This is to certify that the
thesis entitled

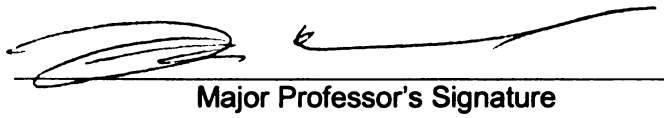
CMOS VARIABLE GAIN AMPLIFIER (VGA) FOR APPLICATION IN RF FRONT END OF A SATELLITE TV TUNER

presented by

NICK JOEL ROSIK

has been accepted towards fulfillment
of the requirements for the

M.S. degree in Electrical Engineering



Major Professor's Signature

06 May 2004

Date

CMOS VARIABLE GAIN AMPLIFIER (VGA) FOR APPLICATION IN RF FRONT END OF A SATELLITE TV TUNER

By

Nick Joel Rosik

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Electrical and Computer Engineering

2004

ABSTRACT

CMOS VARIABLE GAIN AMPLIFIER (VGA) FOR APPLICATION IN RF FRONT END OF A SATELLITE TV TUNER

By

Nick Joel Rosik

RF front ends are a very highly researched topic in the world of wireless communications. Research is focused in this area because the front end is arguably the most vital component in making a communication system function properly. The front end has three functions. The first is to amplify or attenuate the RF input signal to provide a constant amplitude output to the analog to digital converter (ADC). Second, it must provide programmable channel selection by means of a phase-locked loop (PLL). And third, to prevent aliasing in the ADC, it must also low-pass filter the base-band output. The main building blocks which compose the front end are: the variable-gain amplifier (VGA); Mixer; PLL; and low-pass filter. The primary goal of this thesis is to design and simulate a VGA in all CMOS technology for use in a satellite TV tuner. Previous solutions for VGA's in this application have used more expensive and less integrated Silicon Bipolar or GaAs technologies. Therefore, the benefit to completing this design in all CMOS is to reduce cost and to improve overall tuner integration. The design and simulations will be accomplished using a 0.35μm CMOS technology and there are a number of specifications which the VGA design must meet including Matching, Noise Figure (NF), Linearity (IIP3), Input Return Loss (IRL), and Dynamic Range.

Dedicated to my mother, Theresa Rosik. Thank you for your love, support and for instilling me with a desire to learn.

ACKNOWLEDGEMENTS

I would like to thank Conexant Systems and Jim Moniz for providing me with the internship opportunity in the Boston Design Center that made completing this thesis possible.

I would also like to express my gratitude to Dr. Donnie Reinhard and Dr. Leo Kempel for their assistance and support of my thesis.

Finally, I would also like to give a special thanks to my family and friends for all of their love and support. My brother Ray has been an excellent advisor, mentor, and brother and the knowledge I have gained from him has been invaluable. He taught me to believe in myself, even when things seem impossible, and without all of his help, I could not have completed this thesis. To my Uncle Tom, thank you for all of your support along the way. I will certainly miss all of our Sunday dinners, conversations and Red Wing games. For my father, Robert, I hope that you feel proud knowing that both of your sons have been successful in earning masters degrees in electrical engineering. Most of all, I am extremely grateful to my mother Theresa, who has made many sacrifices to ensure that I had the finest education and the best chance to succeed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Background and Motivation

In the last decade, satellite communication has been developing rapidly. Due to its improved performance and reduced cost, satellite TV can offer a highly attractive alternative to standard cable TV. Additional improvements in receiver integration will further reduce cost and improve performance.

A satellite TV system is composed of several highly complex functional blocks. A typical satellite TV system is illustrated as a block diagram in Figure 1.1 [1].

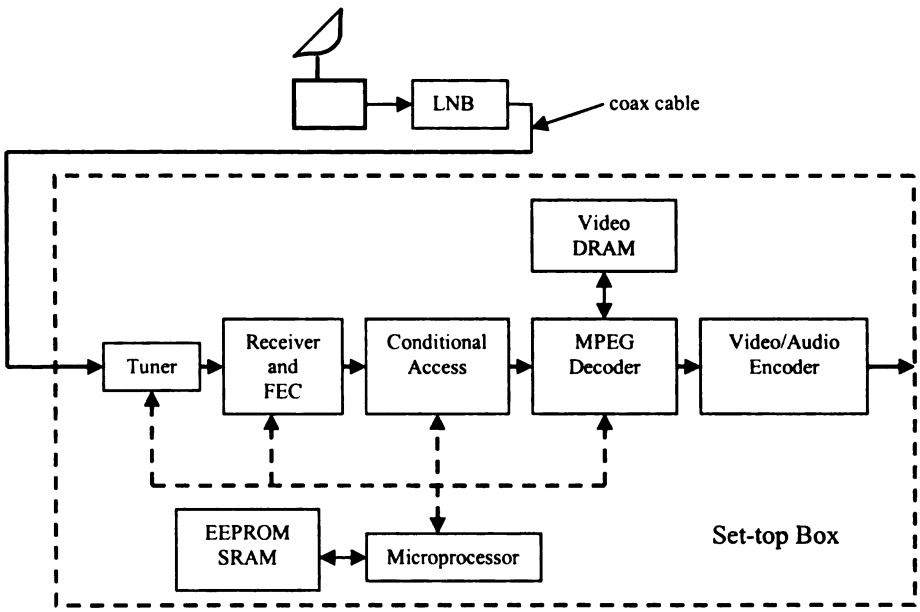


Figure 1.1: Block Diagram of Satellite TV System

The first blocks of the system, located at the dish are the antenna and the low noise block converter (LNB). These blocks receive the input satellite in the K-band (11-12 GHz) and down convert to 925 MHz to 2175 MHz. The down converted signal is then routed to the set-top box inside the house on a coax cable. The next block in the system is the tuner,

which includes the variable gain amplifier (VGA), mixer, phase locked loop (PLL) and base-band filters. The tuner has three functions. First, the tuner must amplify or attenuate the input signal to provide a constant amplitude output for the analog to digital converter (ADC) in the receiver/FEC block. Secondly, it must provide programmable channel selection. Third, the tuner must low-pass-filter the base-band output to prevent aliasing in the ADC. The tuner is also commonly referred to as the analog front end. This thesis will focus on the VGA and implementation in CMOS. The remaining back-end system blocks include the receiver/FEC, the MPEG decoder and Video/Audio encoder. These blocks include an ADC and digital signal processing circuitry to demodulate and decode the input signal. The blocks forming the backend system are beyond the scope of this thesis.

1.2 Design Challenges

Today's satellite systems are all digital communication systems. With the exception of the analog front end, all of the signal processing is often implemented into high-density CMOS integrated circuits. CMOS is very attractive for these digital circuits due to advantages including very high density, low power, and low cost. The low cost and potential for higher levels of system integration make CMOS also attractive for RF/Analog circuits as well

Despite the many advantages of implementation of RF and analog circuits in CMOS, there are many challenges to maintaining the required performance. These challenges include lower supply voltages, lower cut off frequencies, lower transconductance, and increased crosstalk through the substrate to name a few.

Currently, GaAs, Bipolar, and BiCMOS are the most attractive technologies for RF circuits. These technologies offer improved high frequency performance as defined by the cutoff frequency (f_T). The cutoff frequency is defined as the frequency at which the current gain is extrapolated to fall to unity. The process f_T is an indicator of the frequency band for which the process can be used for RF circuits. For example, SiGe BiCMOS f_T ranges from 25GHz to 120GHz. However, CMOS f_Ts only ranges from 10 to 50 GHz. These lower f_Ts for CMOS quantify why RF design in CMOS has many additional challenges.

A second characteristic which makes RF design in CMOS extremely challenging is device transconductance (g_m). Because of the physics of the devices, the transconductance of bipolar transistors is much greater than that of CMOS transistors. This is of great importance for amplifier design, such as the VGA required for the analog front end, because amplifier gain is proportional to the transconductance of the device. The gain in turn has a direct correlation to specifications such as noise factor and linearity which characterize the circuits performance. Hence overcoming the device physics limitations makes high frequency VGA design in CMOS technology challenging.

A third challenging aspect to this project is the fact that the satellite system specifications require the VGA to have a very high dynamic range in addition to having nearly constant broadband operation from 925MHz to 2175MHz. There are many published papers which demonstrate the possibility of realizing CMOS VGAs in the gigahertz frequency range [2] – [4]. However, the majority of these publications deal with either narrowband applications in the range of a 100MHz to 300MHz or applications which have a very limited dynamic range. Therefore, these designs fail to be viable

solutions for this type of broadband design. There are very few papers which demonstrate the ability to provide the required dynamic range over such a wide frequency range.

Design specification trade-offs are arguably the most challenging task in designing the VGA. These tradeoffs occur in noise figure and linearity. For example, a VGA design which is optimized to provide the maximum gain and minimum noise figure will inherently have poor linearity. Conversely, a design which is optimized to provide the best linearity will subsequently have a smaller maximum gain and higher noise figure. These contradictions occur because transistors are intrinsically non-linear devices. However, they can be approximated as linear over a certain range and that range is where amplifiers are designed to operate. The range in which the linear approximation holds is determined by the gain of the transistor. As the gain increases, the input range at which the linear approximation holds is decreased. Also, as will be discussed in a later chapter, noise figure has a positive correlation to the gain. Therefore, optimizing the design so that the specifications for these three criteria are met can be rather exhausting.

