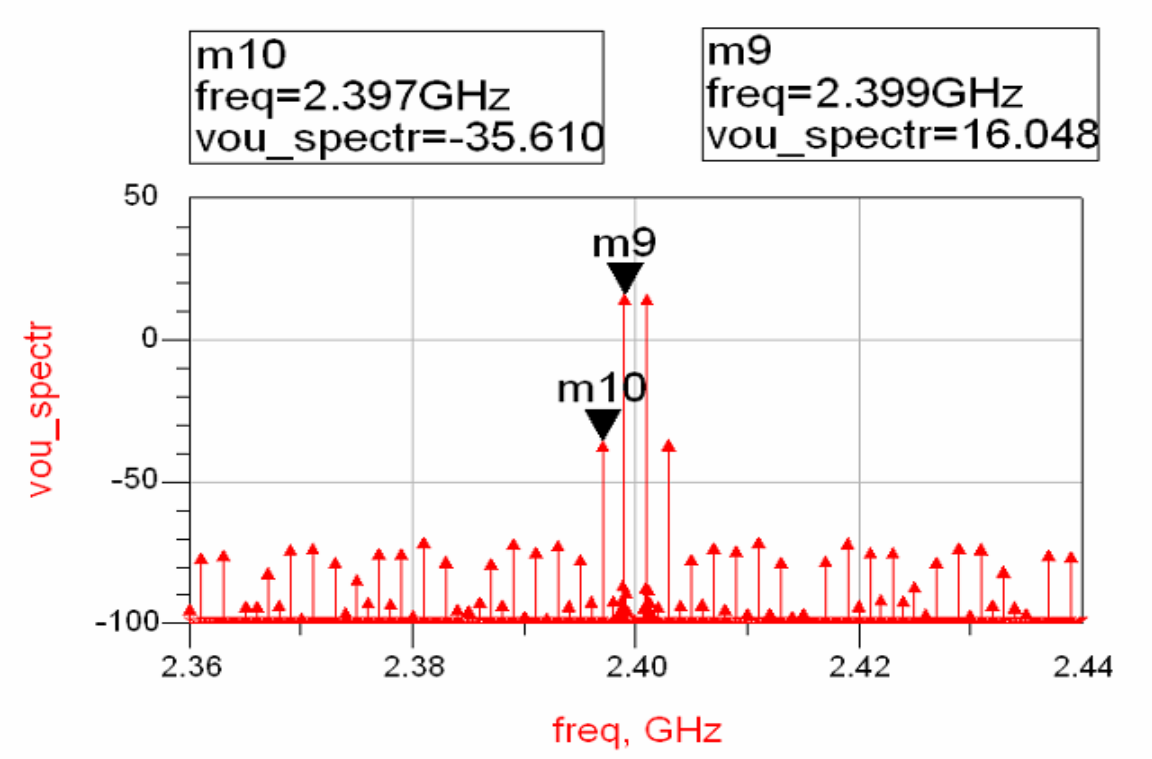


Figure 5.9 Spectrum Near the Operation Frequency

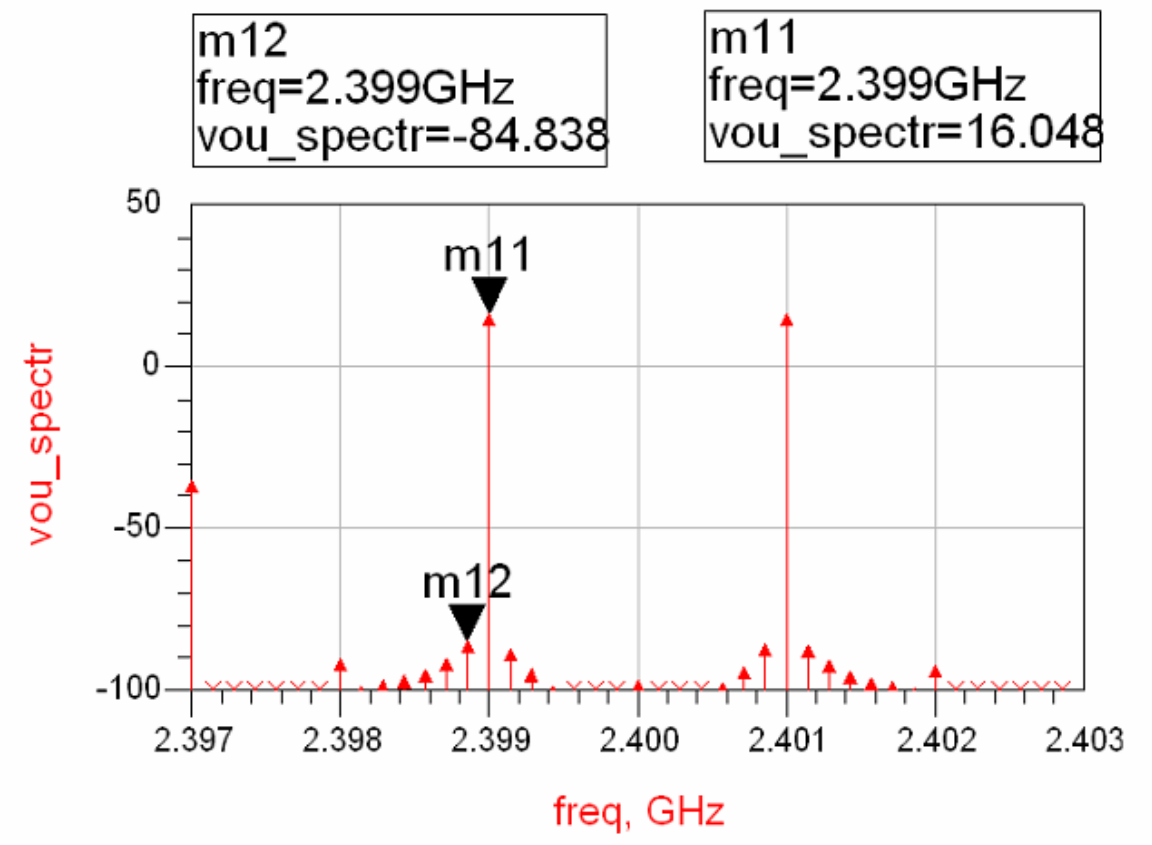


Figure 5.10 Effect of In-Band Interference

B. PA Working with Dynamic Current Switching and Gain Compensation

Figure 5.11 gives the output waveform if the biasing current changes dynamically and gain is compensated. Comparing to Figure 5.7, there are spikes when the biasing point varies, although the curve is still relative smooth. Figure 5.12 shows the frequency spectrum of Figure 5.11. Zooming into the operation frequency, Figure 5.13 shows the first- and third-harmonic components are 15.917 dBm and -19.235 dBm respectively. So the IMD3 is 35.152 dBc. This number is less than we get from part A above, confirming that this system is less linear than the previous one. Figure 5.14 gives the less important in-band interference estimate.

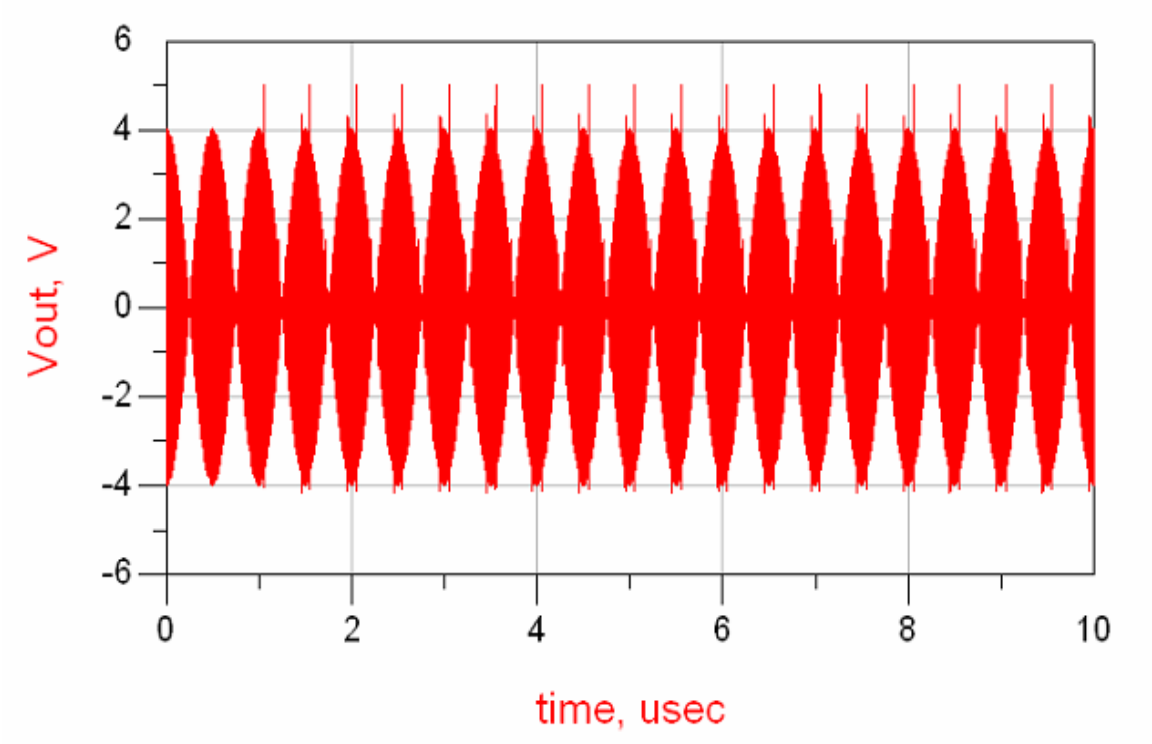


Figure 5.11 Output Waveform with Current Switching and Gain Control

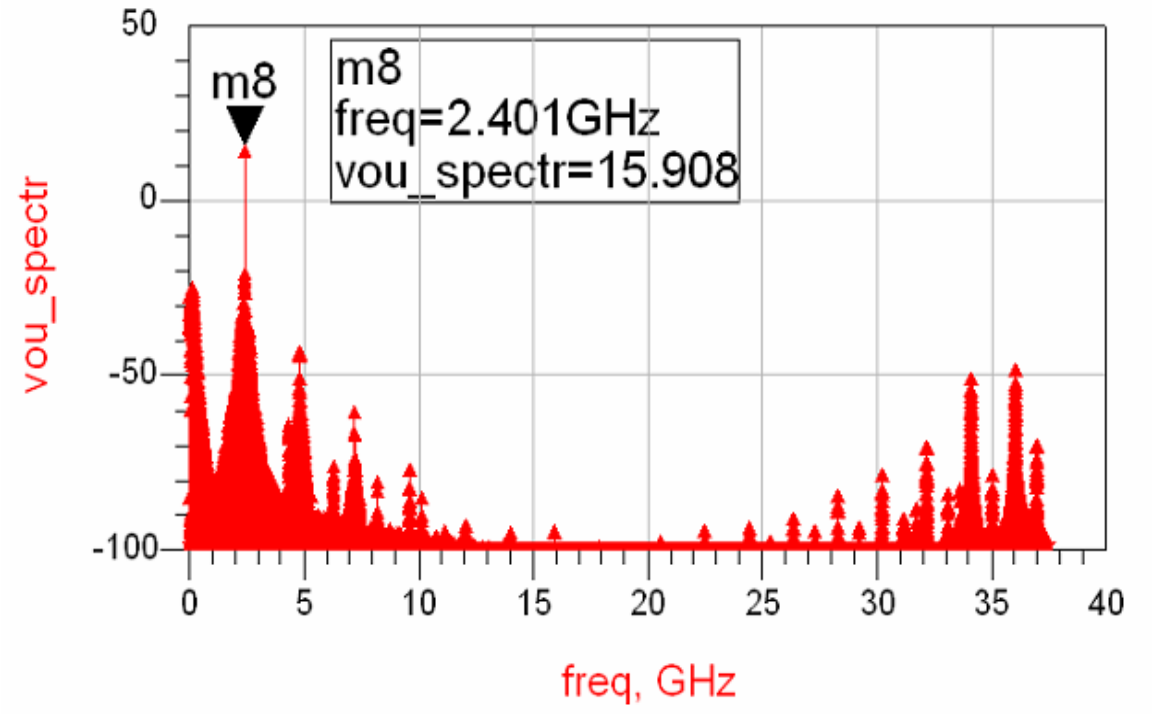


Figure 5.12 Spectrum of Output with Current Switching and Gain Control

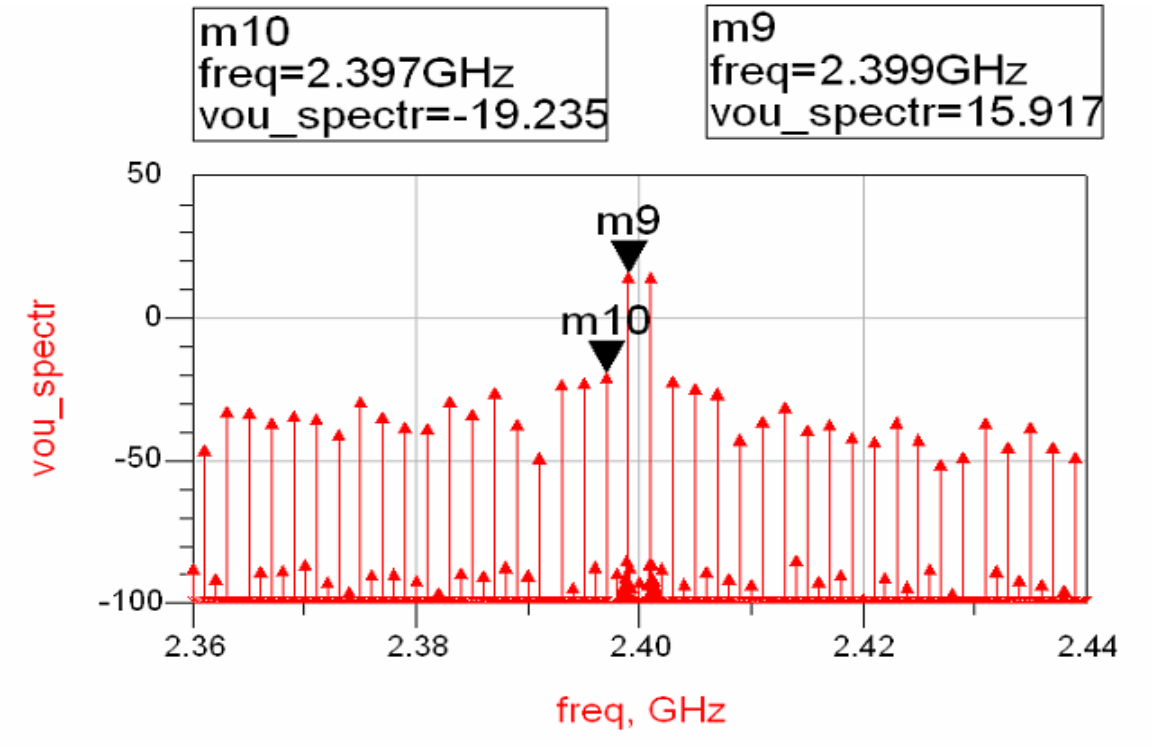


Figure 5.13 Spectrum Near the Operation Frequency

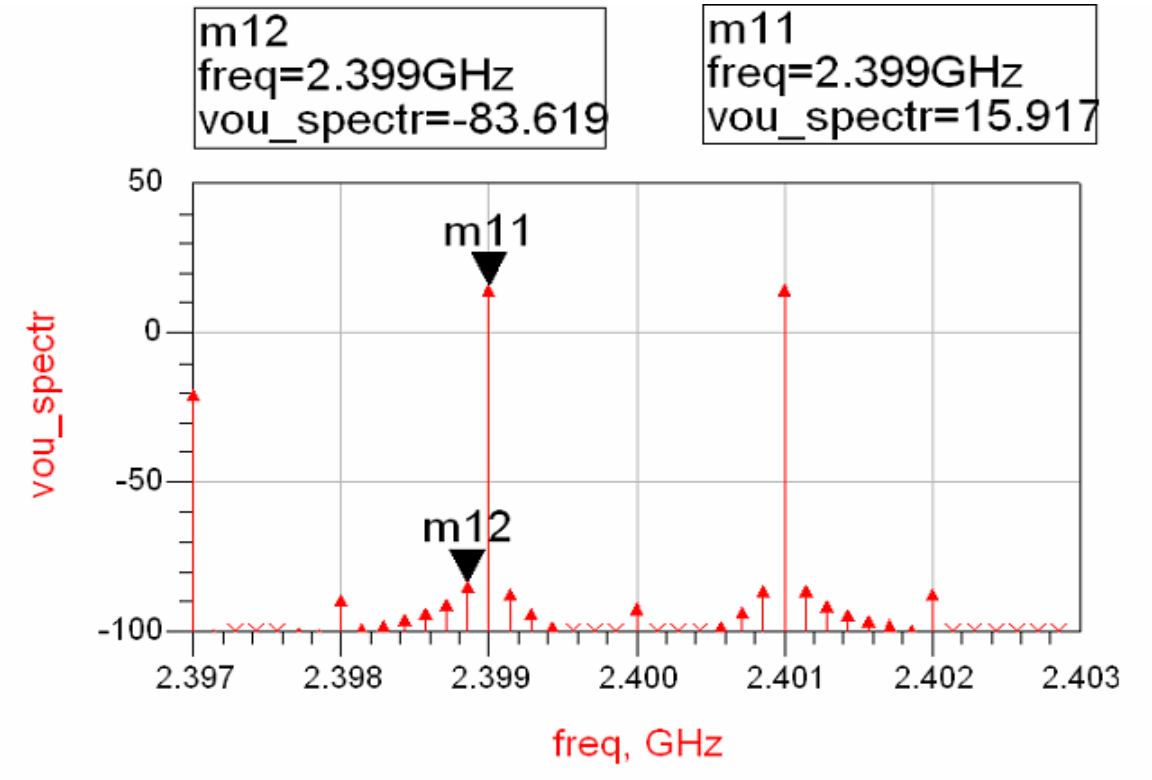


Figure 5.14 Effect of In-Band Interference

C. PA Working with Dynamic Current Switching and Without Gain Compensation

In general, it is usually desirable to not introduce a gain-control scenario. Such function requires extra circuits to control the overall gain, which usually will include some feedback. These circuits must be carefully designed to avoid oscillation. Feedback will also change the load impedance, causing additional distortion.