A final challenge in completing the VGA design is making it fully monolithic. This is a difficult task because in order to ensure that specifications such as matching and low noise figure are maintained the use of inductors and capacitors may be required. However, a couple problems occur when these monolithic circuit elements are used. One problem deals with size. Capacitors and inductors that are implemented on-chip are rather large in size and can take up a significant area of the chip. Therefore, there is only a limited range of values which capacitors and inductors can be designed for to ensure a

reasonable size. A second and more troubling problem is the fact that monolithic inductors suffer from a low Q, which can have a negative effect on the performance.

1.3 Design Specifications

The design specifications for the all CMOS VGA presented in this thesis were developed so as to offer performance competitive to all SiGe bipolar VGAs in a BiCMOS process. The requirements are as follows:

- Must be broadband and cover the frequency range of 925 MHz to 2.175 GHz
- A Dynamic Range of -25 dB to ~ +7 dB
- Noise Figure with a 6 to 7 dB roof at maximum gain
- Linearity: 8 dBm input-referred third order intercept (IIP3) floor at low gain
- Input Return Loss of ~ 10dB or greater
- Vcc current drain of ~ 20 mA or less
- Feed a 150 Ω differential load at output (Mixer)

1.4 Thesis Organization

The thesis is organized into four chapters. Chapter 2 gives a theoretical discussion of noise, linearity, and microwave circuit theory. Chapter 3 discusses a previously published CMOS VGA, its shortcomings and the final VGA design which takes the form of a 6 stage current steering VGA. The simulation results of the final design are detailed in Chapter 4. The final chapter, Chapter 5, gives a brief summary of the results and possible areas of future work.

Chapter 2: Theory

2.1 Introduction

The performance of a RF VGA is generally characterized by the following metrics: noise, linearity, and input return loss. Sufficient understanding of these quantities is essential to analyze the performance of the VGA. Therefore, the theory associated with each topic is presented in the following sections.

2.2 Noise

Noise is a random phenomenon which plays a crucial role in microelectronic design and can be particularly troubling to communication systems with signal power levels on the order of -80 to -100 dBm such as with satellite TV. The reason noise is significant in such an application is because it limits the minimum signal which the system can process with acceptable quality. Also, because of its random and time-varying nature, it is difficult to characterize and is usually incorporated into circuit analysis using statistical models such as power spectral densities (PSDs). A PSD is a measure of the average power of a waveform in a one-hertz bandwidth [8]. Noise comes in one of two forms: environmental noise or device electronic noise. The following discussion focuses on the latter, which is caused by small current and voltage fluctuations generated within CMOS devices themselves.

Thermal noise and flicker noise are the noise mechanisms which effect CMOS devices the most. They are uncorrelated, and therefore they can be analyzed independently with their results added via superposition. Thermal noise is caused by the

random motion of electrons in a conductor which introduce fluctuations in the voltage measured across it. This type of noise occurs in linear resistors as well as the channel of MOSFETs. Circuit noise calculations are performed by modeling the PSDs of these elements as a series voltage source for the resistor, and as a parallel current source connected between drain and source for the MOSFET. The PSD for the noise in a resistor is given by [6]:

$$\overline{V_n^2} = 4kTR * \Delta f \tag{2-1}$$

where V_n^2 is the mean square noise voltage generated by resistor R in a bandwidth Δf, k is Boltzmann's constant, and T is the absolute temperature. Similarly, the PSD of the noise generated by the MOS channel is defined as [6]:

$$\overline{I_n^2} = 4kT(\frac{2}{3}g_m) \tag{2-2}$$

where g_m is the transconductance of the transistor.

Flicker noise is the second form of noise common to CMOS devices and it arises from random trapping of charge at the oxide-silicon interface of MOSFETs. Since flicker noise depends on the "cleanness" of the oxide-silicon interface, the average power of it isn't easily predicted, and its PSD is roughly modeled as a voltage source in series with the gate and is given by [6]:

$$\overline{V_n^2} = \frac{K}{WLC_{ox}} * \frac{1}{f} \tag{2-3}$$

where K is a process dependent constant, W and L are the width and length respectively of the transistor, and C_{ox} is the gate oxide capacitance. It is often referred to as 1/f noise due to its inverse frequency dependence and therefore, it is most significant at lower

frequencies. Figure 2.1 illustrates a typical noise voltage PSD for a MOSFET including both thermal and flicker noise contributions [9].

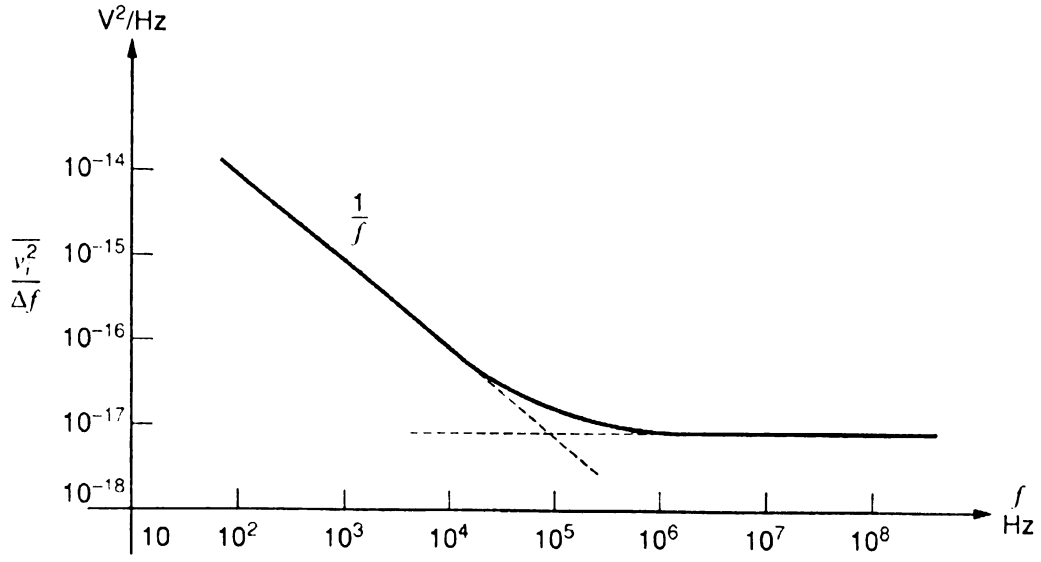


Figure 2.1: Typical equivalent input noise voltage PSD for a MOSFET [9].

The input-output response to noise in a circuit is generally the most important aspect to designers. Therefore, noise is usually characterized using two port theory with less emphasis on the individual models discussed previously. In order to fairly compare the performance of different circuits independent of gain, it is also usually calculated in an "input referred" fashion. It is commonly done this way, because the output noise is proportional to gain, and it can therefore be a misleading comparison between circuits with different gains [8]. By defining noise in the input referred fashion, a noisy circuit is modeled by a noiseless circuit with two input noise generators, consisting of a series voltage source and a parallel current source. The noise generators then encompass all the effects of noise sources in the circuit. Figure 2.2 shows this representation [8].

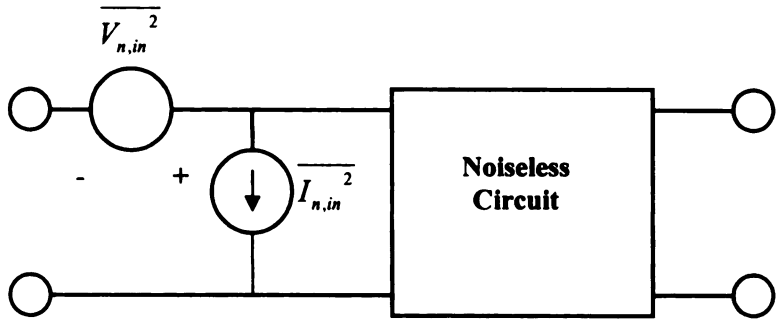


Figure 2.2: Representation of noise by input noise generators.

The current source is included in the model to represent the effects of a finite input impedance. If the model included a voltage source alone, it would imply that circuits with large source impedances would have no output noise, which is untrue.

Since the majority of VGAs, including the one presented in this thesis, use differential pairs to perform signal amplification, the noise behavior of such an amplifier is important [8]. Figure 2.3 shows a basic differential pair circuit and its representation with input-referred noise sources.

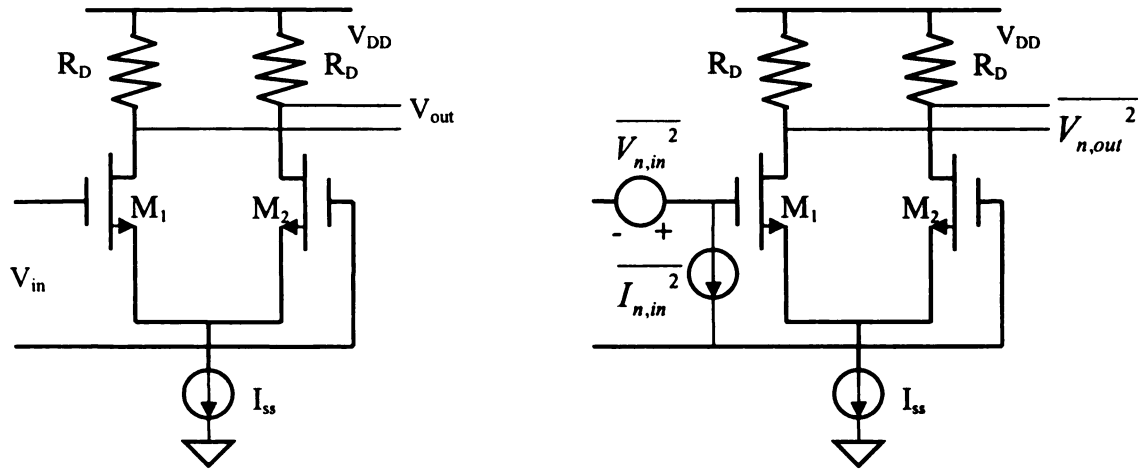


Figure 2.3: Differential pair and circuit including input referred noise sources.