On the other hand, gain control may reduce the overall nonlinearity due to large gain variations, so we need to know the characteristics of the PA with only the bias-current switching. Figure 5.15 gives the output waveform with current switching and without gain control. As we expected, when input level is large, the PA works at full current, causing the gain to be even higher. The envelope no longer looks sinusoidal and there are large distortions in additional to the spikes we see in Case B.

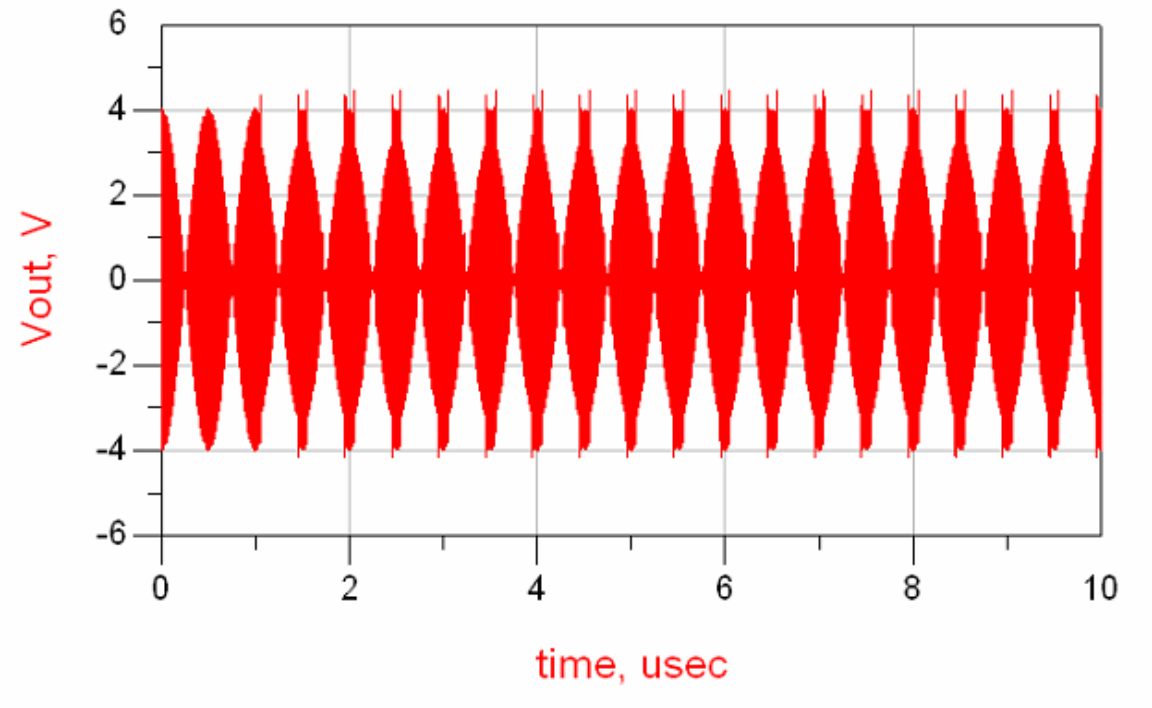


Figure 5.15 Output Waveform with Current Switching and Without Gain Control

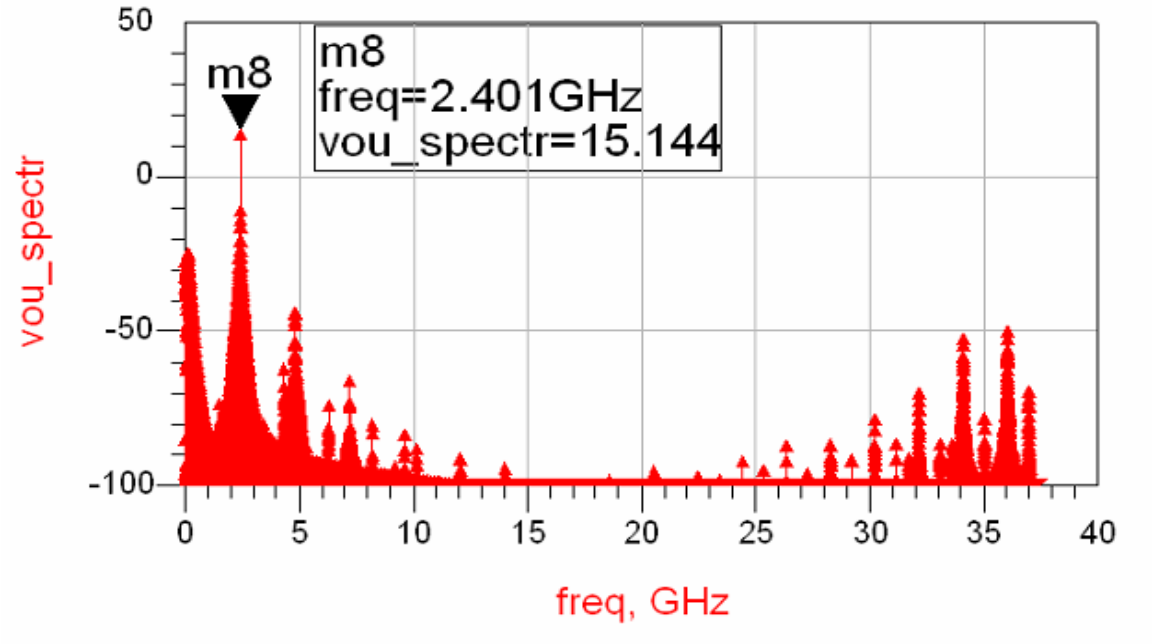


Figure 5.16 Spectrum of Output with Current Switching and Without Gain Control

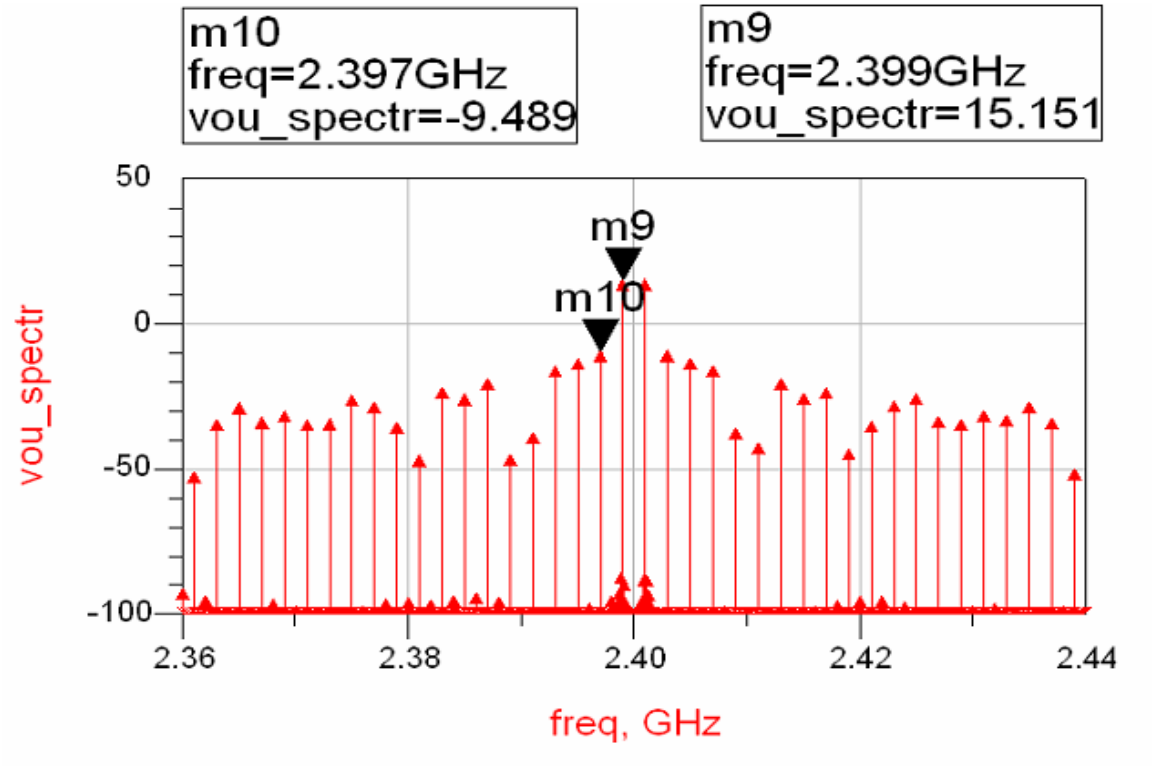


Figure 5.17 Spectrum Near the Operation Frequency

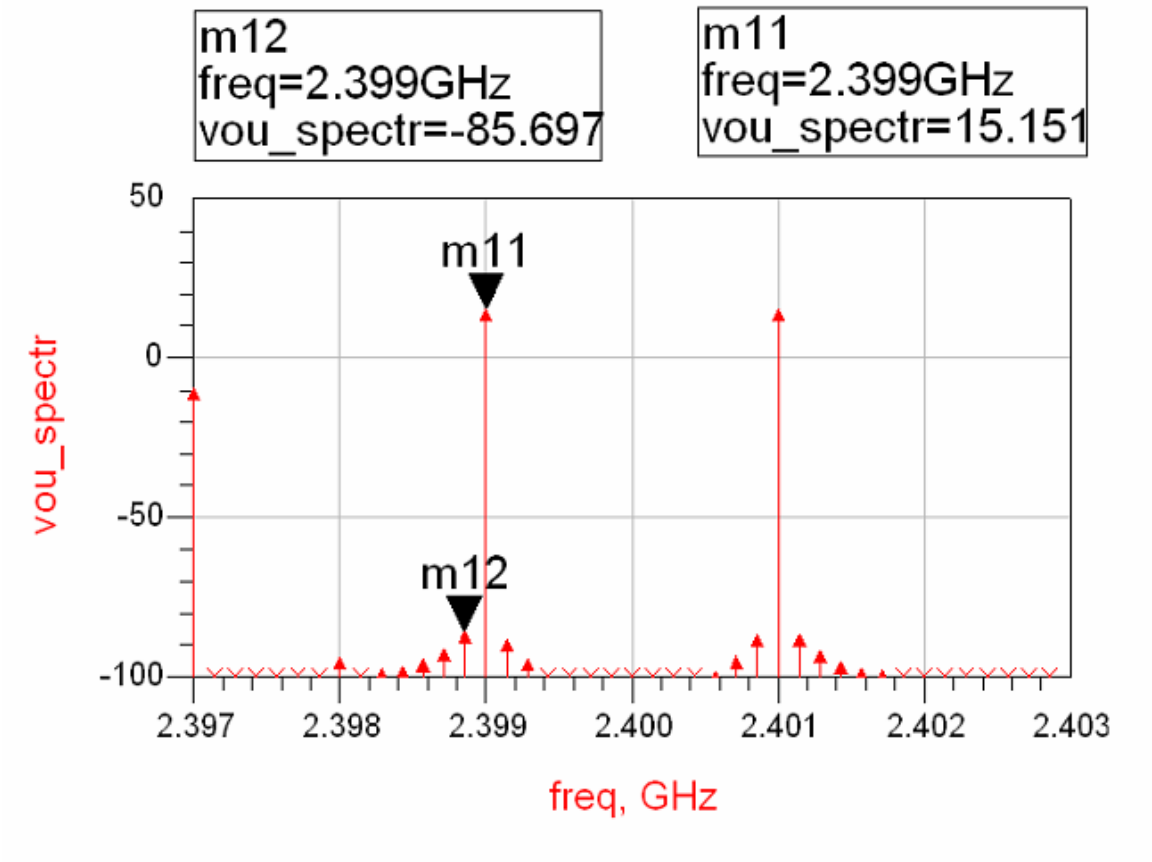


Figure 5.18 Effect of In-Band Interference

Following the same procedure, we can get the frequency domain spectrum, shown in Figures 5.16 and 5.17. The IMD3 in this scenario is 24.64 dBc, which is the worst of the three situations. It is better to introduce the gain compensation to improve the linearity.

Table 5.1 Summary of Linearity and Efficiency at 2 MHz Bandwidth

	Gain (dB)	IMD3 (dBc)	Average PAE (%)
w/o g_cont w/o switching	30.0	51.6	6.7
w/ g_cont w/ switching	29.1	35.1	11.5
w/o g_cont w/ switching	29.1	24.6	11.5

The linearity and efficiency of the three situations are summarized in Table 5.1. The first row is the scenario with the PA working at full current with neither switching nor gain control (w/o g_cont, w/o switching). The second row is the scenario with the PA working with switching and gain control (w/ g_cont, w/ switching). The last row is the scenario with the PA working with switching and without gain control (w/ g_cont, w/o switching). Employing gain control and switching improves the efficiency by 71.6% in this particular input scenario. PA works in full-current, half-current and quarter-current modes 21%, 67% and 12% of the time, respectively. Although the linearity is degraded, it is acceptable because the biasing current is smaller. And using gain control improves the IMD3 by more than 10 dB compared to the no-gain-control scenario.

D. 20-MHz Bandwidth Simulation

WLAN may employ OFDM modulation with a 20-MHz bandwidth. The following gives the transient simulation for a 20-MHz offset.