Identifying the individual sources of noise in a circuit is the first step in calculating the input noise generators. For the differential pair, transistors M_1 and M_2 each contribute a thermal and flicker noise component to the output. In addition, the two resistors labeled R_D add thermal noise. Since these noise sources are uncorrelated, their effects can be calculated individually using superposition. Note that the following calculations exclude any noise from the tail current source because its effects are usually negligible. Figure 2.4 is the schematic of the differential pair including the noise sources.

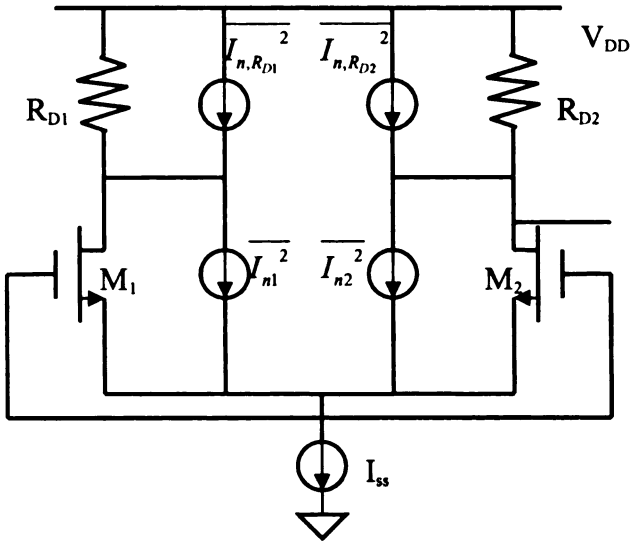


Figure 2.4: Differential pair with noise sources.

Note that the current sources labeled $\overline{I_{n1}^2}$ and $\overline{I_{n1}^2}$ include both the thermal and flicker noise of M_1 and M_2, eventhough they are calculated independently. Beginning with the thermal component of M_1, if $R_{D1} = R_{D2} = R_D$ then the noise at the output is calculated by:

$$\overline{V_{n,out}^2} = \overline{I_{n1}^2} * R_D^2 \tag{2-4}$$

Similarly, the output noise due to the thermal component of M_2 is:

$$\overline{V_{n,out}^2} = \overline{I_{n2}^2} * R_D^2 \tag{2-5}$$

Then, the sum of all the thermal noise components including the thermal noise from resistors R_{D1} and R_{D2} is given by:

$$\overline{V_{n,out}^{\ 2}} = 8kT(\frac{2}{3}g_m R_D^{\ 2} + R_D) \tag{2-6}$$

Finally, the total output noise of the circuit including both the thermal and flicker components of $M_{1,2}$ is:

$$\overline{V_{n,out,tot}^{\ 2}} = 8kT(\frac{2}{3}g_m R_D^{\ 2} + R_D) + \frac{2Kg_m^{\ 2}R_D^{\ 2}}{WLC_{ox}}\frac{1}{f} \tag{2-7}$$

From the total output noise voltage, the input noise voltage is calculated by dividing the output by the square of the differential gain, $g_m^2R_D^2$:

$$\overline{V_{n,in,tot}^{\ 2}} = 8kT(\frac{2}{3g_m} + \frac{1}{g_m^{\ 2}R_D}) + \frac{2K}{WLC_{ox}}\frac{1}{f} \tag{2-8}$$

Lastly, since the input noise current and the input noise voltage are correlated, the current is obtained by dividing the voltage by the square of the input impedance of the differential pair. Therefore, the input noise current is given by:

$$\overline{I_{n,in,tot}^{\ 2}} = \frac{8kT(\frac{2}{3g_m} + \frac{1}{g_m^{\ 2}R_D}) + \frac{2K}{WLC_{ox}}\frac{1}{f}}{|Z_{in}|^2} \tag{2-9}$$

where the input impedance, Z_{in}, is:

$$|Z_{in,diff}| = \frac{1}{wC_{gs}} \tag{2-10}$$

and ω is the radian frequency of the input and C_{gs} is the gate-source capacitance of $M_{1,2}$.

For many analog circuits, simply calculating the signal to noise ratio (SNR) from input noise generators is sufficient to analyze the noise performance of a system. However, in RF designs, such as the VGA, a circuit's noise performance is usually

characterized using noise figure (NF). Physically, the NF is a measure of how much the SNR ratio degrades from the input to the output of a system and it is defined as being the logarithm of the ratio of total output noise power to the noise power at the output due to the input source alone. It is usually expressed in terms of signal to noise ratios (SNR) as [6]:

$$NF = 10 * \log \frac{SNR_{in}}{SNR_{out}} \tag{2-11}$$

For simulation purposes however, the NF is more conveniently calculated as [6]:

$$NF = 10 * \log \frac{\overline{V_{n,out}^{\ 2}}}{A^2} * \frac{1}{4kTR_s} \tag{2-12}$$

where A represents the gain of the circuit, and R_s is the source impedance. It is important to observe from (2-12) that noise figure is a function of the source impedance and it changes if the source impedance changes. Standard port impedances for RF devices and coaxial cables are 50Ω or 75Ω. 50Ω is best for maximum power handling capability of coax, while 75Ω yields minimum attenuation. Minimum noise figure corresponds to highest receiver sensitivity, which is a performance target for all RF circuits. One of the reasons why noise figure is the standard performance metric of RF systems is because of the computational convenience which it provides for multi-stage designs. In such systems, the NF for the entire system can be computed from the NF of each individual stage. Figure 2.5 shows a system which consists of a cascade of 2 noisy stages.

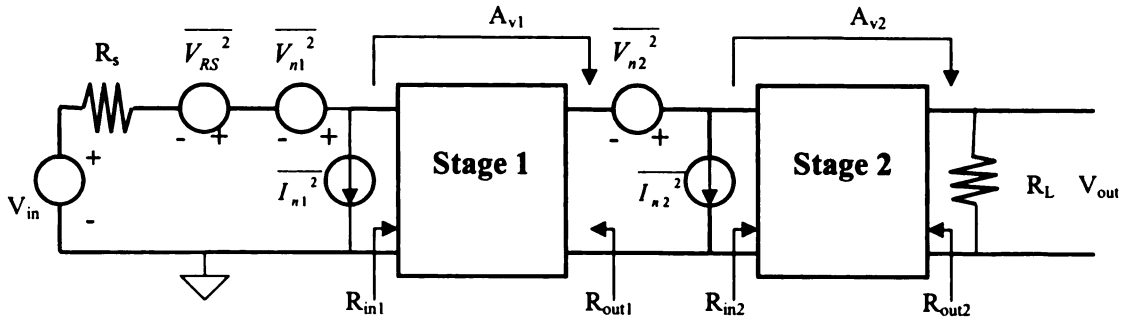


Figure 2.5: Cascade of two noisy stages.

For this system, the overall NF for the two stages is calculated as:

$$NF_{tot} = NF_{1,Rs} + \frac{NF_{2,Rout1} - 1}{A_p} \tag{2-13}$$

where $NF_{1,Rs}$ is the noise figure of stage 1 with respect to a source impedance R_s and A_p is the available power gain. More generally, for m stages, the total NF is given by the Friis equation as:

$$NF_{tot} = 1 + (NF_1 - 1) + \frac{NF_2 - 1}{A_{p1}} + \ldots + \frac{NF_m - 1}{A_{p1} \ldots A_{p(m-1)}} \tag{2-14}$$

This result is very useful because the individual building blocks of a tuner, such as the VGA, Mixer, PLL and Low Pass Filters are usually designed separately. In addition, the Friis equation also suggests that the noise performance of a multi-stage system is dominated by the first stage because the noise contributed by subsequent stages decreases as the gain increases. For the satellite tuner, this fact means that the VGAs noise figure dominates the overall NF.

2.3 Linearity

Linearity is the second concept which is important to the performance and characterization of a VGA. By definition, a system is said to be linear if and only if its outputs can be expressed as a linear combination of individual inputs [6]. For example, for inputs $x_1(t)$, $x_2(t)$ and all constants a and b to a system, the following conditions must be satisfied in order for the system to be linear:

$$x_1(t) \rightarrow y_1(t) \qquad x_2(t) \rightarrow y_2(t) \tag{2-15}$$

$$ax_1(t) + bx_2(t) \rightarrow ay_1(t) + by_2(t) \tag{2-16}$$

If these conditions are not satisfied, then the system is non-linear. Non-linearities can occur in a variety of ways and they are usually difficult to analyze, so therefore most designs are based on linear systems whenever possible. However, in general most systems are not unconditionally linear. In fact, almost all real systems become non-linear when large enough signals are applied to them. Yet, this troublesome fact is overcome by realizing that most systems exhibit linear behavior for small signals. Thus, if the amplitude of the input signal is limited, a non-linear system can be approximated by a linear system.

The input-output characteristics of MOS transistors are an example of the non-linear device described above and are especially important in characterizing the dynamics of the VGA. A MOSFET is linear for an input range which is determined by the transistor transconductance (g_m) of the particular transistor. It is a function of the drain current (I_D) and is given by [9]:

$$g_m = \sqrt{2I_D KP \frac{W}{L}} \tag{2-17}$$

For a MOS differential pair, the overall differential input-output characteristics are given by the following equation [9]:

$$V_{od} = -R_D \frac{W}{L} \frac{KP}{2} V_{id} \sqrt{\frac{4I_{ss}}{KP\frac{W}{L}} - V_{id}^{\,2}} \tag{2-18}$$

where KP is the process constant, I_{ss} is the tail bias current, V_{id} is the differential input voltage, and W and L are the width and length respectively of the transistors. The bias current I_{ss} is equal to twice the drain current for each of the individual transistors. Figure 2.6 illustrates the transfer function for one value of bias current, I_{ss}.

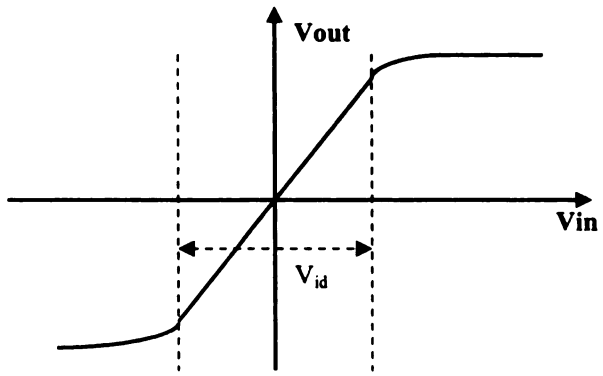


Figure 2.6: Input-Output Characteristics of a Differential Pair.

As Figure 2.6 shows, there is a differential input range, V_{id}, for which the transfer function is approximately linear and it is given by:

$$|V_{id}| \leq \sqrt{\frac{2I_{ss}}{KP\frac{W}{L}}} \tag{2-19}$$

An important observation that can be made from (2-19) is that the area of the transistors has a negative impact on linearity, as opposed to the positive influence which it has on noise.

A common technique which is utilized to improve the linearity of a differential pair is to reduce the dependence of the gain of the circuit upon the input level by making the gain fairly independent of the transistor bias currents (I_{ss}) [8]. This is usually accomplished via source degeneration with linear resistors, and the circuit can take on one of the following two forms shown in figure 2.7.

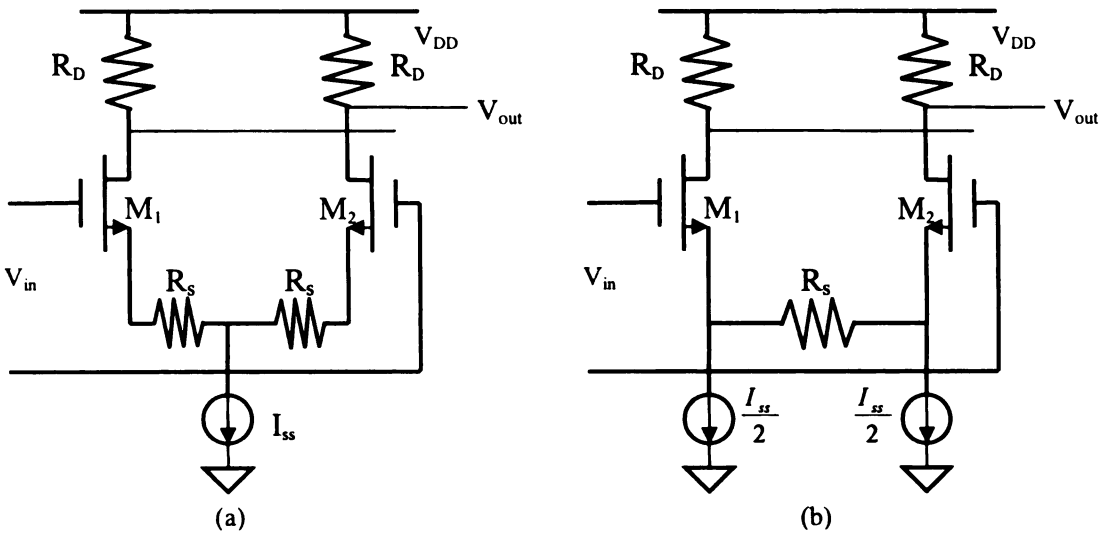


Figure 2.7: 2 Forms of Source Degeneration for a Differential Pair.

The source degeneration of figure 2.7 has the effect of reducing the signal swing which is applied between the gate and source of the transistors. It thereby decreases the effective transconductance for the pair, which is given by:

$$G_m = \frac{g_m}{1 + g_m R_s} \tag{2-20}$$

and the new differential gain of the circuit is then calculated as:

$$A_v = \frac{V_{out}}{V_{in}} = -\frac{g_m}{1 + g_m R_D} R_D \tag{2-21}$$

As the product g_mR_s grows large, the overall transconductance approaches $1/R_s$ which becomes more linear as intended. However, there are drawbacks to employing such a technique. One of the major drawbacks is that while linearity is improved, the gain is decreased and noise figure is increased unintentionally. Also, a second drawback is that the addition of the degeneration resistors decreases the available headroom for the circuit. If careful consideration is not given to the size of the degeneration resistors, it can cause the transistors to fall out of saturation, thus altering performance. For designs which use CMOS processes with minimum features sizes above 0.35μm and with corresponding power supplies above 3.3 V, this is not necessarily a critical concern. However, since power supply voltages decrease with feature size, this can be a very serious issue for designs using a 0.18μm process and below. The topology of figure 2.7b eliminates the headroom concern which the topology of figure 2.7a poses. However, there is still a limit to the maximum size of the degeneration resistor which can be used. This is because if too large of a resistor is used, the assumption of symmetry fails and it creates an open circuit between the sources of the two transistors, which causes the circuit to cease operating as a differential amplifier.

In most designs it is possible to simply characterize a circuit by its linear input-output response, however for RF amplifier circuits such as the VGA, the intrinsic non-linearities can not be entirely overlooked. This is because non-linearities often cause adverse effects which can impact a circuits performance. Consider the system which is modeled by the following expression:

$$y(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \tag{2-22}$$

where α_1, α_2, and α_3 are gain constants. If an input x(t) = A cos ωt is applied to the system, then the output, y(t), simplifies to:

$$y(t) = \frac{\alpha_2 A^2}{2} + (\alpha_1 A + \frac{3\alpha_3 A^3}{4})\cos\omega t + \frac{\alpha_2 A^2}{2}\cos 2\omega t + \frac{\alpha_3 A^3}{4}\cos 3\omega t \qquad (2\text{-}23)$$

where the input frequency present at the output is called the fundamental tone and the higher order frequencies are called harmonics. A couple of observations can also be made from (2-23). First, for ideal fully differential circuits which are odd symmetric, even order harmonics resulting from α_j with even j are eliminated. Second, the amplitude of the n^{th} harmonic grows approximately in proportion to A^n. Therefore, for small values of A, the system yields only a scaled version of the fundamental tone. However, as the magnitude of the input, A, increases, the higher order harmonics will dominate the output.

It is therefore the harmonics which cause a non-linear systems adverse effects. The two most important effects with respect to RF design are gain compression and intermodulation. Gain compression results from the fact that the output and gain vary as signal amplitude increases. In most circuits, this variation is a saturating function of the input, because the gain approaches zero for sufficiently high input signal levels. The 1-dB compression point quantifies this effect and it is defined as the input level at which the output small signal gain drops by 1dB. Figure 2.8 illustrates this phenomenon.

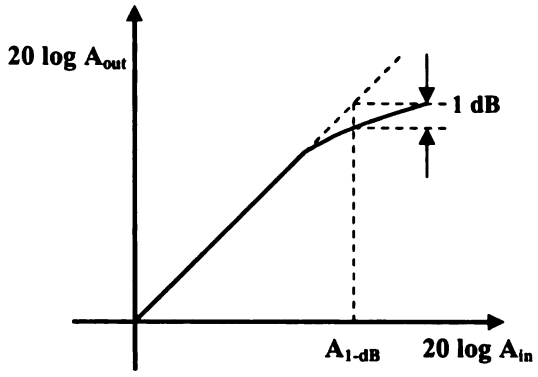


Figure 2.8: Illustration of 1-dB Compression Point.

The 1-dB compression point can be calculated from the system gain constants (α_1 and α_3) by the following expression:

$$A_{1dB} = \sqrt{0.145\left|\frac{\alpha_1}{\alpha_3}\right|} \tag{2-24}$$

The second significant non-linear effect is called intermodulation (IM) distortion and it is the result of when two signals with a small difference in frequency are applied to a non-linear system. In this special case, the frequency spectrum of the output will include components which are not harmonics of the input frequency. In order to see this concept more clearly, consider the non-linear system of (2-22). However, assume that in this instance that a two-tone input x(t) = Acos $\omega_1 t$ + Acos $\omega_2 t$, is applied to the system. The output frequency spectrum will contain components at the following frequencies: $\omega_1, \omega_2, \omega_1 \pm \omega_2, 2\omega_1 \pm \omega_2$, and $2\omega_2 \pm \omega_1$. The third-order IM products at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ have an amplitude of $3\alpha_3 A^3/4$. These are the terms which are particularly troublesome because they appear in close proximity to the fundamental tones. The other

higher order harmonics are usually eliminated by band-pass or low-pass filters and therefore do not introduce any signal distortion.

This type of IM distortion usually occurs in RF systems when a weak signal is accompanied by two strong interferers and they experiences third-order non-linearity. When this happens, one of the IM products can fall in the desired channel and consequently distort the wanted signal. Figure 2.9 illustrates this situation.