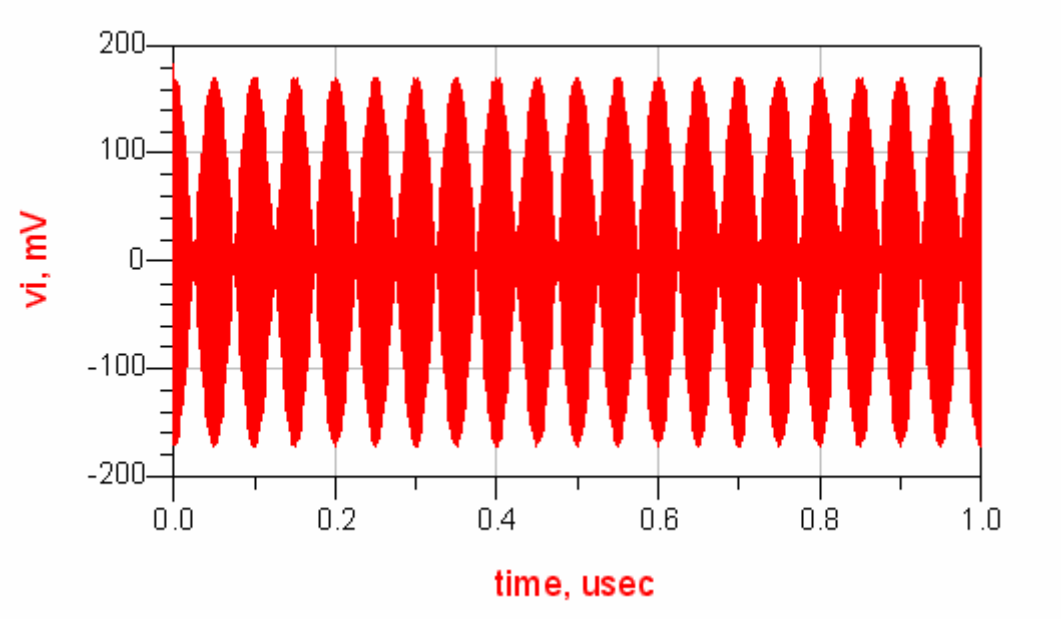


Figure 5.19 Input Waveform for a 20-MHz Offset

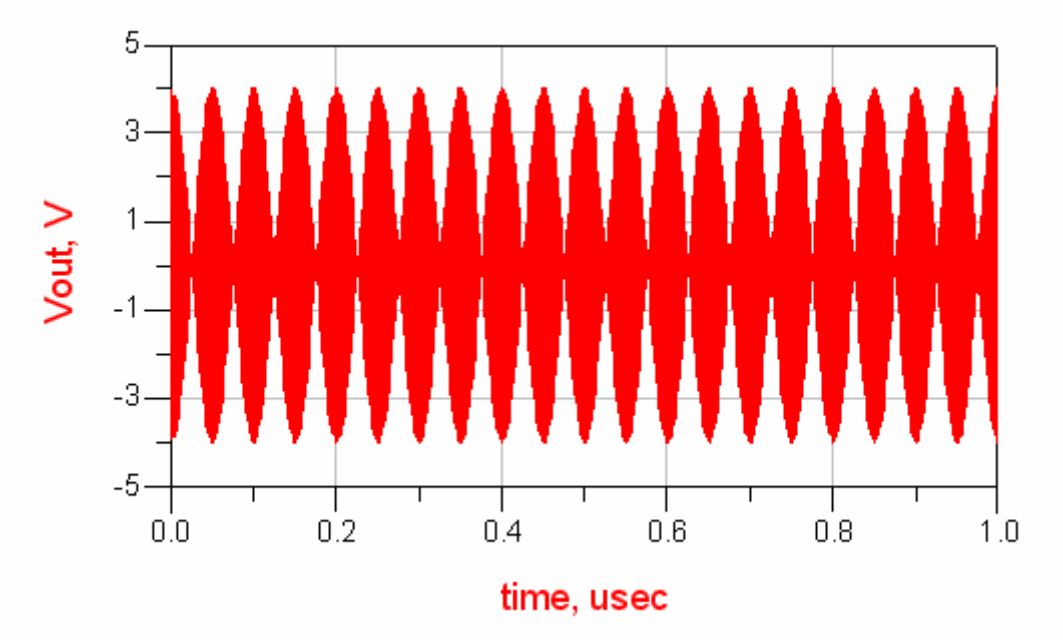


Figure 5.20 Output Waveform for a 20-MHz Offset at Full Current

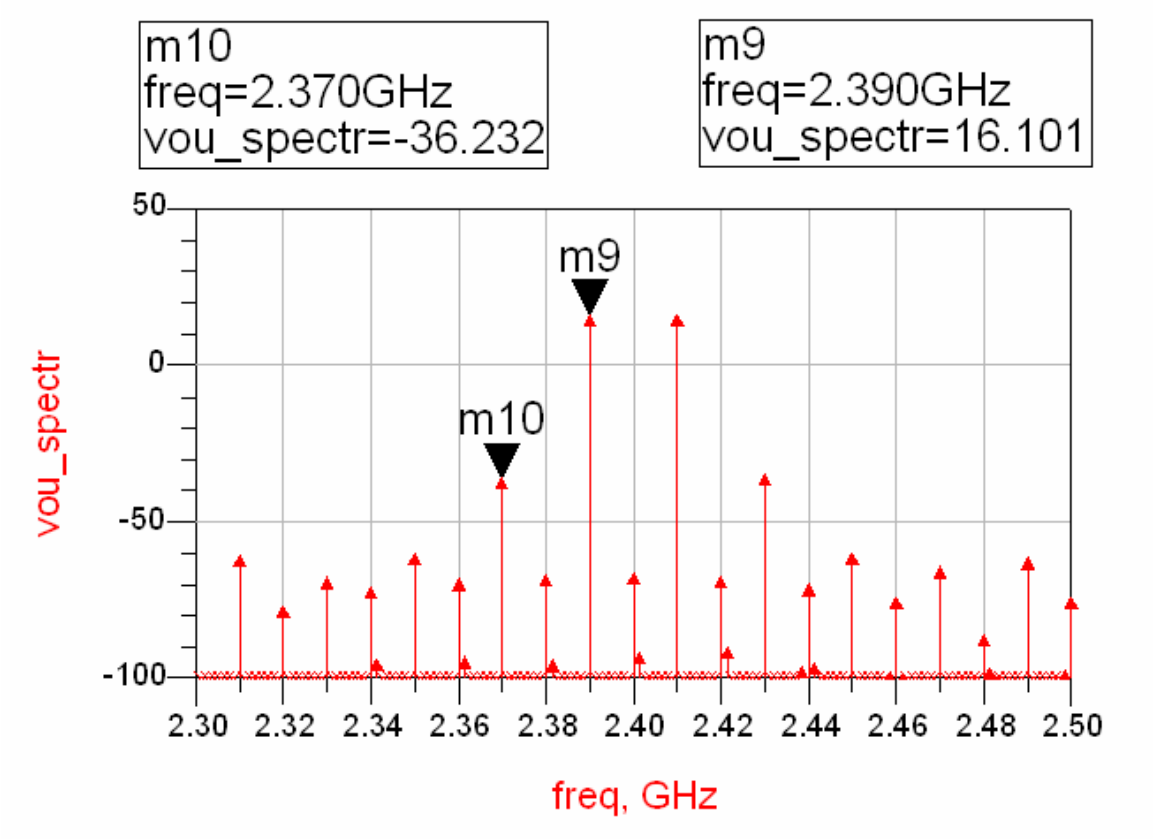


Figure 5.21 Spectrum Near the Operation Frequency

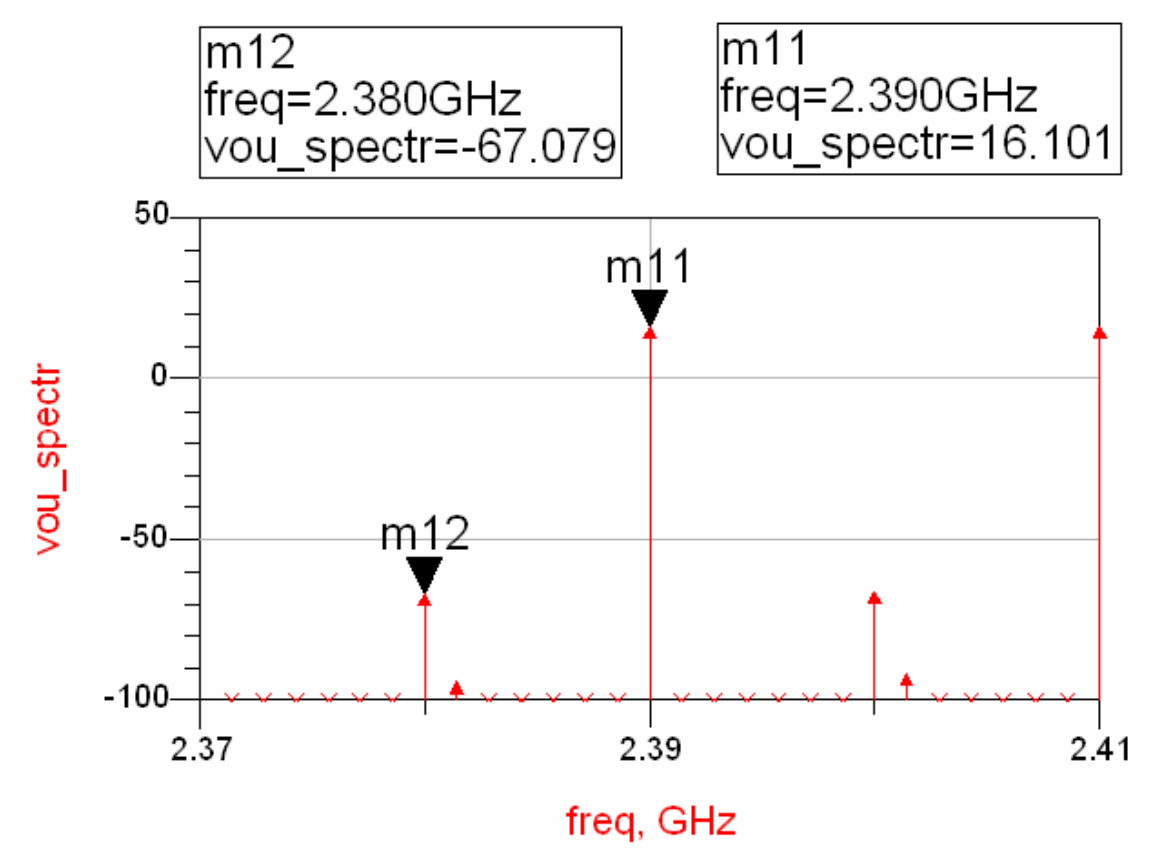


Figure 5.22 Effect of In-Band Interference

Figures 5.20–5.22 give the output waveform, the corresponding spectrum and the effect of in-band interference as in 2-MHz bandwidth situation. The envelope is smooth and the linearity is high. The IMD3 is 52.3 dBc, comparable to the 2-MHz case. The in-band effect is very limited.

Figures 5.23–5.25 show the results of a PA with gain control and current switching. The time domain output waveform shows many spikes; the linearity is not as good in Figure 5.20, as confirmed by the IMD3 of about 36.4 dBc. Figures 5.26–5.28 present the situation with the PA working without gain control but with current switching. Linearity is even worse; the IMD3 is only about 27 dBc. The efficiency improvement is roughly the same as in the 2-MHz bandwidth situation. Table 5.2 summarizes the results.

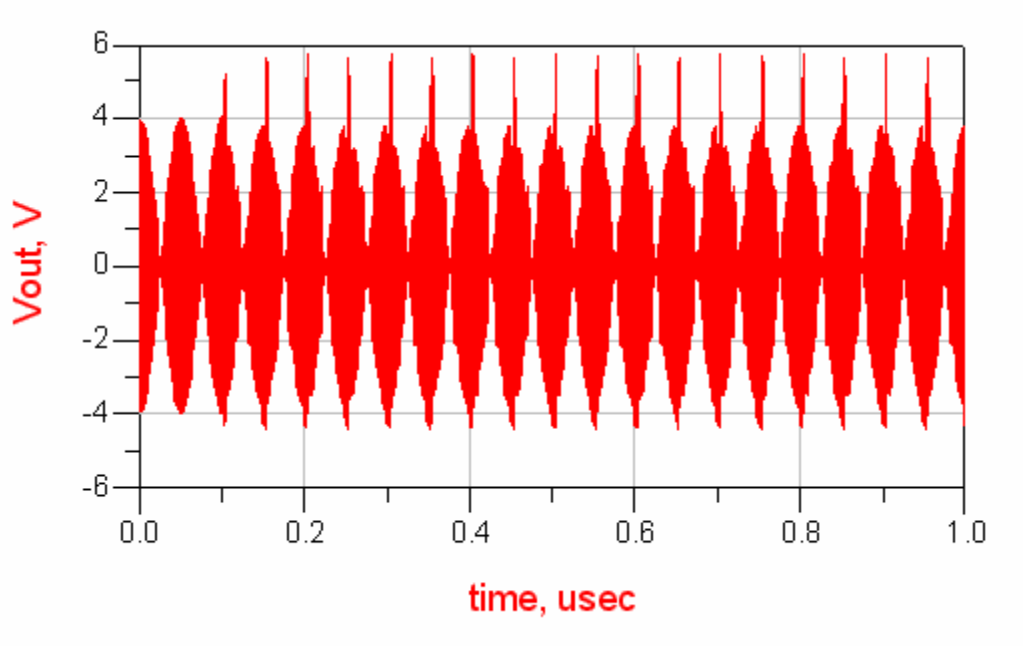


Figure 5.23 Output Waveform with Current Switching and Gain Control with 20 MHz of Bandwidth

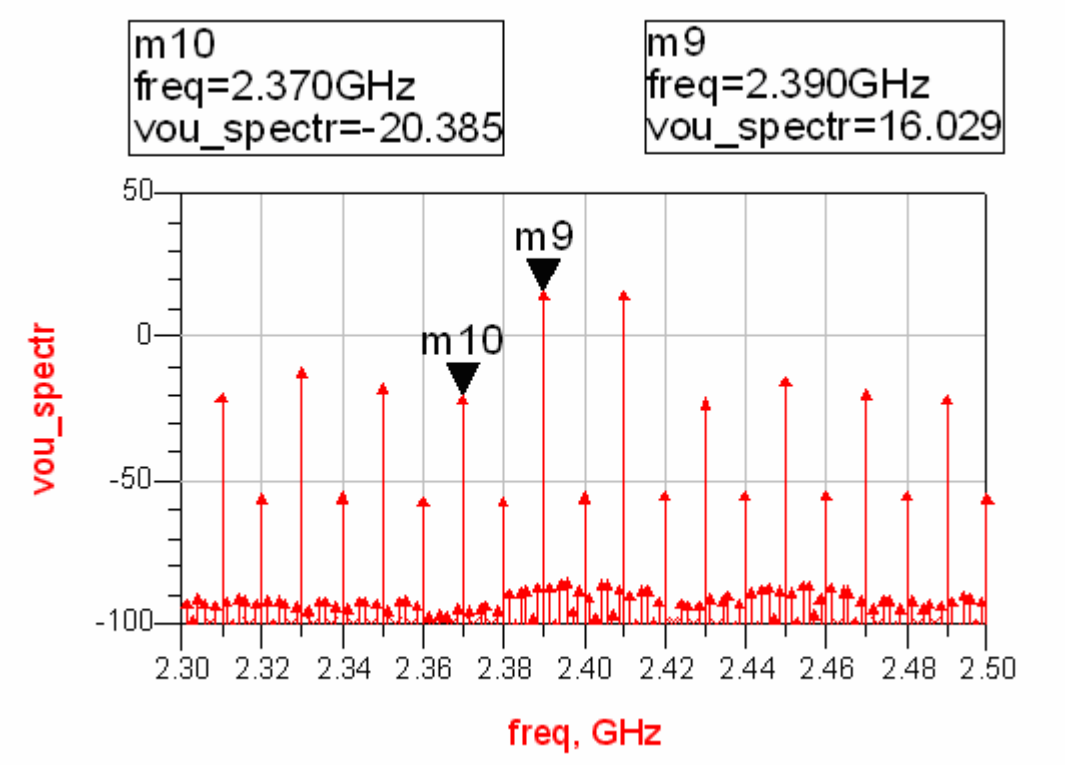


Figure 5.24 Spectrum Near the Operation Frequency

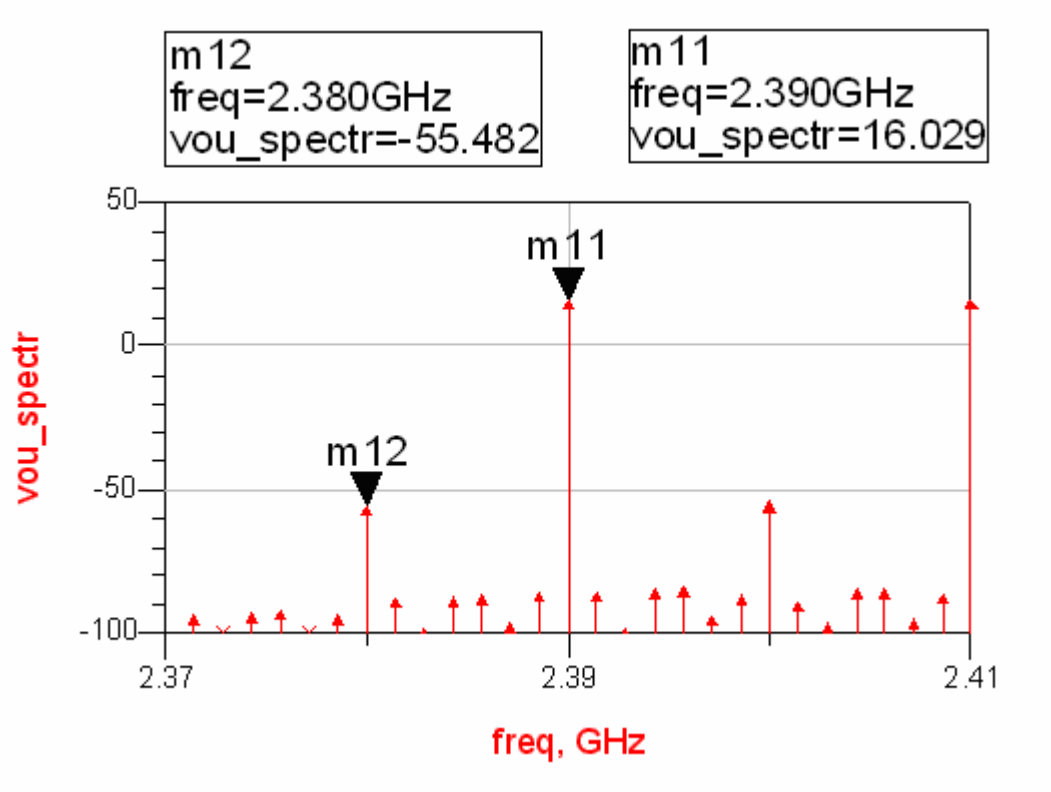


Figure 5.25 Effect of In-Band Interference

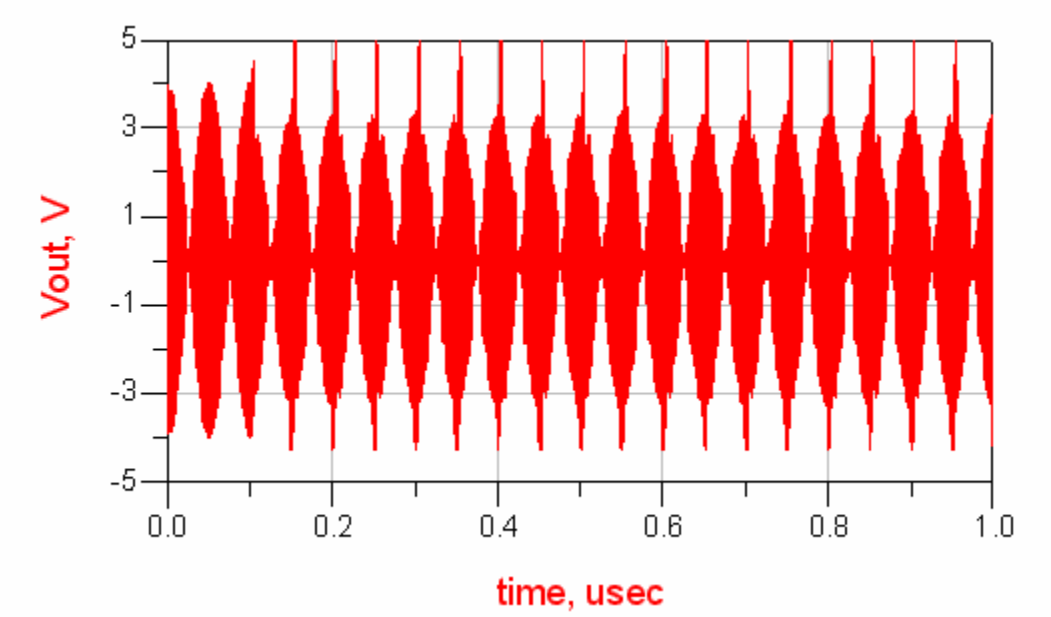


Figure 5.26 Output Waveform with Current Switching and Without Gain Control with 20 MHz of Bandwidth

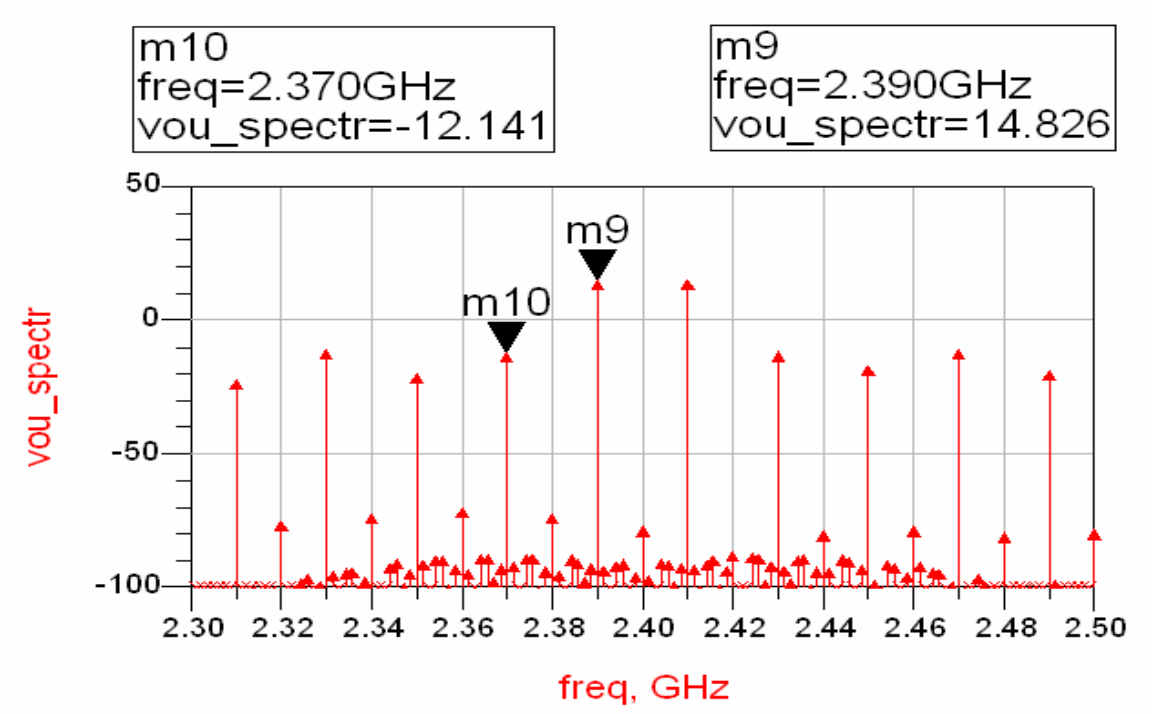


Figure 5.27 Spectrum Near the Operation Frequency

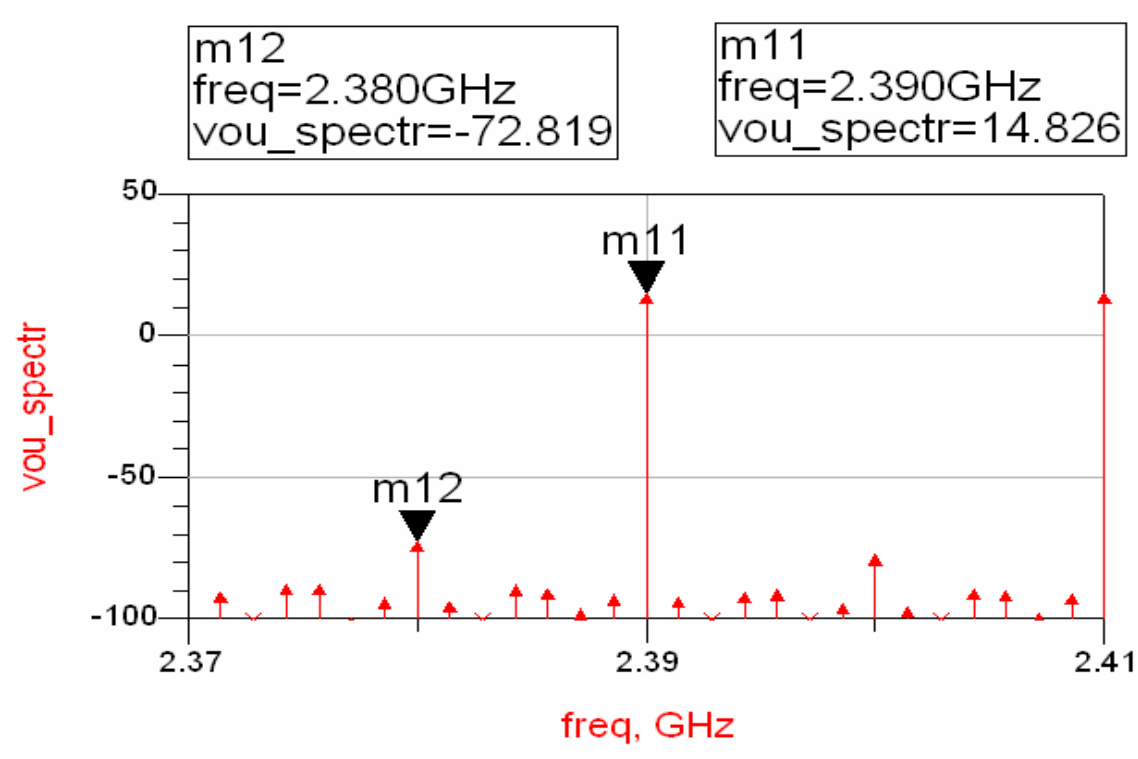


Figure 5.28 Effect of In-Band Interference

Table 5.2 Summary of Linearity and Efficiency at 20 MHz Bandwidth

	Scenario	Gain (dB)	IMD3 (dBc)	Average PAE (%)
1	w/o g_cont w/o switching	30.0	52.3	6.7
2	w/ g_cont w/ switching	29.1	36.4	11.5
3	w/o g_cont w/ switching	29.1	27.0	11.5

5.3 Power Detector Simulation

The power-detector circuit diagram is shown in chapter 4. Here we present the time-domain simulation results. The input is still the two-tone sinusoidal waveform.

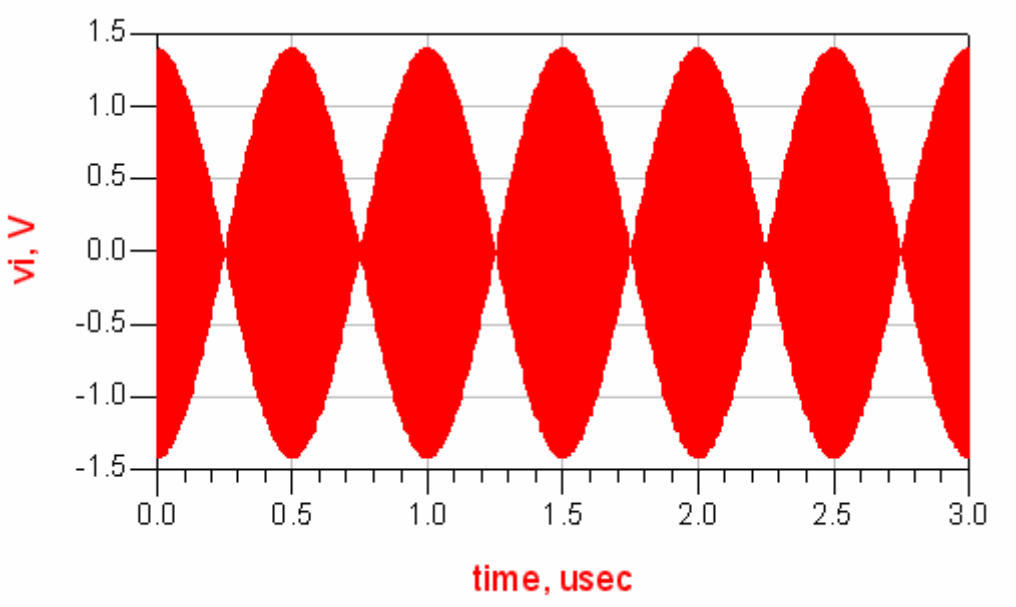


Figure 5.29 Power Detector Input with 2 MHz of Bandwidth

The frequency is 2.401 GHz and 2.399 GHz, with an offset of 2 MHz. Both have 1.5 dBm input power, which is comparable to the output voltage in Figure 5.7. The output waveform is given in Figure 5.30. The extracted output can be calibrated to the

corresponding output power and fed into the control circuit to accomplish the current switching and gain compensation.

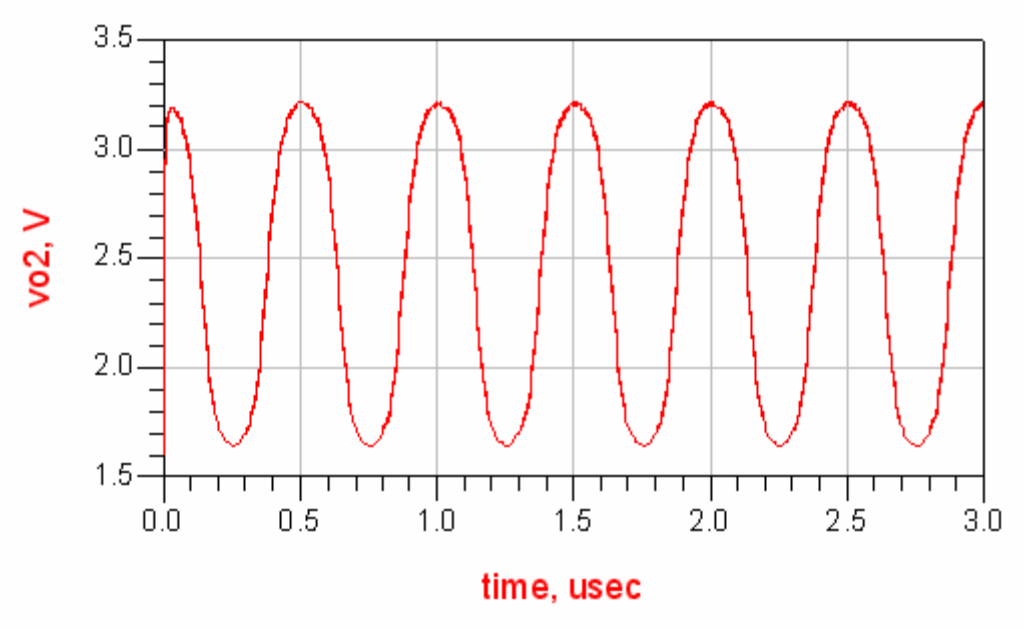


Figure 5.30 Power Detector Output with 2 MHz of Bandwidth

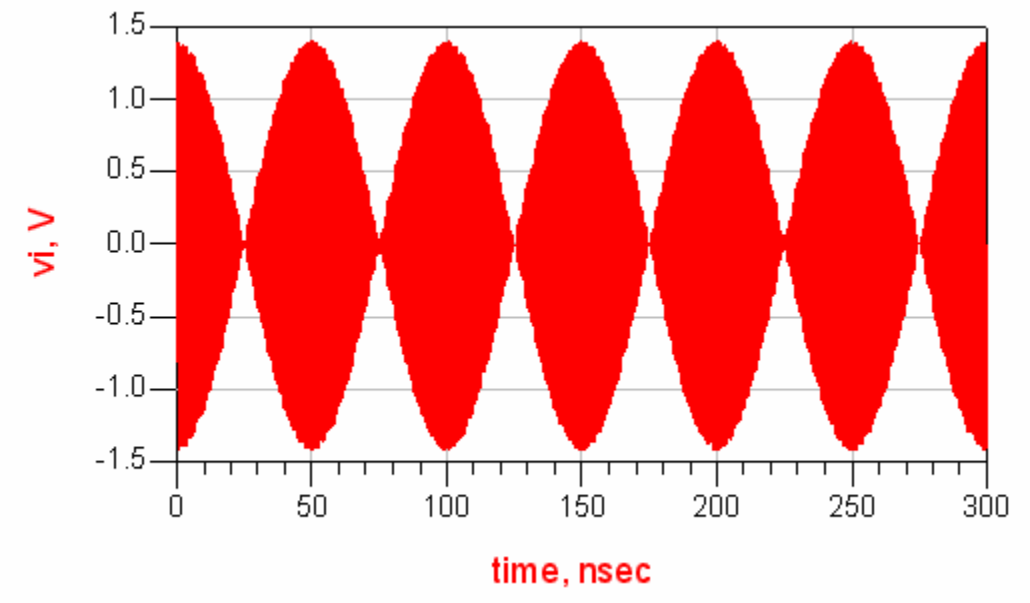


Figure 5.31 Power Detector Input with 20 MHz of Bandwidth

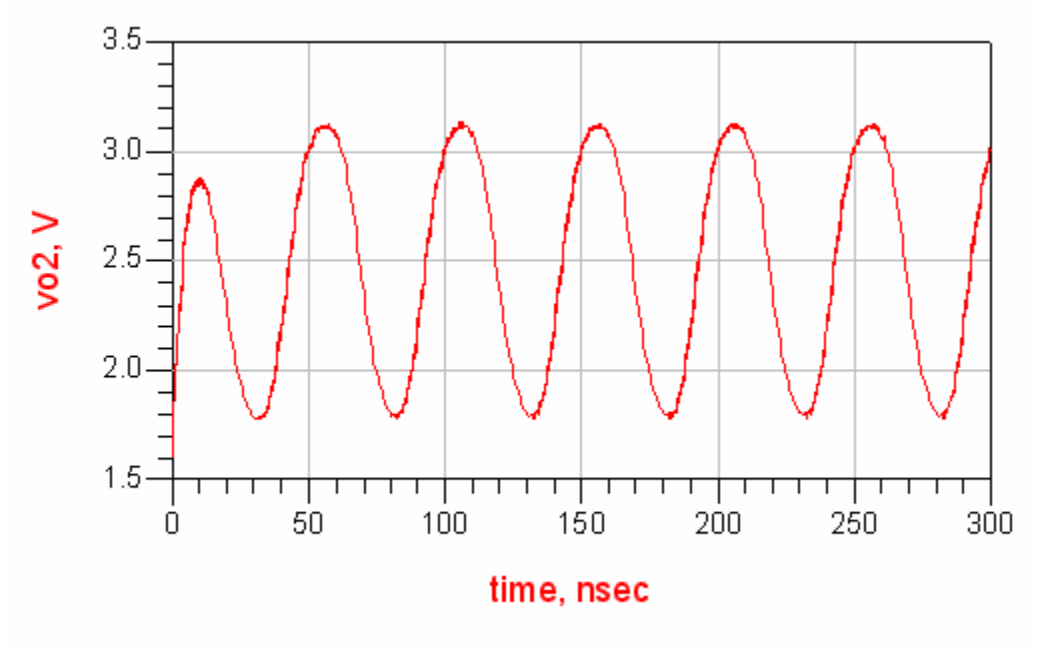


Figure 5.32 Power Detector Output with 20 MHz of Bandwidth

For the 20-MHz case, the simulation results are given in Figures 5.31 and 5.32. The input power level is again 1.5 dBm.

5.4 Simulation PA with Power Detector

Putting the PA, power detector and driver circuits together, we can design a system that can automatically adjust the biasing current to improve the efficiency. The gain control improves linearity without sacrificing much power. The block diagram of such a system is shown in Figure 5.33. The circuits inside the blocks were discussed in chapter 4. In the driver block, two identical driver circuits control the biasing currents at full, half or quarter. In fact, it is easy to add the third such circuit to bias the PA at three quarters current. In the simulation, we still choose the same input as 5.2: f_1 = 2.401 GHz, P_1 = -14 dBm, f_2 = 2.399 GHz and P_2 = -14 dBm. The offset frequency is 2 MHz, which is the bandwidth of the baseband.

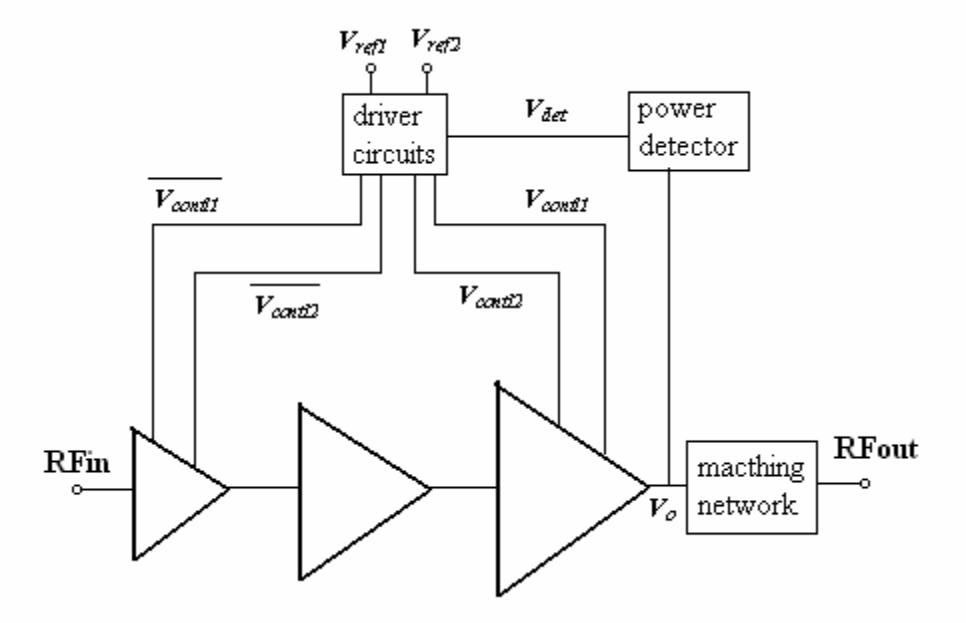


Figure 5.33 Block Diagram of the PA with Power Detector and Driver

In domain simulation, we can get the waveform at each node in Figures 5.34–5.41. Figure 5.34 is the input waveform, the waveform at RFin. Figure 5.36 presents the waveform at RFout; linearity will be analyzed later. V_{ref1} and V_{ref2} are 1.79 and 3 V respectively.