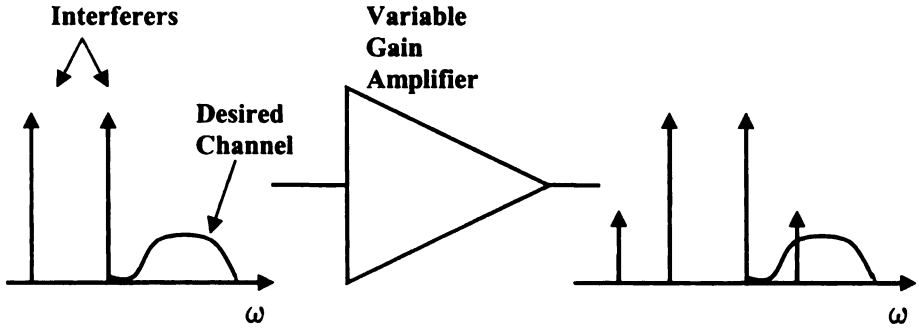


Figure 2.9: IM distortion of a signal by two strong interferers.

The third-order intercept point (IP_3) is the parameter which characterizes the distortion and it is usually measured with a two-tone test in which two tones close in frequency with a sufficiently low A are applied to a system. Initially, A is chosen small enough so that the gain of the system is relatively constant and the higher order terms are negligible. Then, as the value of A is increased, a point is eventually reached where the magnitude of the fundamental terms (αA) equals the magnitude of the IM products ($3\alpha_3 A^3/4$). This point is called the third-intercept point. Figure 2.10 graphs on a log scale the fundamental tones and IM products for this type of test.

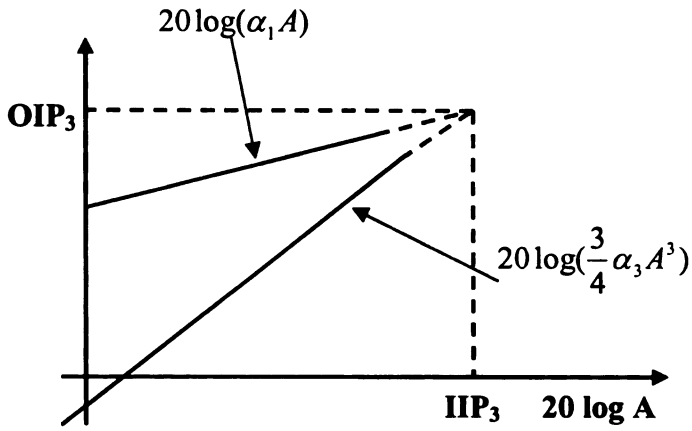


Figure 2.10: Outputs of a two-tone intermodulation test.

The x-axis coordinate of the intercept is called the input-referred third-order intercept point (IIP_3), and the y-axis coordinate is called the output-referred third-order intercept point (OIP_3). IIP_3 is how the linearity of RF systems is usually characterized and it also serves as the means by which the linearity of different circuits can be compared. In many cases the actual point is beyond the allowable input range of the circuit, so therefore, it can be obtained mathematically using linear extrapolation on a logarithmic scale. From the mathematical perspective, the IIP_3 can be calculated from the system gain factors, α_1 and α_3, as follows:

$$A_{IP3} = \sqrt{\frac{4}{3}\left|\frac{\alpha_1}{\alpha_3}\right|} \tag{2-25}$$

However, traditionally signal levels are expressed in terms of powers (dBm), and IIP_3 is usually estimated experimentally without extrapolation using the following formula:

$$IIP3_{dBm} = \frac{\Delta P_{dB}}{2} + P_{in/dBm} \tag{2-26}$$

where ΔP_{dB} refers to the difference in power between the IM products and fundamental tones at the output, and P_{in} is the input signal power. Figure 2.11 illustrates this graphical technique for obtaining the IIP_3 of a circuit using (2-26).

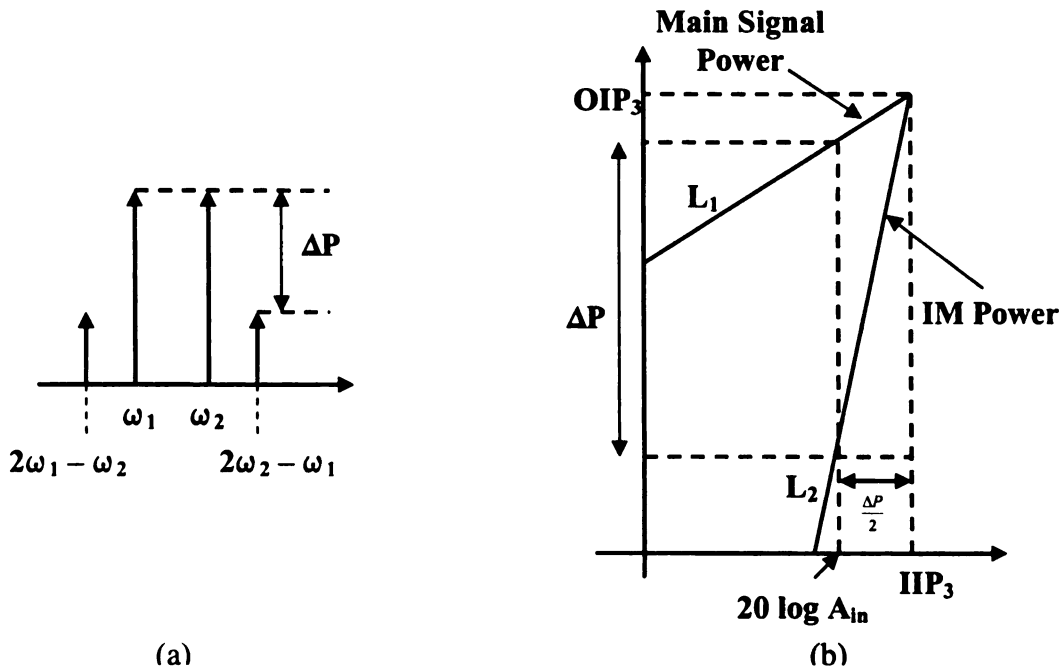


Figure 2.11: (a) Frequency components and (b) graphical interpretation of (2-26).

It is also important to note that the 1dB compression point and IIP_3 are related. The relation of the two phenomena is given by the following equation:

$$\frac{A_{1-dB}}{A_{IIP3}} = \frac{\sqrt{0.145}}{\sqrt{\frac{4}{3}}} \approx -9.6dB \qquad (2\text{-}27)$$

In addition to single stage linearity calculations, it is also important to understand the effects of non-linearities on multi-stage systems such as the satellite tuner. Figure 2.12 shows an example of a cascade of two non-linear stages.

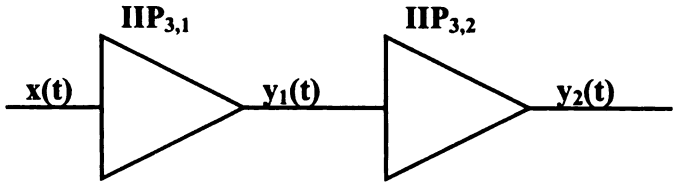


Figure 2.12: Cascade of 2 Non-Linear Stages.

For such a system, consider the case where the output of the first stage is given by (2-23) and the second stage by the following expression:

$$y_2(t) = \beta_1 y_1(t) + \beta_2 y_1^{\,2}(t) + \beta_3 y_1^{\,3}(t) \tag{2-28}$$

where $\beta 1$, $\beta 2$, and $\beta 3$ are the gain constants of the second stage. Then, the IIP_3 calculation for the two stage system follows as [6]:

$$\frac{1}{A_{IIP3}^{\,2}} = \frac{1}{A_{IIP3,1}^{\,2}} + \frac{3\alpha_2\beta_2}{2\beta_1} + \frac{\alpha_1^{\,2}}{A_{IIP3,2}^{\,2}} \tag{2-29}$$

where $A_{IIP3,1}$, and $A_{IIP3,2}$ represent the IIP_3 points for the first, and second stages respectively. This analysis can also be extended to include systems which consist of more than two stages. In such a generalization, the second term from (2-29) becomes negligible and the IIP3 of the system can be calculated by this equation [6]:

$$\frac{1}{A_{IIP3}^{\,2}} = \frac{1}{A_{IIP3,1}^{\,2}} + \frac{\alpha_1^{\,2}}{A_{IIP3,2}^{\,2}} + \frac{\alpha_1^{\,2}\beta_1^{\,2}}{A_{IIP3,3}^{\,2}} + \ldots \tag{2-30}$$

An important observation which results from (2-28) is that when the gain of the first stage (α_1) increases, the overall IIP_3 decreases, meaning that the system has a overall smaller linear input range. Therefore it is concluded from this observation that in order to keep approximately the same overall linear input range as the first stage, the latter stages

must become more linear and have a higher IIP_3 than the first stage. In other words, the linear response of a multi-stage system is dominated by the latter stages. More specifically for this thesis, the IIP_3 of the subsequent stages in the tuner such as the mixer are more critical than the VGAs IIP_3 in determining the overall linearity of the tuner.

2.4 Microwave Circuit Theory

A third and final theoretical topic which is important to the RF VGA is matching and input return loss. In general, most circuit analysis is done using Kirchoff's voltage and current laws (KVL and KCL). However, these concepts only represent a restricted version of the more general concepts covered by Maxwell's Equations. KVL and KCL are based upon lumped circuit elements and are adequate to analyze circuits which operate at low enough frequencies so that the dimensions of the circuit elements are much smaller than a wavelength [10]. On the other hand, when frequencies become high and the wavelengths become comparable to the size of the circuit element, lumped circuit theory is no longer valid and transmission line theory takes over. This thesis presents a design which operates in the gigahertz range; hence, transmission line theory plays a crucial role in the design characterization.