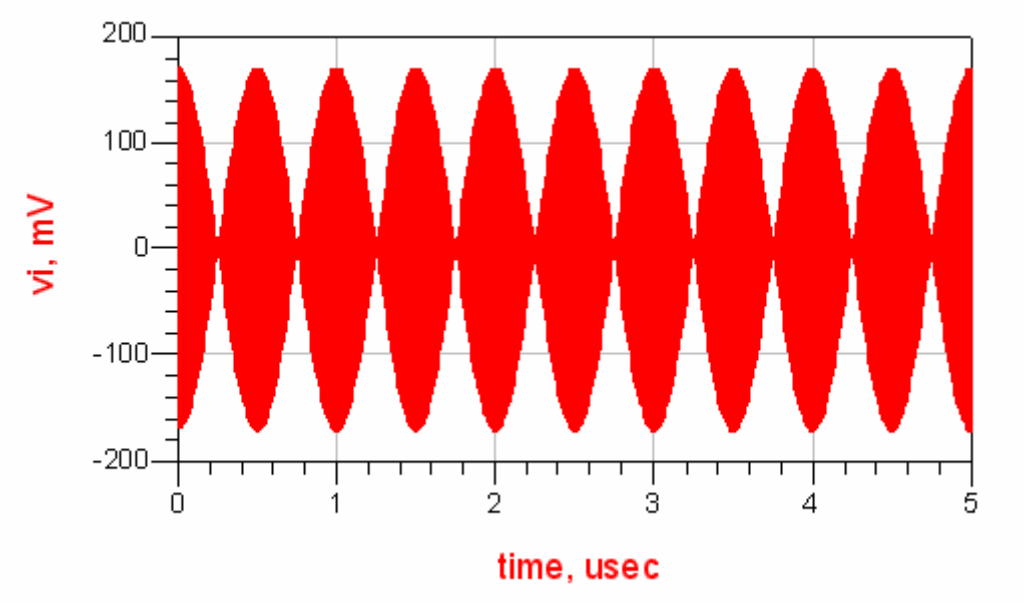


Figure 5.34 Waveform at RFin

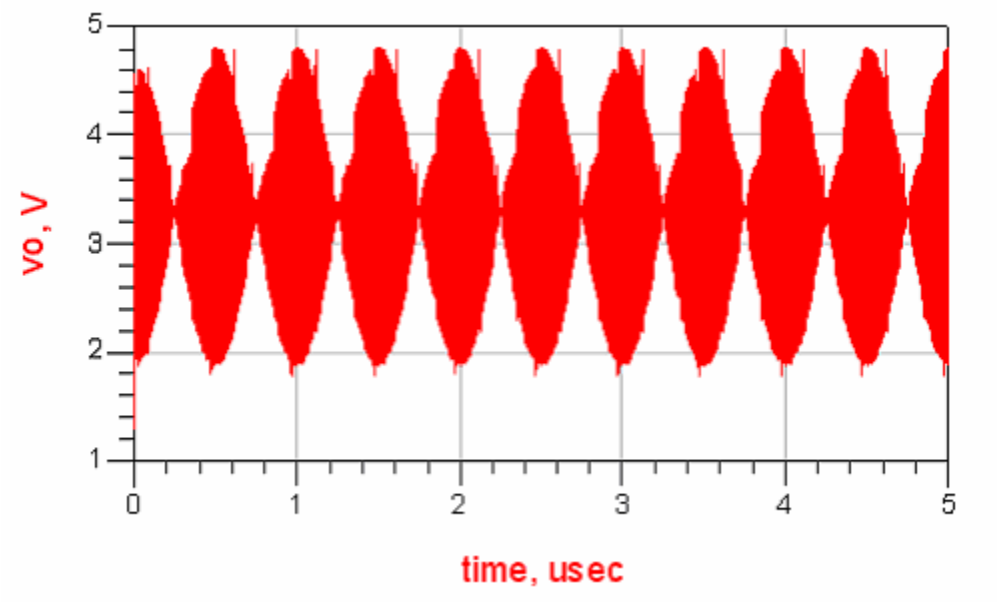


Figure 5.35 Waveform at V_o

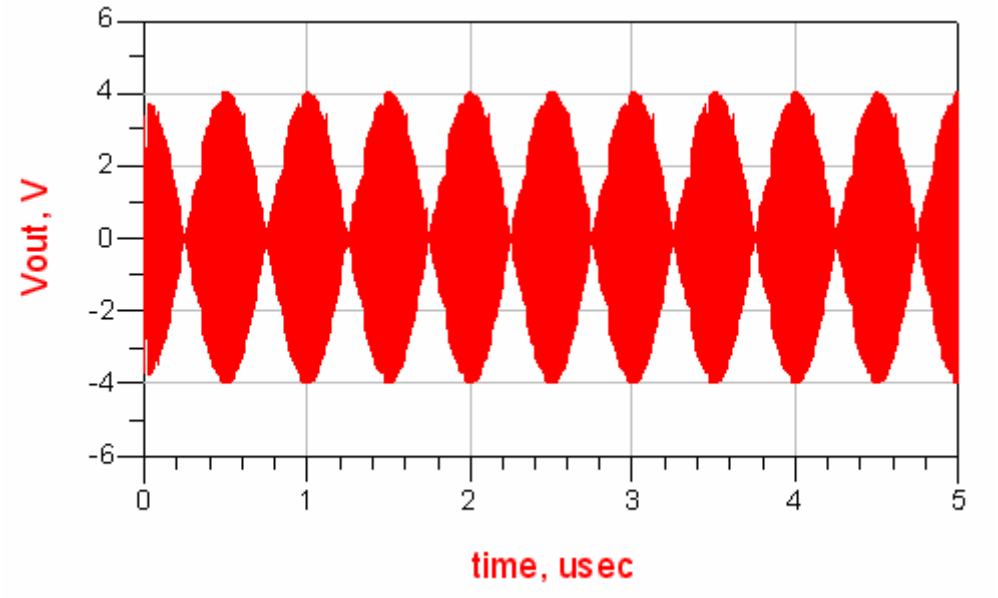


Figure 5.36 Waveform at RFout

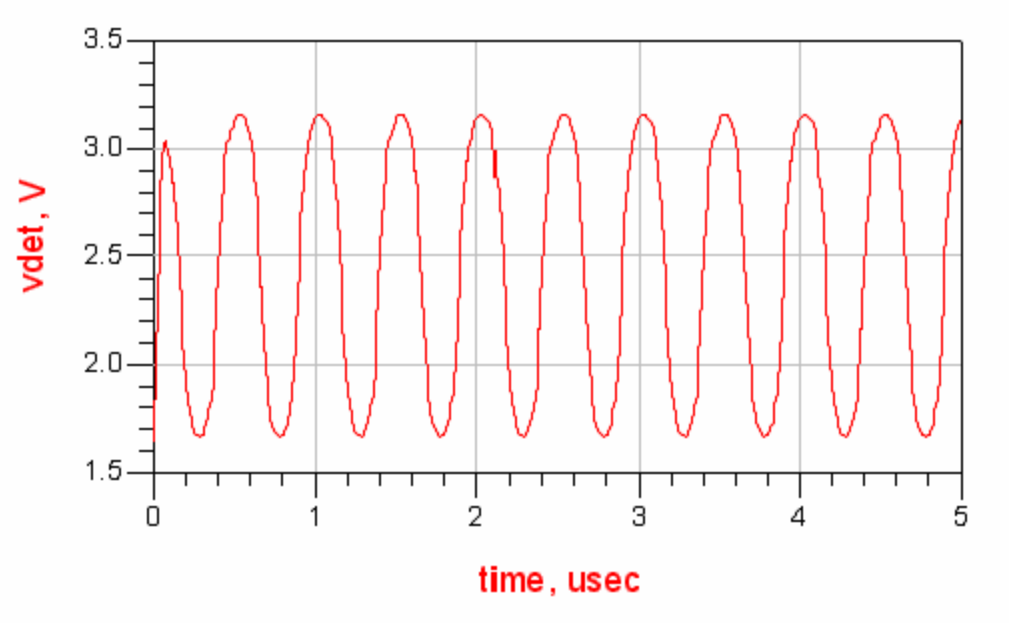


Figure 5.37 Waveform at V_{det}

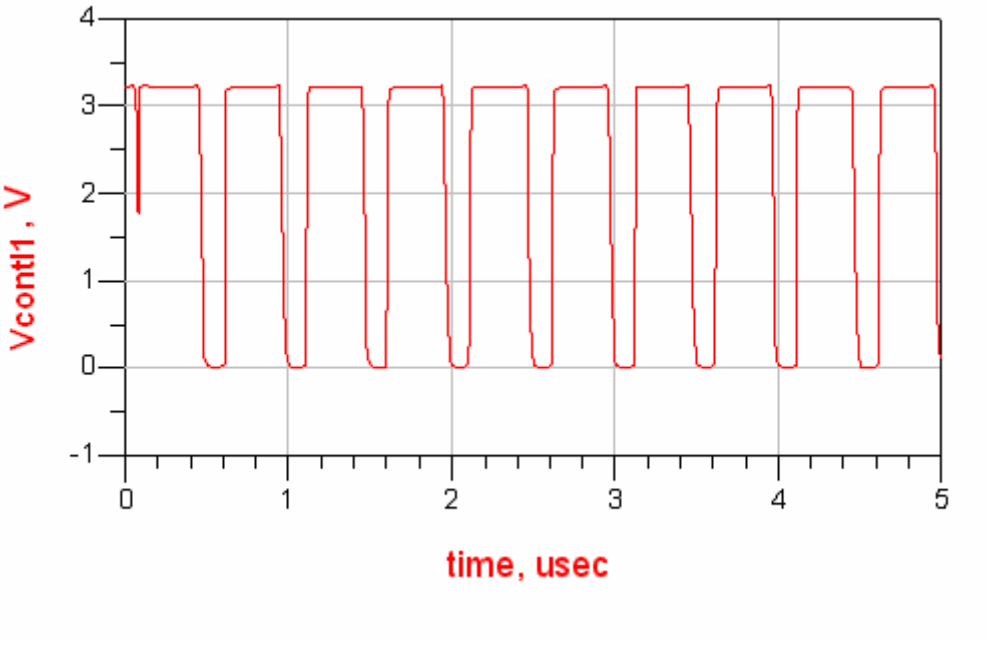


Figure 5.38 Waveform at $V_{contrl1}$

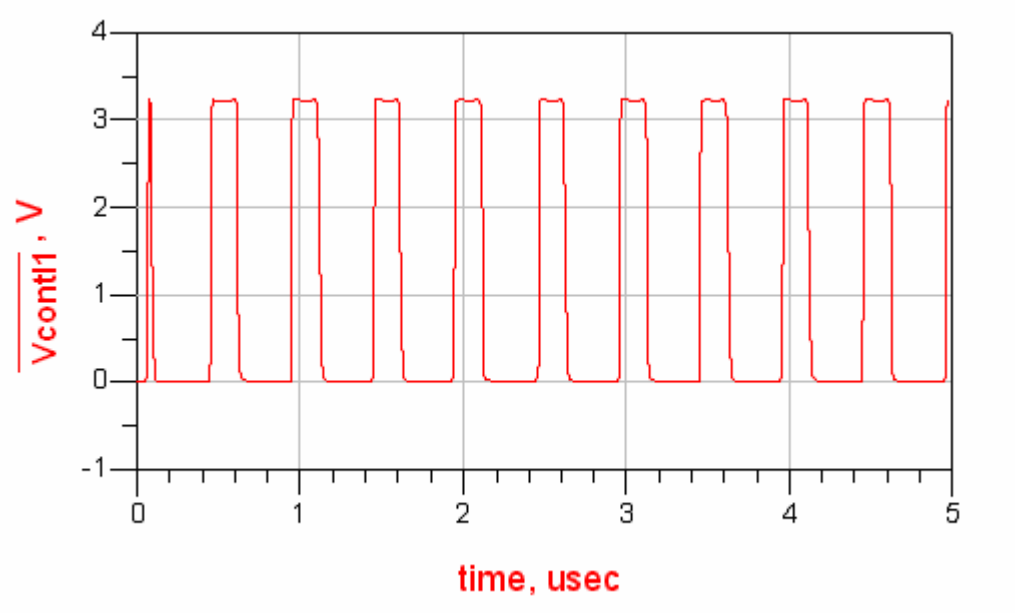


Figure 5.39 Waveform at $\overline{V_{contrl1}}$

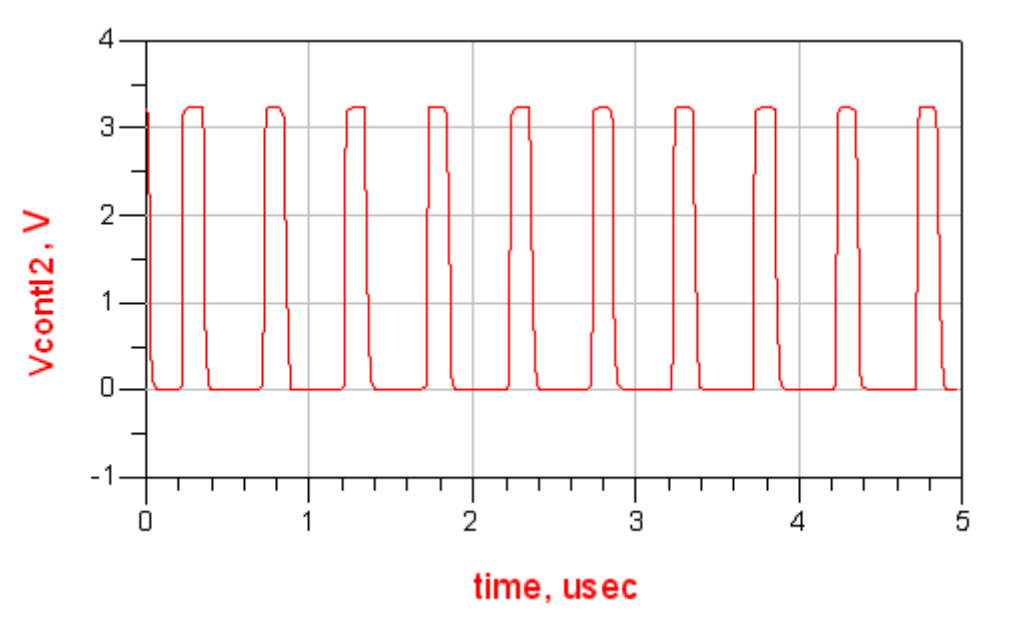


Figure 5.40 Waveform at $V_{contrl2}$

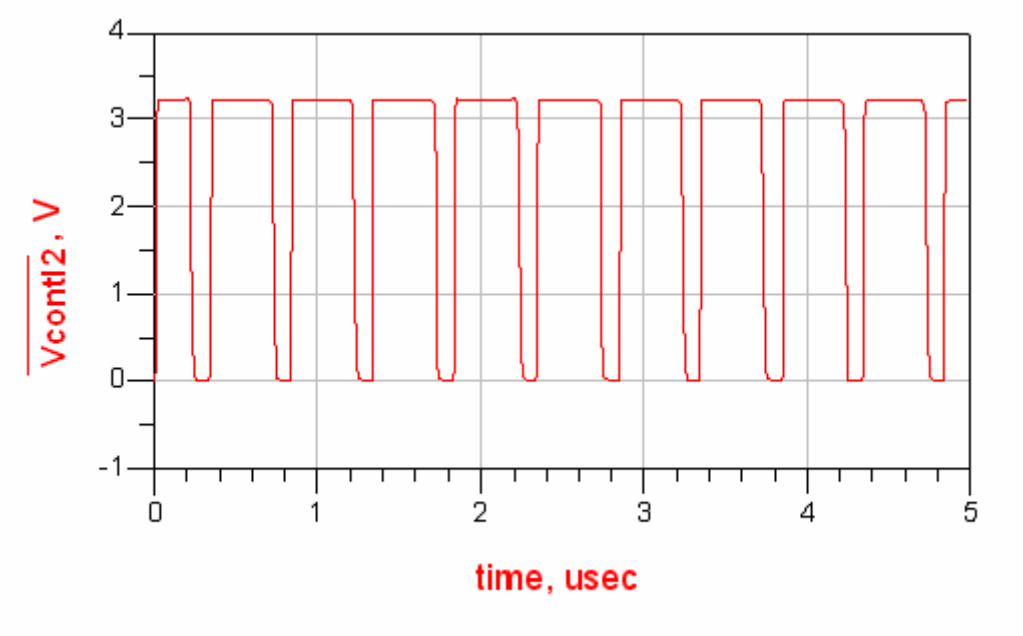


Figure 5.41 Waveform at $\overline{V_{contrl2}}$

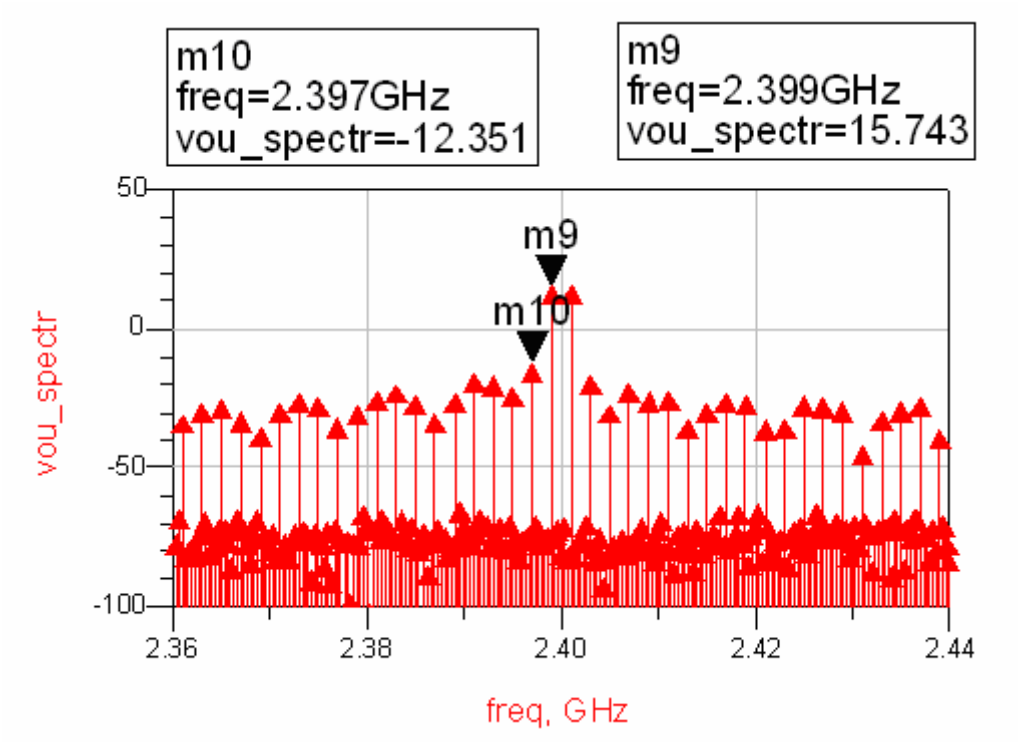


Figure 5.42 Output Spectrum

Transferring the time-domain waveform at RFout to the frequency domain, we can see the linearity. The spectrum is given in Figure 5.42. The IMD3 is about 28.1 dBc, slightly worse than the simulation of section 5.2. This degradation is partly due to the detector-path delay and partly due to feedback-caused instability. Further improvement will adjust the relative delay of V_{contl} and its complementary signal, providing a hysteresis circuit to stabilize the transition. The power saved can be seen from the current flowing through the PA's last stage, shown in Figure 5.43. Without switching the current, the quiescent current will stay at more than 600 mA all the time. With the switching, the average current is much smaller.