Thanks to the small dimensions involved with modern integrated circuit (IC) processes, most on-chip transmission lines can be virtually ignored and analysis can be limited to traditional circuit theory. However, the inputs and outputs between various circuit blocks such as the LNB, VGA and Mixer can require special attention as transmission lines. The high frequency signals which travel on the transmission lines are often described in terms of powers and characterized by a wave-like behavior which is

highly dependant on the values of the impedances of the circuit. It is highly desirable that as much of the signal power as possible is transmitted from the source to the load, with as little power as possible lost on the transmission line itself. Transmission line theory dictates that this condition only results when the source and load impedances are equal to the characteristic impedance of the transmission line. Otherwise, any mismatch will result in reflections and return loss along the line. However, impedance matching is not a trivial task, especially given the fact that wideband RF circuits often incorporate a combination of resistive, capacitive and inductive components. Therefore, it is important to quantize impedance matches and this is usually done by means of the reflection coefficient (Γ), given by [7]:

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2-31}$$

where Z_L is the load impedance and Z_0 is the characteristic impedance of the transmission line. The reflection coefficient is zero when Z_L is equal to Z_0, or in other words when there is a perfect impedance match. In all other cases, the reflection coefficient is non-zero, resulting in reflections along the transmission line. In addition, since the impedances can have a combination of resistive, capacitive or inductive components, the reflection coefficient will usually consist of a real and imaginary component. Although it may seem straightforward to plot this quantity on a rectangular coordinate system, it is often more convenient to plot it using the Smith Chart which is shown in Figure 2.13.

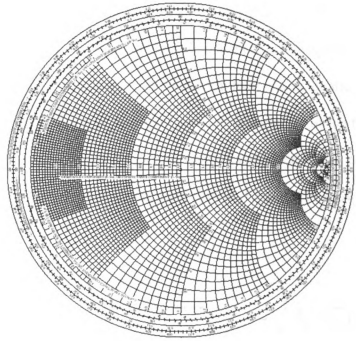


Figure 2.13: The Smith Chart.

The center of the Smith Chart corresponds to a zero reflection coefficient, or a perfect impedance match. While the bottom half and top half of the circle corresponds to capacitive and inductive impedances, respectively.

The Smith Chart is a particularly useful design tool in that it provides insight into the nature of input and output impedances. That insight then allows designers to make modifications if necessary in order to achieve better matching. However, a tool that is used more frequently in order to characterize the small-signal microwave behavior of a system is the scattering parameters, often referred to as the S-parameters. S-parameters are the RF equivalent to the two port parameters (Y-parameters and Z-parameters), which are commonly used to analyze the input output behavior of low frequency circuits. The reason that S-parameters replace the lower frequency two-port parameters is because the Y- and Z-parameters require open and short circuits of the ports in order to determine

their values experimentally. At frequencies above one gigahertz, adequate open and short circuits are difficult to obtain. Therefore, S-parameters solve this problem by exploiting the fact that transmission lines that are terminated in their characteristic impedances give rise to no reflections. In doing so, the input and output variables of the two port system are defined in terms of incident and reflected waves as opposed to the port voltages and currents of the lower frequency two port descriptions. Figure 2.14 shows the S-parameter 2 port description.

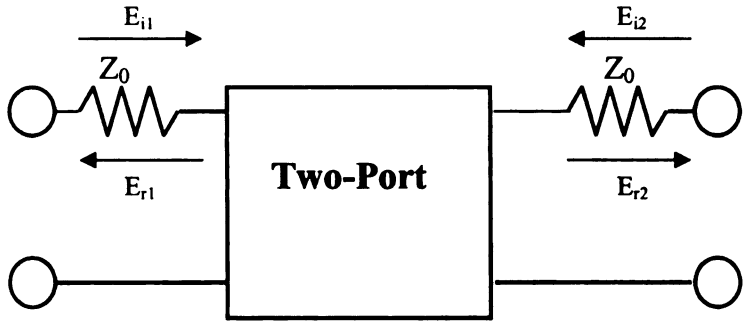


Figure 2.14: S-Parameter 2 Port Description.

From the input and output definitions of Figure 2.14, the two port equations are given by the following [7]:

$$b_1 = s_{11}a_1 + s_{12}a_2 \tag{2-32}$$

$$b_2 = s_{21}a_1 + s_{22}a_2 \tag{2-33}$$

where $a_1 = E_{i1}/\sqrt{Z_0}$, $a_2 = E_{i2}/\sqrt{Z_0}$, $b_1 = E_{r1}/\sqrt{Z_0}$, and $b_2 = E_{r2}/\sqrt{Z_0}$. By terminating the output port in Z_0 and driving the input port, the S_{11} and S_{21} parameters can be calculated as:

$$S_{11} = \frac{b_1}{a_1} = \frac{E_{r1}}{E_{i1}} \tag{2-34}$$

$$S_{21} = \frac{b_2}{a_1} = \frac{E_{r2}}{E_{i1}} \tag{2-35}$$

The parameter S_{11} is termed the input return loss (reflection) coefficient and s_{21} represents a gain coefficient. Also, the remaining coefficients S_{22} and S_{12} can be calculated in the reverse method, by terminating the input port in Z_0 and driving the output port:

$$S_{22} = \frac{b_2}{a_2} = \frac{E_{r2}}{E_{i2}} \tag{2-36}$$

$$S_{12} = \frac{b_1}{a_2} = \frac{E_{r1}}{E_{i2}} \tag{2-37}$$

where S_{22} is the output reflection coefficient, and S_{12} refers to the reverse transmission coefficient. S_{12} determines the amount of isolation from output to input. As a whole, these four parameters then completely characterize the microwave transmission line behavior of a system. It is significant to note that laboratory test equipment used for high frequency (1GHz or more for example) typically provides S-parameters as the fundamental measured quantity.

2.5 Conclusion

The RF VGA has a high degree of complexity and it requires an understanding of several areas of study in order to completely characterize it. Among the more crucial topics which need to be considered are: Noise; Linearity and Microwave Transmission Lines. Knowledge of noise is essential because a systems response to noise details the minimum signal level which can be processed while maintaining acceptable quality. Likewise, a familiarity with the topic of linearity is important because all real systems have a threshold for linearity. The linearity threshold combined with the noise figure

determines the overall dynamic range of a receiver. Therefore, it is crucial to know how the threshold is defined and measured. And, finally, an understanding of microwave transmission line theory is vital to ensure that proper input and output matching is done for maximum transfer of input signal power with minimum loss.

Chapter 3: VGA Design

3.1 Introduction

There are many potential circuit topologies for variable gain amplifiers (VGAs). However, many of the VGA implementations published in recent years lack the performance required of a satellite tuner. This chapter begins by discussing a previously published VGA for satellite applications along with its shortcomings. Then, the VGA which was designed for this thesis is presented. The complete design consists of the following three sub-circuits: a constant g_m bias circuit; a constant gain stage; and five variable gain stages. The design, purpose and functionality of each sub-circuit are discussed in detail and then the final version is presented as a whole.

3.2 Previous Work

One of the more relevant and most recent broadband VGA designs to be published is discussed in [4]. The paper presents a complete broadband satellite tuner integrated circuit fabricated in a 0.18μm CMOS technology. A simplified version of the RF VGA portion of the design is shown in Figure 3.1.

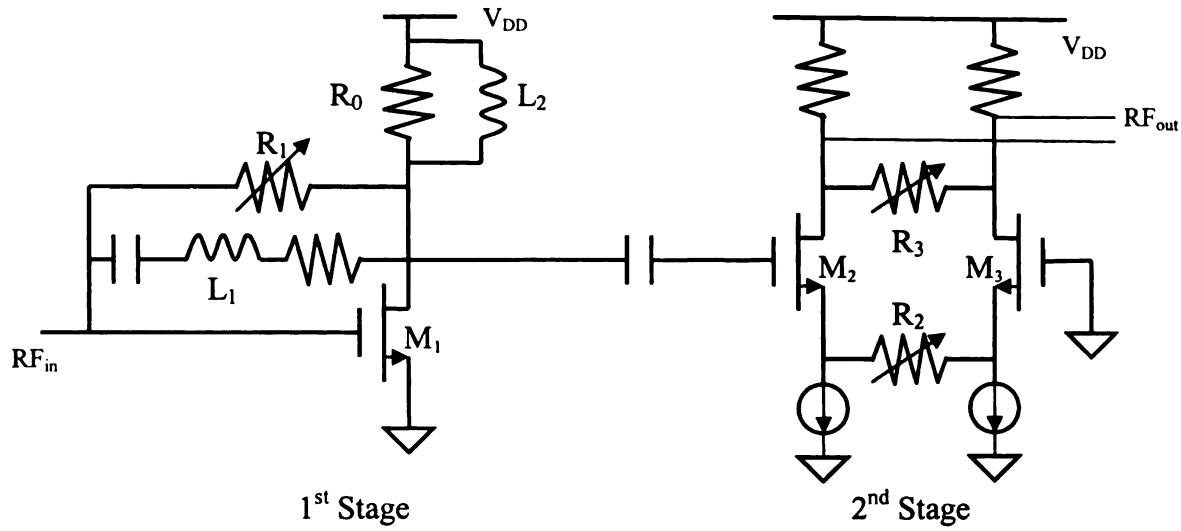


Figure 3.1: Simplified Schematic of VGA from [4].

The VGA is composed of two stages. The first stage consists of M_1, R_0, L_2 and the conventional parallel RLC feedback containing L_1. The resistive and capacitive components of the RLC feedback path are parasitic quantities. The primary function of this stage is to provide a wideband input impedance matching network to minimize reflections. In addition, it improves the overall linearity at low gain and provides moderate gain control via the variable resistor R_1. The output of this stage is coupled to the second stage which consists of M_2, M_3 and the two variable resistors R_2 and R_3. Most of the gain control is done in this part of the circuit using R_2 and R_3. The authors assert that the RF front end consisting of the VGA and Mixer is capable of producing the following results: a 33dB dynamic range; a 6.5 – 10 dB NF; and a +9 dBm IIP_3 from a -20 dBm input.