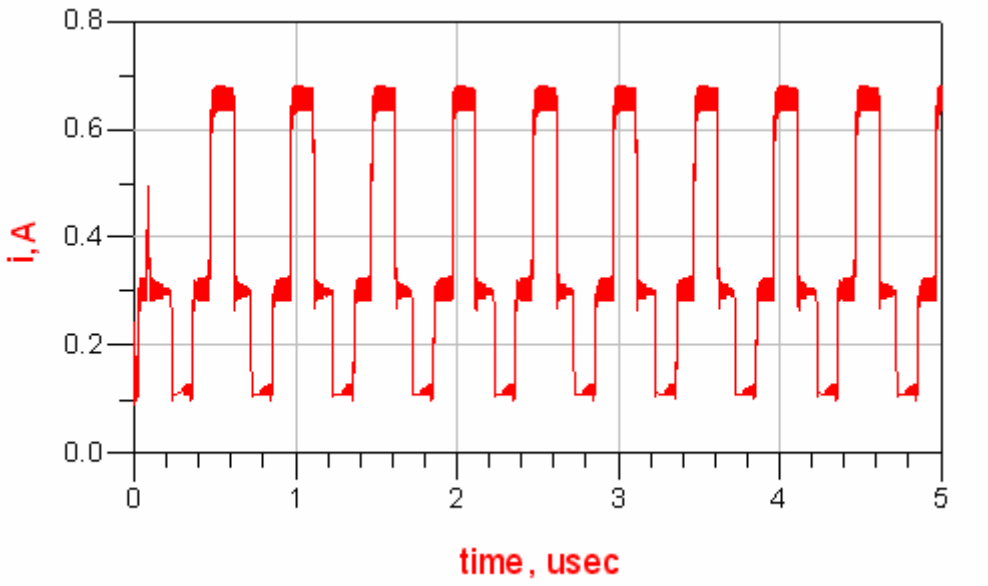


Figure 5.43 Current Flows Through the PA's Last Stage

To compare the case without the gain control, the waveform and spectrum are given in Figures 5.44 and 5.45. The current consumption is given in Figure 5.46. As expected, the IMD3 is even worse without the gain compensation, only about 20.8 dBc. (This is partly due to fact the control voltages do not track each other. In this scenario, the control voltage is not the same as in previous one, ±0.1 V. If the same voltage were applied, the

PA could only work in the half or quarter case. Recalibrating the power level gives the desired the switching points.)

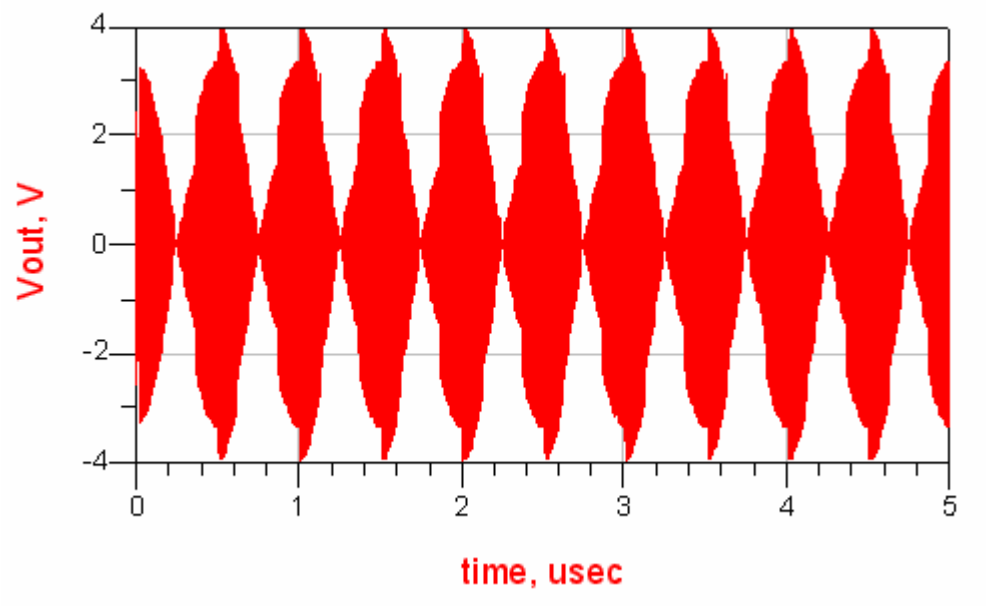


Figure 5.44 Waveform of V_{out} Without Gain Control

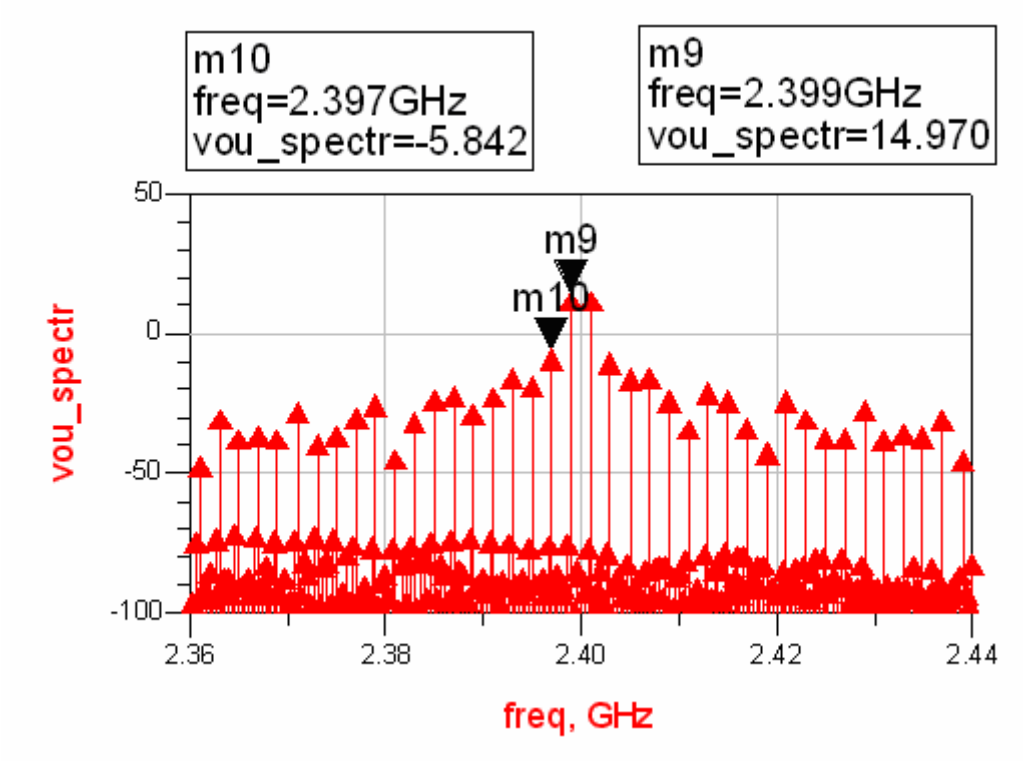


Figure 5.45 Spectrum of V_{out} Without Gain Control

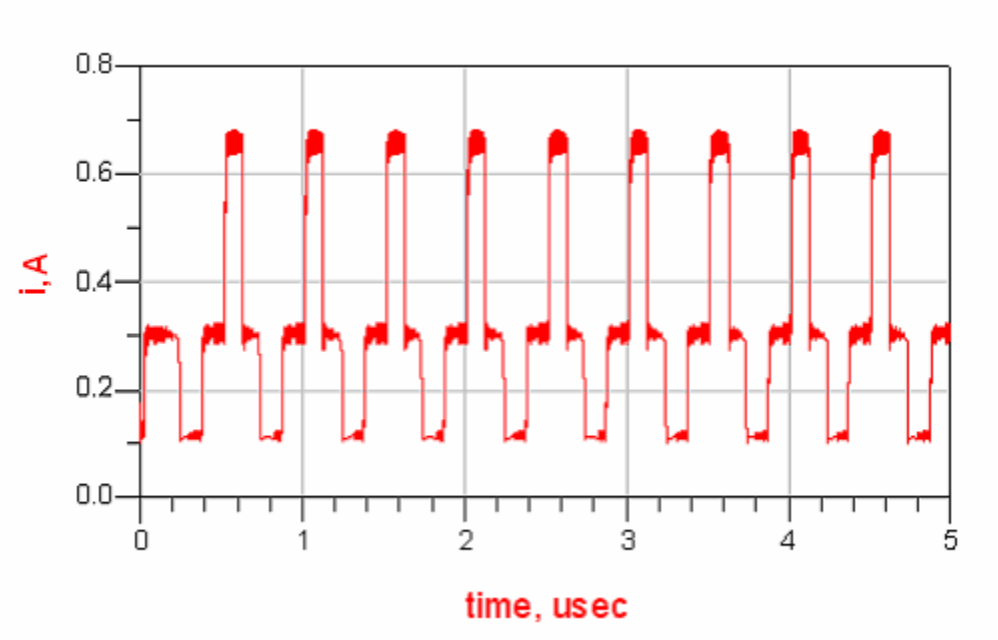


Figure 5.46 The Current Without Gain Control

The average current in Figures 5.43 and 5.46 are 329 mA and 340 mA. This difference is also due to the control voltage fluctuation. Table 5.3 summarizes the PA's characteristics. Comparing scenarios 1 and 2, the efficiency is improved by 62.7%. Comparing scenarios 3 and 2, the linearity is improved by 7.3 dB without sacrificing efficiency. In conclusion, switching the biasing current and providing gain compensation is a choice to meet both efficiency and linearity requirements.

Table 5.3 Summary of Linearity and Efficiency

	Scenario	Gain (dB)	IMD3 (dBc)	Average PAE (%)
1	w/o g_cont w/o switching	30.0	52.3	6.7
2	w/ g_cont w/ switching	29.7	28.1	10.9
3	w/o g_cont w/ switching	29.0	20.8	9.4

If the baseband signal bandwidth is 20 MHz, the simulation gives the following results. Figures 5.47–5.49 present the output waveform, correspondent spectrum and the current flowing through the last stage with the gain control.

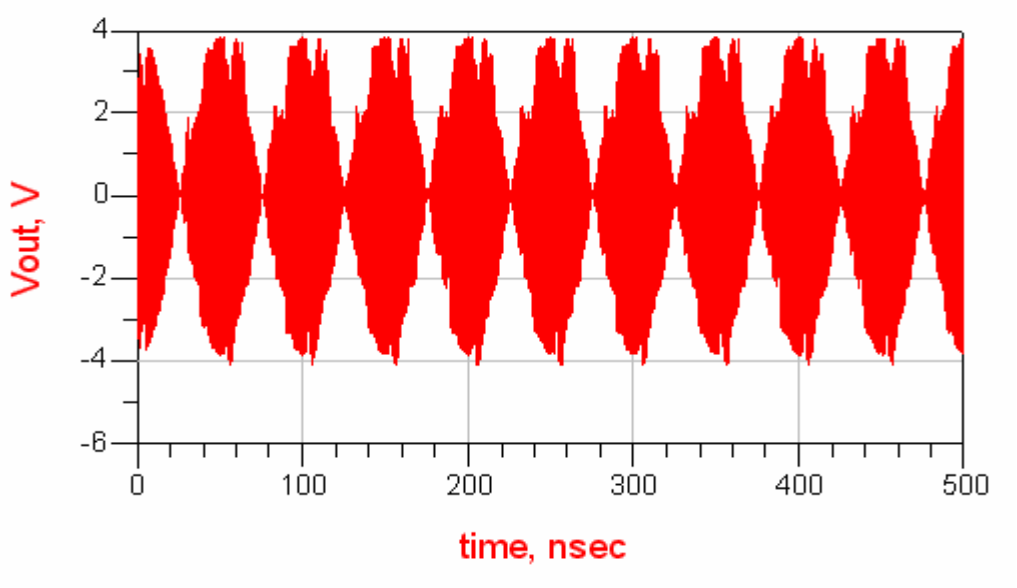


Figure 5.47 20-MHz Bandwidth Output Waveform with Gain Control

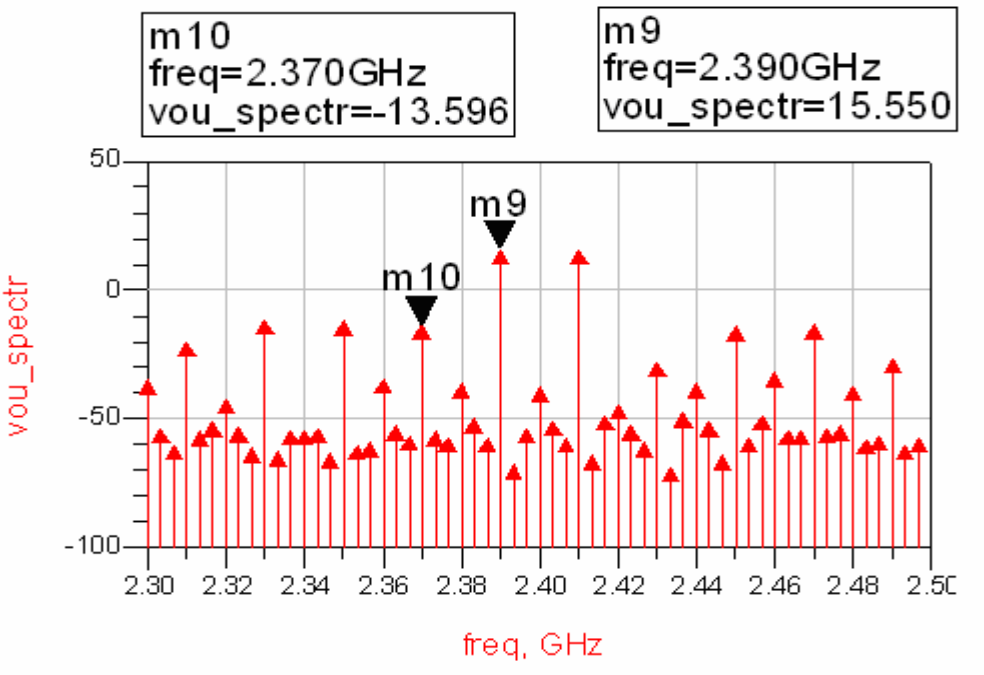


Figure 5.48 20-MHz Bandwidth Output Spectrum with Gain Control

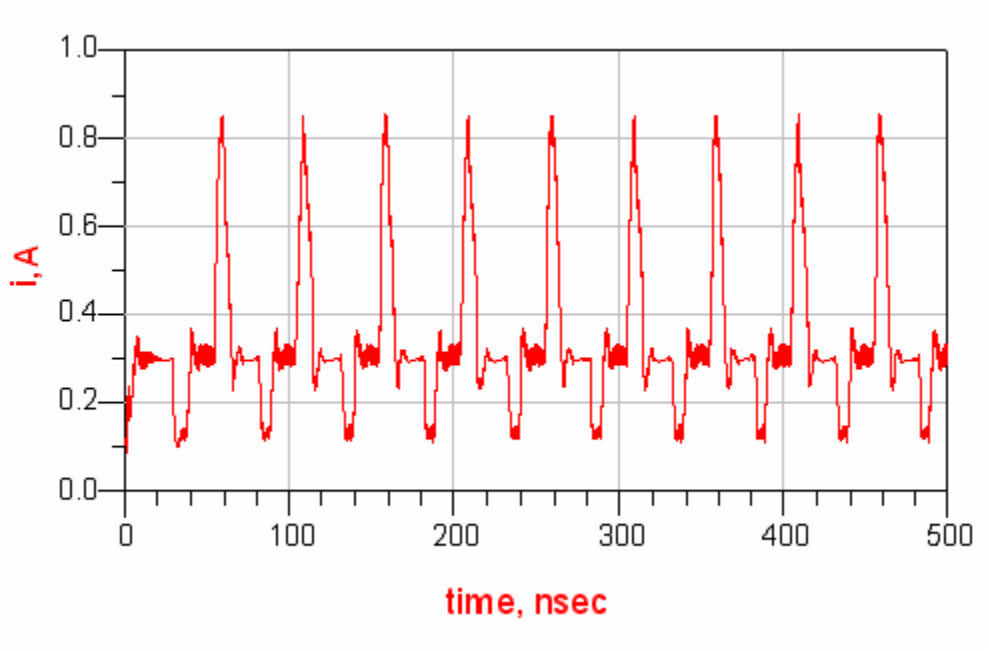


Figure 5.49 20-MHz Bandwidth Last-Stage Current with Gain Control

Figures 5.50–5.52 give the waveform, spectrum and current if there is no gain compensation.

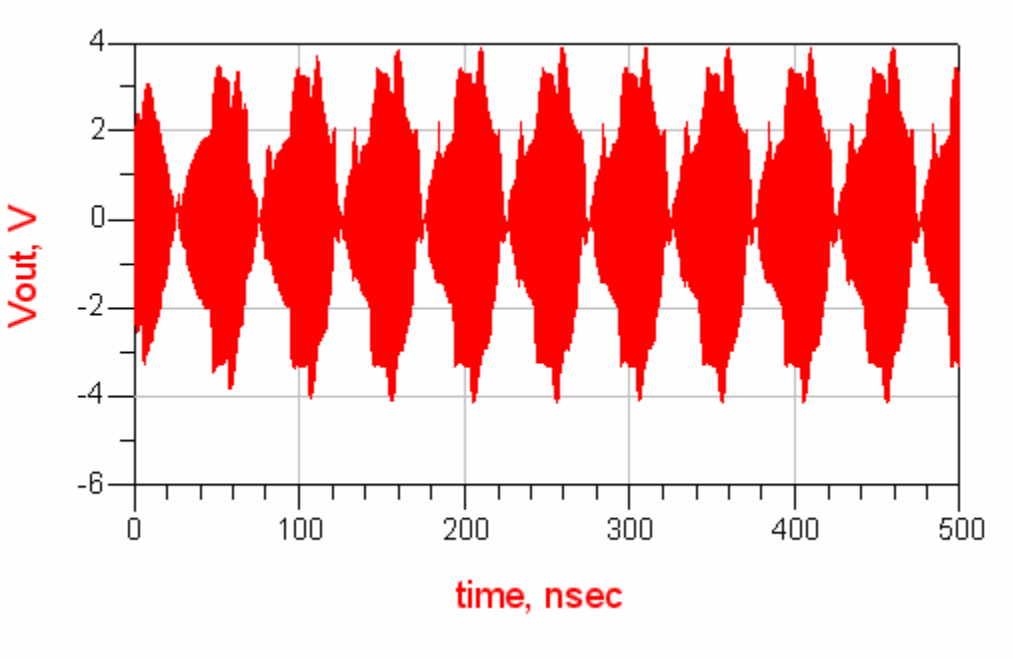


Figure 5.50 20-MHz Bandwidth Output Waveform Without Gain Control

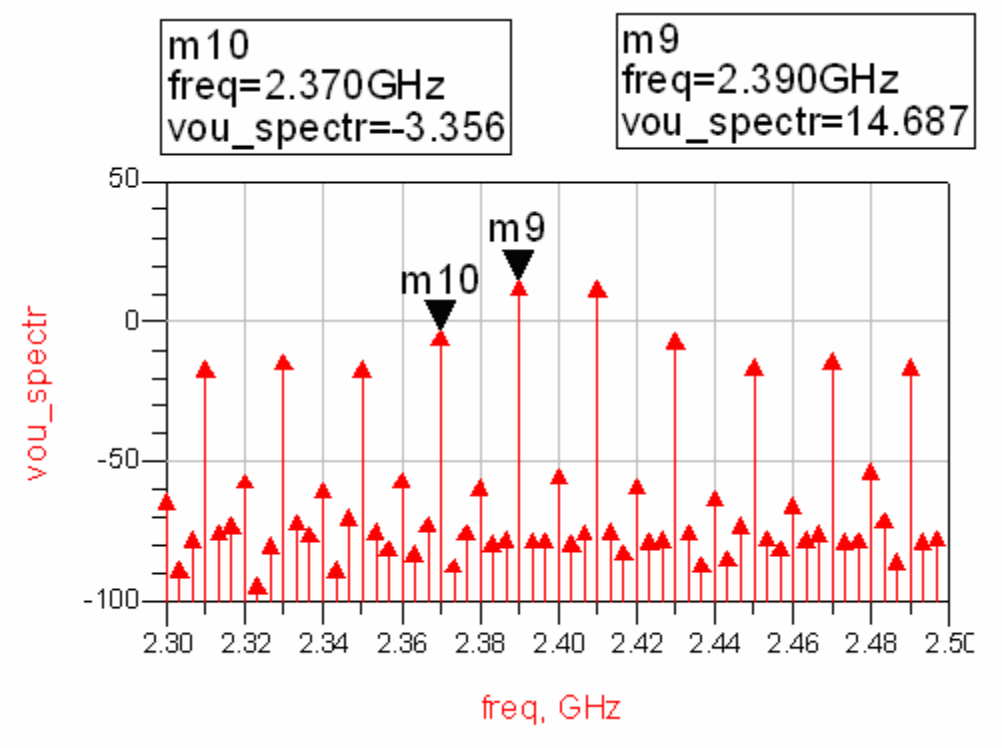


Figure 5.51 20-MHz Bandwidth Output Spectrum Without Gain Control

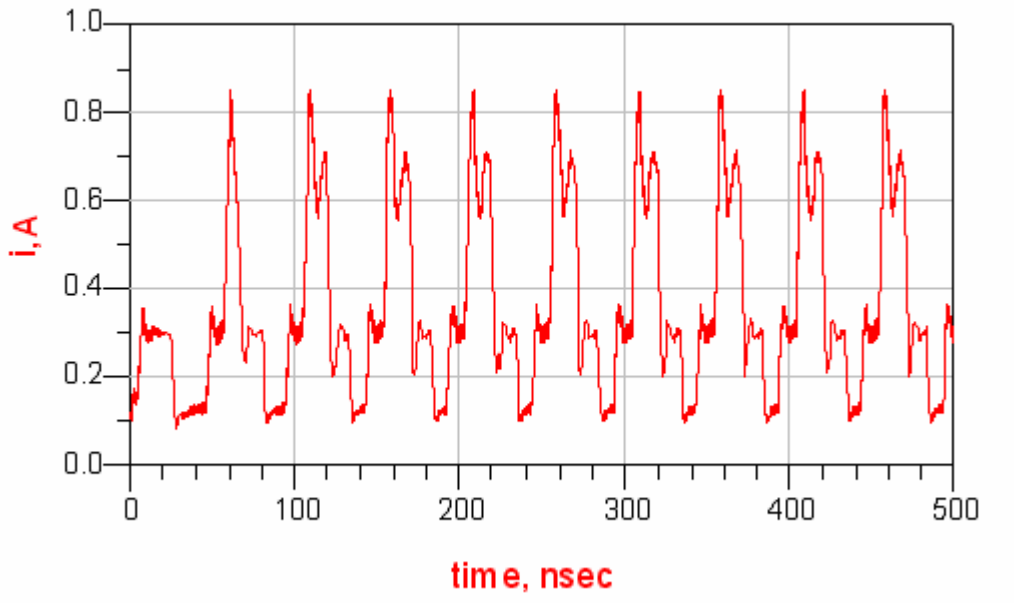


Figure 5.52 20-MHz Bandwidth Last-Stage Current Without Gain Control

It is easy to see that gain compensation improves the system's linearity. As in the 2-MHz bandwidth comparison, the 20 MHz results are listed in Table 5.4.

Table 5.4 Summary of Linearity and Efficiency

	Scenario	Gain (dB)	IMD3 (dBc)	Average PAE (%)
1	w/o g_cont w/o switching	30.0	52.3	6.7
2	w/ g_cont w/ switching	29.6	29.1	10.3
3	w/o g_cont w/ switching	29.0	18.0	8.2

The power detector and driver circuit's quiescent current is about 5.5 mA. There is significant power saved by appling the current switching and gain compensation. In this particular case, the efficiency is improved by 54%.

5.5 OFDM Input Discussion

OFDM is used in 802.11a, 802.11g and the latest, WiMax. The benefits of OFDM are high spectral efficiency, resiliency to RF interference and lower multipath distortion. This is useful because in a typical terrestrial broadcasting scenario there are multipath channels (i.e., the transmitted signal arrives at the receiver using several paths of different lengths). Since multiple versions of the signal interfere with each other ISI, it becomes very hard to extract the original information. OFDM is a multicarrier modulation. That is, signals are transmitted in parallel in many subcarriers. The problem is that this system entails a large peak-to-average ratio. For example, 802.11a/g use 52 channels and the peak-to-average ratio exceeds 13 dB. The Complementary Cumulative Distribution Function (CCDF) is used to characterize the probability of a power larger than the average power. Figure 5.53 shows a typical CCDF for 802.11g [2].

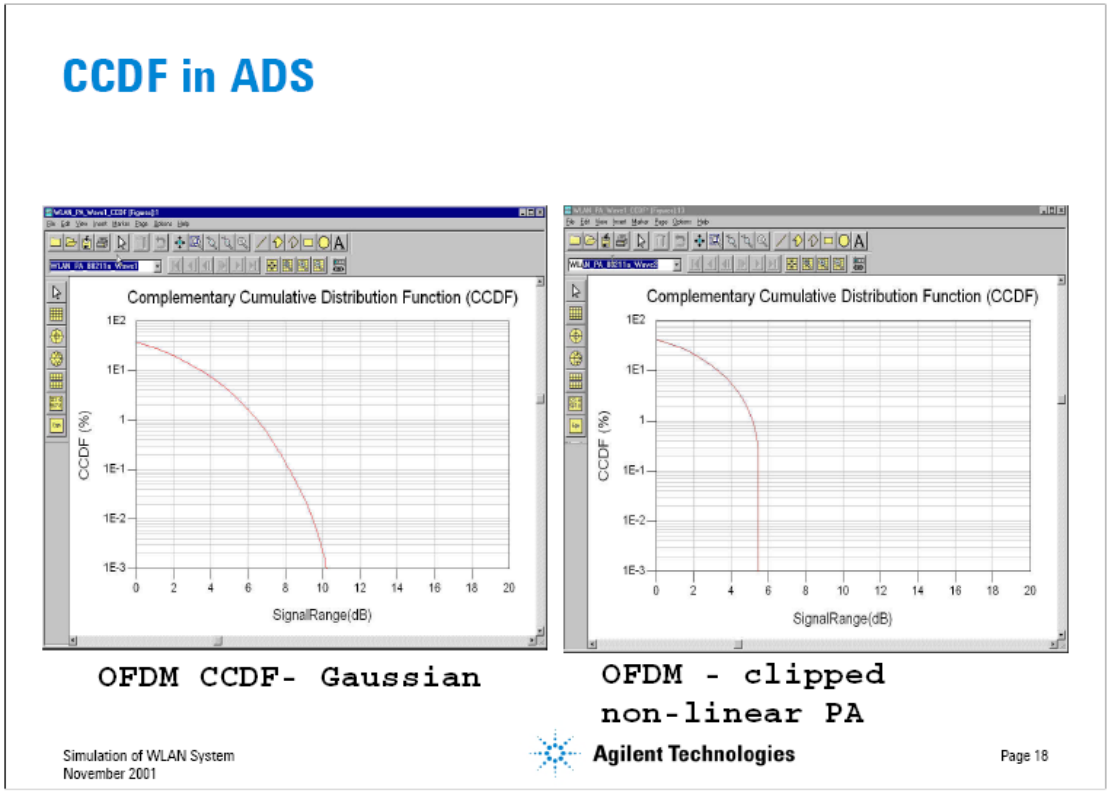


Figure 5.53 A Typical CCDF of an 802.11g Signal Source

From the figure, about 30% of the signal has an instantaneous power equal to average, 8% for a power 4 dB larger and 1.5% for a power 6 dB larger if the PA introduces no clipping. In our design, if we assume the PA's input average power is -14 dBm, when the instantaneous power is 6 dB larger, the PA works in full-current mode. When the instantaneous power is 8 dB smaller than the average power, the PA works in quarter-current mode. Otherwise, the PA works in half-current mode. From the figure, we can conclude most of time, more than 95%, the PA works in half-current mode. The probability of the signal being stronger than -8 dBm (-14 dBm + 6 dB) is less than 2%. Rough estimates suggest almost a 100% efficiency improvement. However, the high linearity requirement may limit such the improvement.

Figures 5.54 and 5.55 show the spectrum of the OFDM mask when the input average power is -14 dBm.

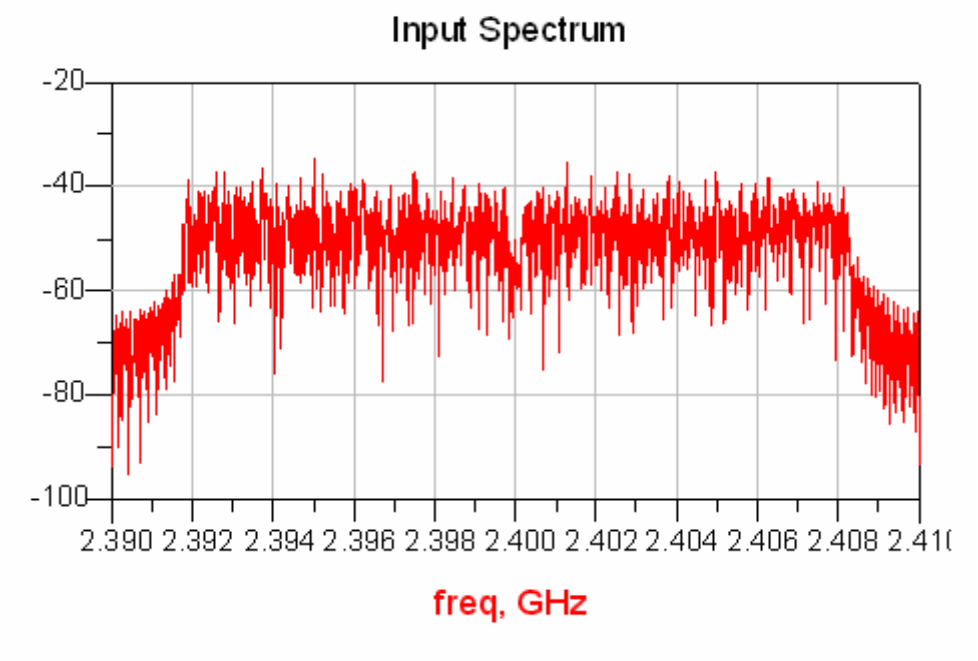


Figure 5.54 OFDM Input Mask

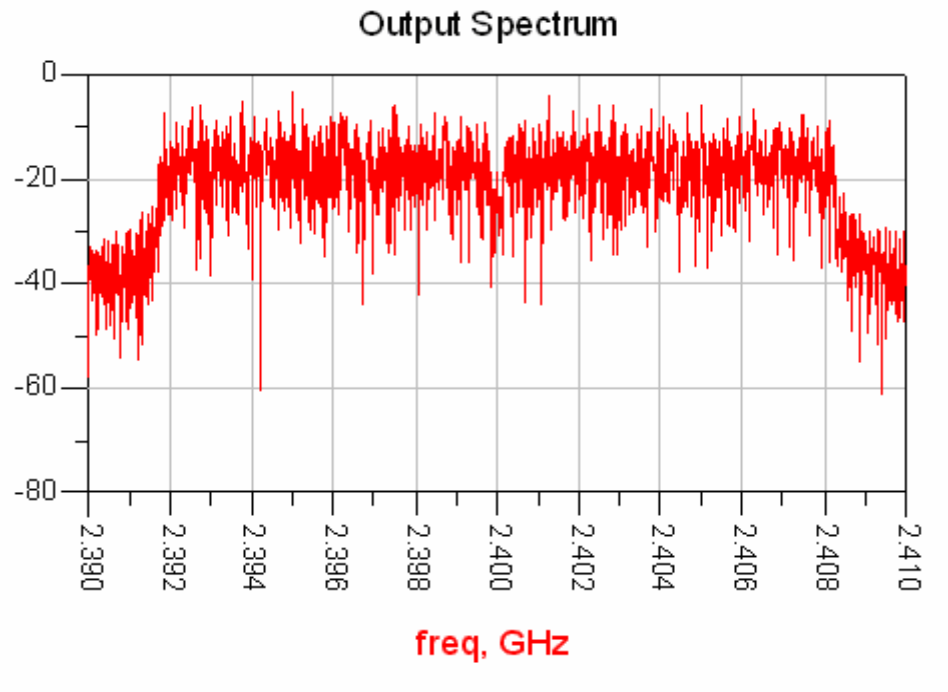


Figure 5.55 OFDM Output Mask

5.6 Summary of the PA's Characteristics

Figure 5.56 gives the return loss of the input and output, which describe the quality of the match. As stated before, the input return loss is poor, which is the reason the input is insensitive to the feedback resistance. The output return loss is very small.