While the results presented in [4] are desirable, the design does have some drawbacks which limit its effectiveness. One of the biggest problems with this topology is the difficulty which exists in realizing the variable resistances. While the authors don't

explicitly say how the variable resistances are realized, they are most likely implemented using MOSFETs operated in their ohmic region. The resistance is then varied by changing a control voltage which is applied to the gate of the transistor. In order to provide the approximate 30 dB of dynamic range which is required, each resistor would have to be capable of attaining a 10x variation, which is a very difficult task to accomplish. A second drawback to this design is that it is composed of a cascade of two stages, with the second stage providing most of the gain control. According to the Friis equation which is discussed in Chapter 2, this will result in a higher overall system noise figure. Therefore, in order to minimize the noise figure, the first stage should provide most of the gain control. Finally, a third difficulty which arises deals with the parallel RLC feedback. The RLC feedback relies on parasitic elements which are difficult to predict. Also, implementing high quality monolithic inductors is rather challenging. Therefore, the remaining sections of this chapter present a VGA architecture with improved performance implemented in a 0.35 μm CMOS process.

3.3 Constant g_m Bias Circuit

The VGA designed for this thesis is composed of three sub-circuits, and each will be analyzed individually. The first sub-circuit consists of the constant g_m bias circuit. As the previous chapter detailed, the biasing current (I_{ss}) of a differential amplifier is one of the most important parameters in controlling the performance of the VGA. This is because the biasing current directly controls the transconductance (g_m) of the amplifying transistors of the VGA, which in turn plays a crucial role in determining the gain, noise figure, linearity, and input impedance of the circuit. Therefore, since performance is so

highly dependant on this current, it is essential that it is stabilized and independent of variations in power supply voltage, process and temperature. One of the ways in which this can be accomplished is by using a current biasing network such as the one presented in [11] in which the transistor transconductances are matched to the conductance of a resistor. Figure 3.2 shows the schematic of this circuit.

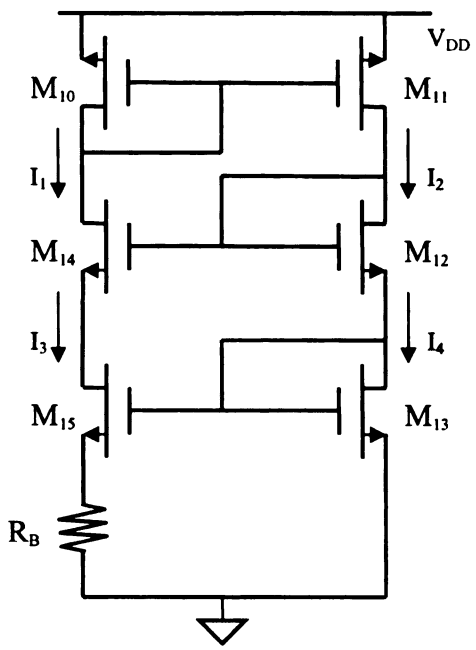


Figure 3.2: Constant g_m bias circuit.

Transistors M_{10} and M_{11} of the bias circuit are designed to be the same size. Then, the current mirror formed by M_{10} and M_{11} forces the currents I_1 and I_2 to be equal. Similarly, currents I_3 and I_4 are also equal. So, if kirchhoff's voltage law is applied to the loop containing M_{13}, M_{15} and R_B, the following expression results:

$$V_{GS13} = V_{GS15} + I_{D15}R_B \tag{3-1}$$

Then, subtracting the threshold voltage (V_{TH}) from both sides, and using the fact that the currents of M_{13} and M_{15} are equal, results in the re-written equation:

$$\sqrt{\frac{2I_{D13}}{KP(\frac{W}{L})_{13}}} = \sqrt{\frac{2I_{D13}}{KP(\frac{W}{L})_{15}}} + I_{D13}R_B \tag{3-2}$$

By rearranging the terms of (3-2) and recalling that $g_{m(M13)} = \sqrt{2KP(\frac{W}{L})I_{M13}}$, the transconductance of M_{13} is related to the resistance R_B by the following:

$$g_{m(M13)} = \frac{2\left[1 - \sqrt{\frac{(\frac{W}{L})_{13}}{(\frac{W}{L})_{15}}}\right]}{R_B} \tag{3-3}$$

Therefore, the transconductance of M_{13} is independent of power-supply voltage as well as process and temperature variations, and it is simply determined by the size ratios of M_{13}, M_{15} and the value of the resistance R_B. It is important to note that all of the differential pair biasing currents of the VGA are just scalar multiples of the current generated by the biasing network. Therefore, the currents of the differential pairs are also independent of variations in power supply voltage, process and temperature which is the desired effect. However, the preceding derivation is ideal and it ignores many second order effects such as the body effect, transistor output impedance, and the process and temperature dependency of on-chip resistors. The neglected second order effects explain why small variations of the biasing currents exist in reality. However, the variation is minimized by this circuit.

3.4 Constant Gain Stage

The second sub-circuit of the VGA is the constant gain (attenuation) stage. The function of this circuit is to generate the lower gain (attenuation) limit which is needed

for the tuner to process large valued RF signals. As Chapter 2 indicated, large input signals cause real systems to exhibit non-linear behavior. Therefore, since this stage processes the largest input signals, it must be the most linear and in turn define the upper limit of IIP_3 for the system. In addition, the circuit must provide uniform gain over the entire frequency range.

The second requirement is troublesome especially at high gains, because the performance of analog circuits changes at high frequencies due to the effect of device and load capacitances. However it still requires significant attention at lower gains. One of the most important high frequency effects is known as the "Miller Effect" which is derived from Millers Theorem.

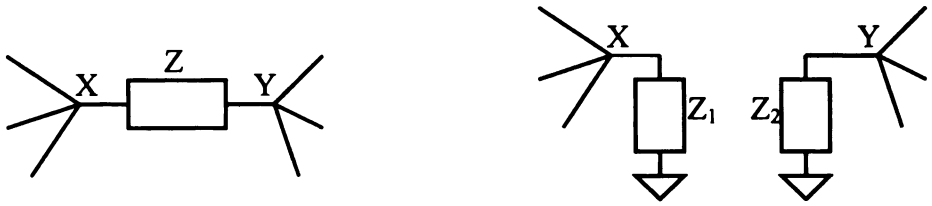


Figure 3.3: Illustration of Miller's Theorem.

Millers Theorem states that if the circuit of Figure 3.3(a) can be converted to that of Figure 3.3(b), then the resulting impedances Z_1 and Z_2 are given by the following expressions:

$$Z_1 = \frac{Z}{(1 - A_V)} \tag{3-4}$$

$$Z_2 = \frac{Z}{(1 - A_V^{-1})} \tag{3-5}$$

where A_V is the voltage gain from node X to Y. While Miller's Theorem can be applied to any circuit in which any finite impedance appears in parallel with the main signal path, the Miller Effect is specific to the case when a capacitance is connected across two nodes (X and Y) which have an inverting voltage gain between them. Figure 3.4 illustrates this type of a situation.

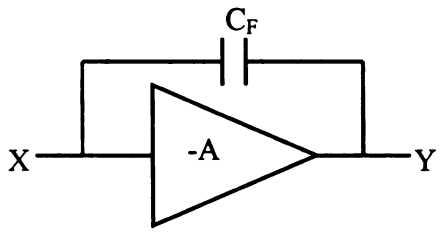


Figure 3.4: The Miller Effect.

The capacitance C_F can be viewed as two capacitances (C_1 and C_2) at the input and output nodes respectively by Millers Theorem. Figure 3.5 shows the equivalent circuit.

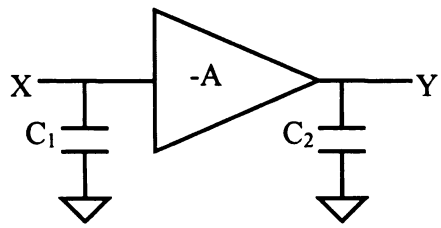


Figure 3.5: Equivalent Circuit of Figure 3.4.

The value of C_1 and C_2 are given by $(1+A)*C_F$ and $(1+A_v^{-1})*C_F$ respectively. The scale factors $(1+A)$ and $(1+A_v^{-1})$ are frequently termed the Miller multiplication factors. The transfer function (V_o/V_{in}) for the circuit exhibits two poles, each of which is determined by the total capacitance seen from each node to ground multiplied by the total resistance seen at each node to ground. The Miller multiplication of C_F moves the poles closer to

the origin and results in a reduction of gain and limits the useable bandwidth. Since Miller multiplication is gain dependant, it generally has its greatest consequence at high gains.

In order to relate this to a common source amplifier, which is the goal of this discussion, it is necessary to analyze the high frequency behavior for the circuit. Figure 3.6 (a) shows a single ended common source amplifier and (b) its high frequency small signal model.