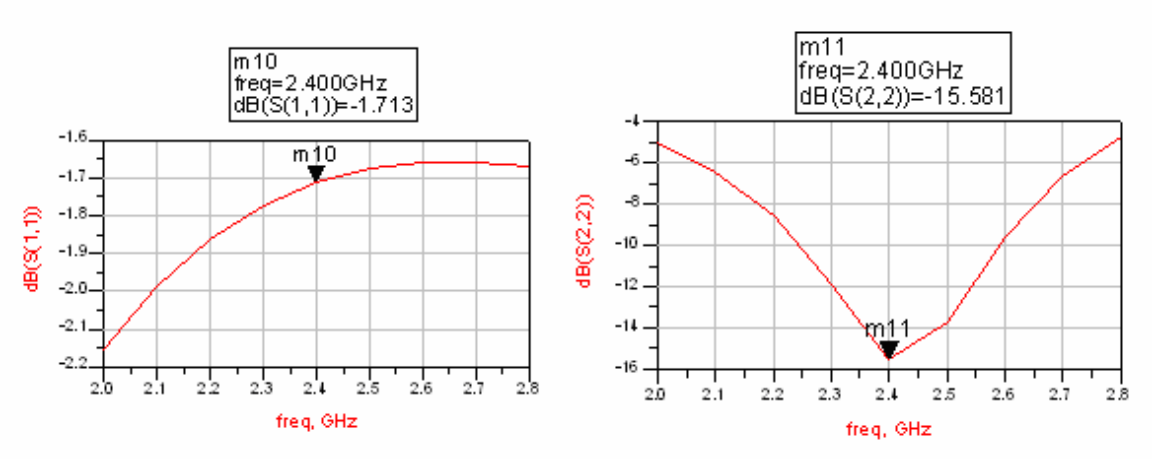
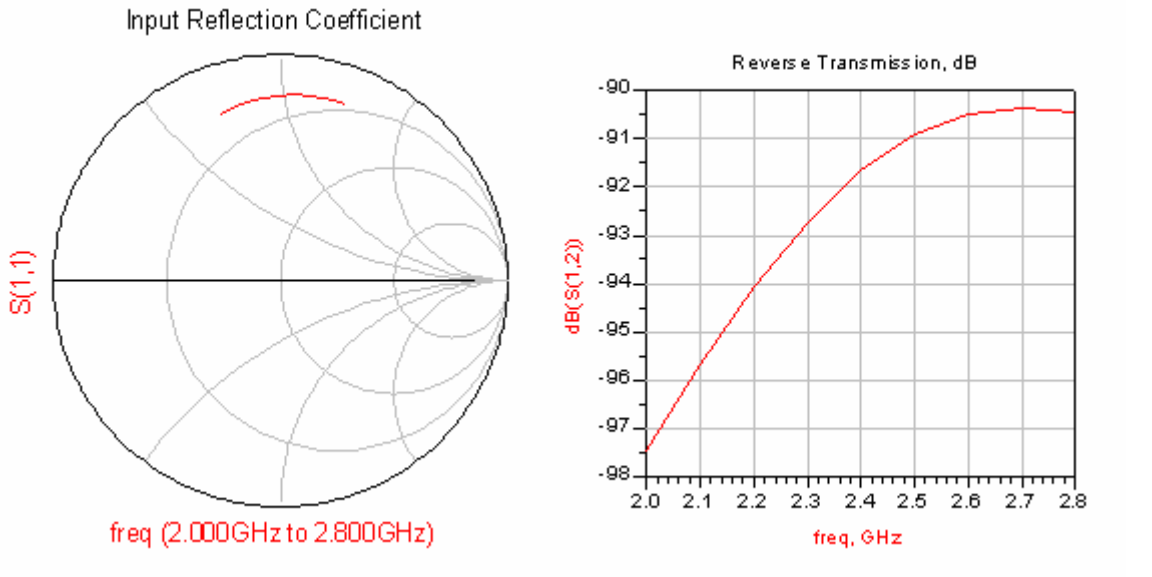


Figure 5.56 Input and Output Return Loss

The s-parameter of the PA is shown in Figure 5.57



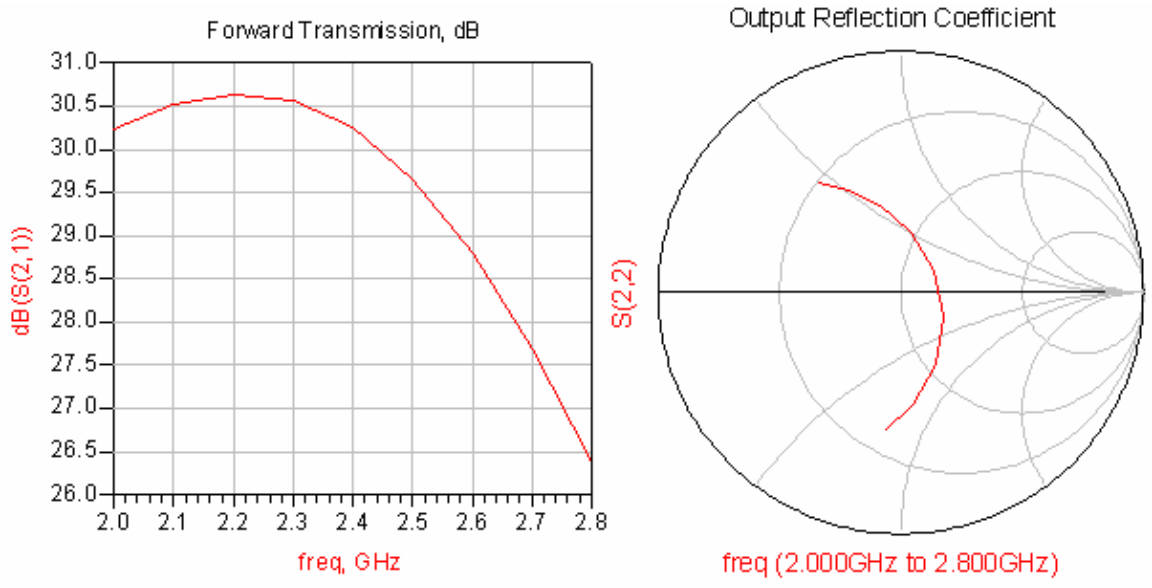


Figure 5.57 S-Parameter of the PA

References

[1] B. Bisal, R. Eline, and L. M. Franca-Neto, "RF System and Circuit Challenges for WiMAX," *Intel Technology Journal*, vol. 08, Aug., 2004.

[2] http://eesof.tm.agilent.com/pdf/wireless_networking_03.pdf

Chapter 6 Conclusion

The rapid development of mobile applications calls for new low-power RF designs. However, lower power consumption usually means less linearity. The trade-off requires a systematic study of the relationship between the noise figure, linearity and power consumption. LNA design analysis gives a good starting point to optimize the overall noise figure and linearity.

The PA consumes most of the power in RF transceiver systems. Efficiency improvements will not only extend battery life, but improve system reliability due to lower heat generation. The idea of dynamically switching the biasing current with gain control greatly improves the PA's efficiency by 76% in the ideal case and more than 50% if a real detector is included. On the other hand, the linearity will not be as good as a less efficient system. The gain-control algorithm compensates for some of the linearity loss. All in all, dynamic current switching plus gain control provides a good trade-off between efficiency and linearity.

References

Chapter 1

[1] http://www.gsmworld.com

[2] http://www.cdg.org/worldwide/cdma_world_subscriber.asp

[3] http://commerce.motorola.com/consumer/QWhtml/phone_comp.html

[4] http://www.lginfocomm.com/products/product_detail.php?p=26

[5] http://www.sonyericsson.com/us/spg.jsp?page=start

[6] http://www.nokia.com/phones/

[7] S. Mahdavi and A. Abidi, "Fully integrated 2.2-mW CMOS front end for a 900-MHz wireless receiver," *IEEE J. Solid-State Circuits*, vol. 37, pp. 662-669, May 2002.

[8] J. Ryynanen, K. Kivekas, J. Jussila, L. Sumanen, A. Parssinen, and K. A. I. Halonen, "A singal-chip multimode receiver for GSM900,DCS1800, PCS1900, and WCDMA," *IEEE J. of Solid-State Circuits*, vol. 38, pp. 594-602, Apr. 2003.

[9] D. Manstretta, M. Brandolini, and F. Svelto, "Second-order intermodulation mechanisms in CMOS downconverters," *IEEE J. Solid-State Circuits*, vol. 38, pp. 394-406, Mar. 2003.

[10] D. J. Cassan and J. R. Long, "A 1-V transformer-feedback low-noise amplifier for 5-GHz wireless LAN in 0.18-um CMOS," *IEEE J. Solid-State Circuits*, vol. 38, pp. 427-435, Mar. 2003.

[11] C. J. Debono, F. Malberti, and J. Micallef, "On the design of low-voltage, low power CMOS analog multipliers for RF application," *IEEE Tans. Very Large Scale Integration (VLSI) Systems*, vol. 10, pp. 168-174, Apr. 2002.

[12] Q. Li and J. S. Yuan, "CMOS RF low-noise amplifier design for wireless communication," *Proc. 43rd IEEE Midwest Symp. on Circuits and System*, Aug. 2000.

[13] M. Lin, Y.-m. Li, and H.-y. Chen, "A 0.18-um CMOS LNAs with sub-2 dB NF and low power for bluetooth applicatios," *Proc. 4th Int.l Conference on ASIC*, 2001.

[14] R.-C. Liu, C.-R. Lee, H. Wang, and C.-K. Wang, "A 5.8-GHz two-stage high-linearity low-voltage low noise amplier in a 0.35-um CMOS technology," *Symp. Radio Frequency Intergated Circuits (RFIC)*, 2002.

[15] M. S. J. Steyaert, B. D. Muer, P. Leroux, M. Borremans, and K. Mertens, "Low-voltage low-power CMOS-RF transceiver design," *IEEE Trans. Microwave Theory and Techniques*, vol. 50, pp. 281-287, Jan. 2002.

[16] F. Svelto, S. Deantoni, G. Montagna, and R. Catello, "Implementation of a CMOS LNA plus mixer for GPS applications with no external components," *IEEE Trans. Very Large Scale Integration (VLSI) System*, vol. 9, pp. 100-104, Feb. 2001.

[17] C.-Y. Wu and H.-S. Kao, "A 2-V low-power CMOS direct-conversion quadrature modulator with integrated quadrature voltage-controlled oscillator and RF amplifier for GHz RF transmitter applications," *IEEE Trans. Circuits and Systems*, vol. 49, pp. 123-134, Feb. 2002.

[18] K.-H. Cheng, L.-J. Tzou, W.-B. Yang, and S.-S. Sheu, "A CMOS low power voltage controlled oscillator with split-path controller," *The 8th IEEE Int. Conf. Electronic, Circuits and Systems (ICECS)*, 2001.

Chapter 2

[1] H. T. Friis, "Noise figure of radio receiver," *Proc. IRE*, vol. 32, pp. 419-422, July 1944.

[2] A. A. Abidi, G. J. Pottie, and W. J. Kaiser, "Power-conscious design of wireless circuits and systems," *Proc. IEEE*, vol. 88, pp. 1528-1545, Oct. 2000.

[3] D. K. Shaeffer and T. H. Lee, "A 1.5V, 1.5-GHz CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745-759, May 1997.

[4] B. Wang, J. R. Hellums, and C. G. Sodini, "MOSFET thermal noise modeling for analog integared circuits," *IEEE J. Solid-State Circuits*, vol. 29, pp. 833-835, July 1994.

[5] T. H. Lee, *The Design of CMOS Radio Frequency Integrate Circuits*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[6] D. J. Cassan and J. R. Long, "A 1-V transformer-feedback low-noise amplifier for 5-Ghz wireless LAN in 0.18-um CMOS," *IEEE J. Solid-State Circuits*, vol. 38, pp. 427-435, Mar. 2003.

[7] J. R. Long, "Monolithic transformers for silicon RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1368-1382, Sept. 2000.

[8] Y. Aoki, N. Hayama, M. Fujii, and H. Hida, "A 23/3-dB dual-gain low-noise amplifier for 5-GHz-band wireless applications," *IEEE GaAs Digest*, 2002.

[9] H.-W. Chui and S.-S. Lu, "A 2.17 dB NF, 5 GHz band monolithic CMOS LNA with 10 mW DC power consumption," *Symp. on VLSI Circuitd digest of Technical papers*, 2002.

[10] R.-C. Liu, C.-R. Lee, H. Wang, and C.-K. Wang, "A 5.8-GHz two-stage high-

linearity low-voltage low noise amplier in a 0.35-um CMOS technology," *Symp. Radio Frequency Intergated Circuits (RFIC)*, 2002.

[11] C. J. Debono, F. Maloberti, and J. Micallef, "A 1.8 GHz CMOS low-noise amplifier," *8th Intenal. conf. on Electronics, Circuits and Systems*, 2001.

[12] B. Chi and B. Shi, "A 1.8V 2.4 GHZ CMOS on-chip impedance matching low noise amplifier for WLAN applications," *Proc. of the 2003 Int. Symp. on Circuits and Systems*, 2003.

[13] J. Kaukovuori, M. Hotti, J. Ryynanen, J. Jussila, and K. Halonen, "A linearized 2-GHz SiGe low noise amplifier for direct conversion receiver," *Proc. of the 2003 Int. symp. on Circuits and Systems*, 2003.

[14] T.-K. Nguyen and S.-G. Lee, "A 5.2 GHz, 1.7 dB NF low noise amplifier for wireless LAN based on 0.18 /spl mu/m CMOS technology," *Proc. of 2002 3rd Int. Conf. on Microwave and Millimeter Wave Technology*, 2002.

[15] N. O. Sokal and A. D. Sokal, "Class E, a new class of high-efficiency tuned single-ended power amplifiers," *IEEE J. Solid-State Circuits*, vol. 10, pp. 168-176, June 1975.

[16] S. Xie, V. Paidi, R. Coffie, S. Keller, S. Heikman, B. Moran, A. Chini, S. P. Denbaars, U. Mishra, S. Long, and M. J. W. Rodwell, "High-linearity Class B power amplifiers in GaN hemt technology," *IEEE Microwave and Wireless Components Lett.*, vol. 13, pp. 284-286, July 2003.

[17] T. Sowlati and D. M. W. Leenaerts, "A 2.4-GHz 0.18-um CMOS self-biased cascode power amplifier," *IEEE J. Solid-State Circuits*, vol. 38, pp. 1318-1324, Aug. 2003.

[18] C.-C. Yen and H.-R. Chuang, "A 0.25-um 20-dBm 2.4-GHz CMOS power amplifier with an integrated diode linearizer," *IEEE Microwave and Wireless Components Lett.*, pp. 45-47, Feb. 2003.

[19] A. Shirvani, D. K. Su, and B. A. Wooley, "A CMOS RF power amplifier with parallel amplification for efficient power control," *IEEE J. Solid-State Circuits*, vol. 37, pp. 684-693, June 2002.

[20] M. Berkhout, "An integrated 200-W Class-D audio amplifier," *IEEE J. Solid-Sate Circuits*, vol. 38, pp. 1198-1206, July 2003.

[21] F. Fortes and M. J. d. Rosario, "A second harmonic Class-F power amplifier in standard CMOS technology," *IEEE Trans. on Microwave Theory and Techniques*, vol. 49, pp. 1216-1220, June 2001.

[22] K. L. R. Mertens and M. S. J. Steyaert, "A 700-MHz 1-W fully differential CMOS Class-E power amplifier," *IEEE J. Solid-State Circuits*, vol. 37, pp. 137-141, Feb. 2002.

[23] W. H. Doherty, "A new high efficiency power amplifier for modulated waves," *Proc. IRE*, vol. 24, pp. 1163-1182, Sept. 1936.

[24] V. Viswanathan, "Efficeincy enhancement of base station power amplifiers using Doherty technique," in *M.S. Thesis, Dept. of Electrical Engineering*: Virginia Polytechnic Institute and State University, Feb., 2004.

[25] Y. Yang, J. Cha, B. Shin, and B. Kim, "A microwave Doherty amplifier employing envelope tracking technique for high efficiency and linearity," *IEEE Microwave and Wireless Components Lett.*, vol. 13, pp. 370-372, Sept. 2003.

[26] N. Srirattana, A. Raghavan, D. Heo, P. E. Allen, and J. Laskar, "Analysis and

design of a high-effiency multistage Doherty power amplifier for wireless communications," *IEEE Trans. Microwave Theory and Techniques*, vol. 53, pp. 852-860, Mar. 2005.

[27] H. Chireix, "High power outphasing modulation," *Proc. IRE*, vol. 23, pp. 1370-1392, Nov. 1995.

[28] D. C. Cox and R. P. Leck, "A VHF implementation of a LINC amplifier," *IEEE Trans. Commun.*, vol. COM-24, pp. 1018-1022, Sept. 1976.

[29] I. Hakala, D. K. Choi, L. Gharavi, N. Kajakine, J. Koskela, and R. Kaunisto, "A 2.14-GHz Chireix outphasing transmitter," *IEEE Trans. on Microwave Theory and Techniques*, vol. 53, pp. 2129-2138, June 2005.

[30] J. Chen, K. U-yen, and J. S. Kenney, "An envelope elimination and restoration power amplifier using a CMOS dynamic power supply circuit," *2004 MTT-S International*, Jun. 2004.

[31] L. Larson, P. Asbeck, and D. Kimball, "Next generation power amplifiers for wireless communications - squeezing more bits out of fewer joules," *Radio Frequency integrated Circuits (RFIC) Symposium*, June 2005

[32] N. Srirattana, P. Sen, H. M. Park, C. H. Lee, P. E. Allen, and J. Laskar, "Linear RF CMOS power amplifier with improved efficiency and linearity in wide power levels," *Radio Frequency integrated Circuits (RFIC) Symposium, 2005*, June 2005

[33] W.-C. Hua, H.-H. Lai, P.-T. Lin, C. W. Liu, T.-Y. Yang, and G.-K. Ma, "High-linearity and temperature-insensitive 2.4 GHz SiGe power amplifier with dynamic-bias control," *Radio Frequency integrated Circuits (RFIC) Symposium*, June 2005

[34] Y. W. Kim, K. Han, S. Hong, and J. Shin, "A 45% PAE/18mA quiescent current CDMA PAM with a dynamic bias control circuit," *2004 IEEE Radio Frequency Integrated Circuits Symp.*, 2004.

[35] Y. Yang, "High efficiency CDMA power amplifier with dynamic current control circuits," *IEEE Compound Semicondector Integrated Circuit, digest*, 2004.

[36] N. Schlumpf, M. Declercq, and C. Dehollain, "A fast modulator for dynamic supply linear RF power amplifier," *IEEE J. Solid-State Circuits*, vol. 39, pp. 1015-1025, July 2004.

[37] S. C. Cripps, *RF power amplifier for wireless communications*. Boston: Artech House, 1999.

Chapter 3

[1] D. K. Shaeffer and T. H. Lee, "A 1.5V, 1.5-GHz CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 745-759, May 1997.

[2] P. Andreani and H. Sjoland, "Noise optimization of an inductively degenerated CMOS low noise amplifier," *IEEE Trans. on Circuits and System-II: Analog and Digital Signal Precessing*, vol. 48, Sept. 2001.

[3] J. Ou, X. Jin, P. Gray, and C. Hu, "Recent Developments in BSIM for CMOS RF ac and Noise Modeling," *Advances in Analog Circuit Design*, Nice, France, March 1999.

[4] P. Rossi, F. Svelto, A. Mazzanti, and P. Andreani, "Serendipitous noise reduction in inductively degenerated CMOS RF LNAs," *Proceedings of NORCHIP '03*, Nov. 2003.

[5] https://www.mosis.com.

[6] S. Donati, M. A. Alam, K. S. Krisch, S. Martin, M. R. Pinto, and H. H. Vuong, "Physics-based RF noise modeling of submicron MOSFETs," Dec. 1998.

[7] T. Manku, "RF simulations and physics of the channel noise parameter within MOS transistors," *Proc. IEEE Custom Integrated Circuits Conf.*, May 1999.

Chapter 4

Chapter 5

[1] B. Bisal, R. Eline, and L. M. Franca-Neto, "RF System and Circuit Challenges for WiMAX," *Intel Technology Journal*, vol. 08, Aug., 2004.

[2] http://eesof.tm.agilent.com/pdf/wireless_networking_03.pdf.