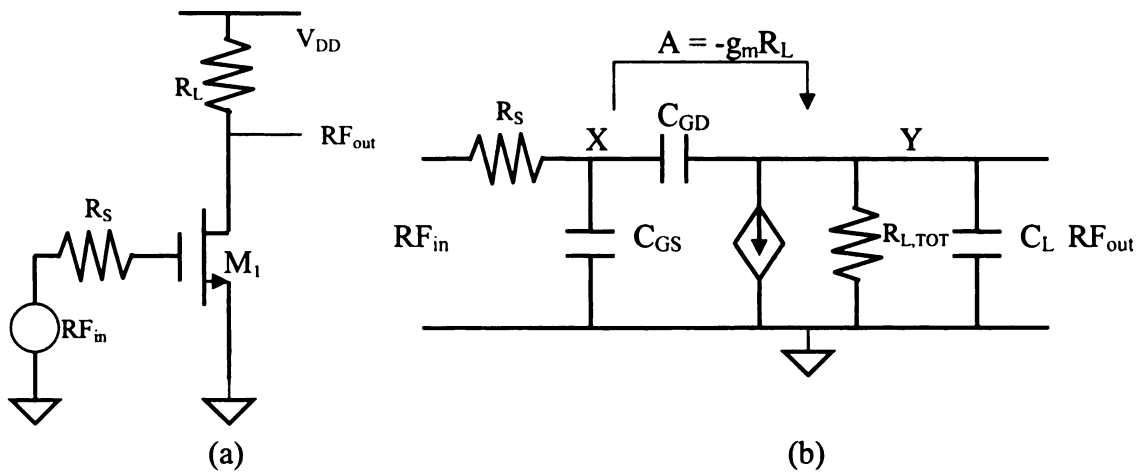


Figure 3.6: (a) Common Source Amplifier (b) High Frequency Small Signal Model.

Figure 3.6 demonstrates the single-ended version for simplicity because the analysis is similar to that of the differential pair. $R_{L,TOT}$ is the parallel combination of R_L and the output resistance (R_{DS}) of M_1. C_L is the equivalent load capacitance and it is the combination of the drain to bulk capacitance (C_{DB}) of M_1 and any capacitive loading seen at the output of M_1. C_{GD} is the gate to drain capacitance of M_1 and it provides a parallel path from the input to the output. Therefore, since common source amplifiers are

characterized by inverting gains this structure is susceptible to the Miller Effect. The transfer function for Figure 3.6 can be roughly estimated by associating one pole with each node. The total capacitance from node X to ground consists of C_{GS} and the Miller multiplication of C_{GD} and is given by the following:

$$C_{in,tot} = C_{GS} + (1 - A_V)C_{GD} \tag{3-6}$$

where $A_v = -g_m R_D$. The total capacitance from node Y to ground is equal to the combination of $C_{L,TOT}$ and the Miller multiplication of C_{GD} and is given by:

$$C_{out,tot} = C_L + (1 - A_V^{-1})C_{GD} \approx C_L + C_{GD} \tag{3-7}$$

Thus, the approximate input and output poles are:

$$\omega_{in} = \frac{1}{R_S[C_{GS} + (1 + g_m R_{L,TOT})C_{GD}]} \tag{3-8}$$

and

$$\omega_{out} = \frac{1}{R_{L,TOT}(C_L + C_{GD})} \tag{3-9}$$

And, the resulting approximate transfer function neglecting the presence of any zeroes is then:

$$\frac{V_{out}}{V_{in}}(s) = \frac{-g_m R_{L,TOT}}{(1 + \frac{s}{\omega_{in}})(1 + \frac{s}{\omega_{out}})} \tag{3-10}$$

An exact transfer function can be computed using KCL at the input and output nodes to yield:

$$\frac{V_{out}}{V_{in}}(s) = \frac{(C_{GD}s - g_m)R_{L,TOT}}{R_S R_{L,TOT}\xi s^2 + [R_S(1 + g_m R_{L,TOT})C_{GD} + R_S C_{GS} + R_{L,TOT}(C_{GD} + C_L)]s + 1} \tag{3-11}$$

The poles which are predicted by the Miller multiplication of C_{GD} approximation technique in equations (3-8) and (3-9) are relatively accurate approximations. However, the output pole which is predicted in (3-9) is only valid if C_{GS} dominates the transfer function. In addition, the zero which is located at $+g_m/C_{GD}$ is not predicted by the Miller multiplication technique. It is instead the result of direct coupling of the input to the output through C_{GD}.

The 3 dB bandwidth can be explicitly calculated from the transfer function generated above; however, this approach is rather difficult. Therefore, the preferred method for estimating the 3 dB bandwidth is to use open circuit time-constants (OCτs). One of the benefits to this technique is that it identifies the elements which are most responsible for bandwidth limitations. The recipe for computing the 3 dB bandwidth from the OCτs is as follows:

- Compute the effective resistance (R_{jo}) facing each jth capacitor with all of the other capacitors removed (open circuited).
- Form the product $\tau_{jo} = R_{jo} C_j$.
- Sum all τ_{jo}.
- Invert the sum.

Since the common source amplifier of Figure 3.6 includes three capacitances, it will have three open circuit time constants. The time constant corresponding to C_{GS} is given by the following:

$$\tau_{GS} = C_{GS} R_{GS} = C_{GS} R_S \tag{3-12}$$

Similarly, the time constants for C_{GD} and C_L follow respectively as:

$$\tau_{GD} = C_{GD} R_{GD} = C_{GD}(R_S + (g_m R_S + 1) R_{L,TOT}) \tag{3-13}$$

and

$$\tau_L = C_L R_{L,TOT} \tag{3-14}$$

And the 3 dB bandwidth is equal to:

$$\omega_{3dB} = \frac{1}{\tau_{GS} + \tau_{GD} + \tau_L} = \frac{1}{C_{GS}R_S + C_{GD}(R_S + (g_m R_S + 1)R_{L,TOT}) + C_L R_{L,TOT}} \tag{3-15}$$

As (3-13) indicates, the time constant corresponding to the gate to drain capacitance is the largest and dominates the denominator of the bandwidth expression. Therefore it is most responsible for the bandwidth limitations of the circuit.

One solution to the Miller Effect problem brought upon by C_{GD} is to isolate or shield the capacitance so that it no longer appears in a direct parallel path from input to output. By eliminating the direct parallel path, it reduces the size of the pole by suppressing the Miller multiplication so that it no longer dominates the bandwidth expression and therefore the Miller Effect is less significant. A cascode configuration effectively performs this shielding. Figure 3.7 illustrates a single-ended cascode common source stage and its small signal equivalent circuit.

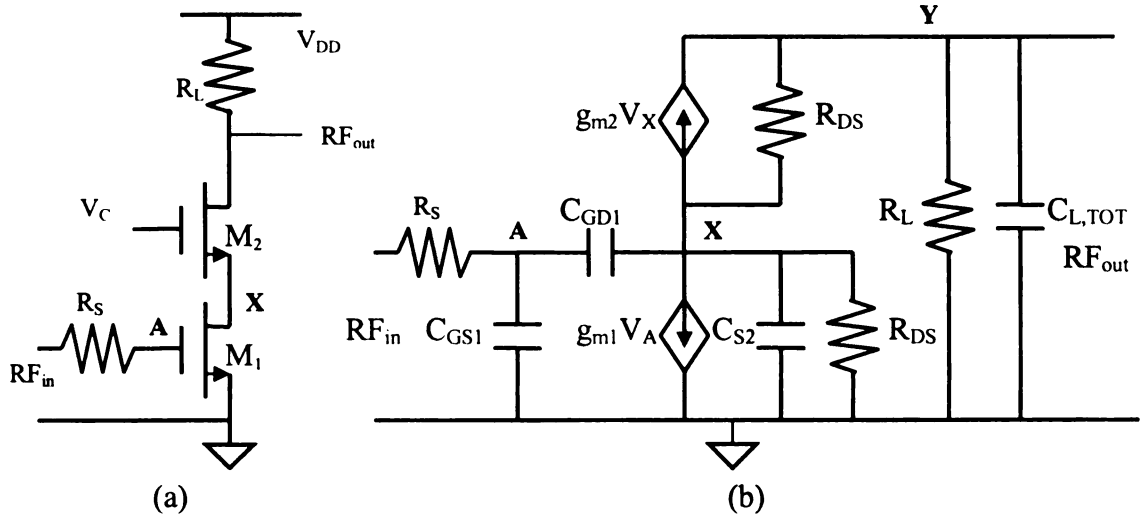


Figure 3.7: (a) Cascode Common Source Amplifier (b) High Frequency Model.

C_{S2} is the total capacitance seen at node X and it is given by the following expression:

$$C_{S2} = C_{DB1} + C_{SB2} + C_{GS2} \tag{3-16}$$

Similarly, $C_{L,TOT}$ is the total capacitance seen at the output node (Y) and it is equal to the following:

$$C_{L,TOT} = C_{GD2} + C_{DB2} + C_L \tag{3-17}$$

In this type of configuration, M_1 generates a small signal drain current which is proportional to RF_{in}, and then M_2 acts as a current buffer and simply routes the current to the load (R_L). Notice that the gate to drain capacitance of M_1 is no longer directly connected from input (A) to output (Y). The gain from node A to node X determines the Miller Effect of $C_{GD1}$and it is approximately equal to $-g_{m1}/(g_{m2}+g_{mb2})$. If M_1 and M_2 are approximately the same size, the gate to drain capacitance of M_1 (C_{GD1}) is multiplied by roughly 2 instead of the large voltage gain (A) of a simple common source stage. Then, the pole which corresponds to node A is approximated by:

$$\omega_{p,A} = \frac{1}{R_S[C_{GS1} + (1 + \frac{g_{m1}}{g_{m2} + g_{mb2}})C_{GD1}]} \tag{3-18}$$

Nodes X and Y also contribute poles to the input-output transfer function, and their values are respectively given as:

$$\omega_{P,X} = \frac{g_{m2} + g_{mb2}}{2C_{GD1} + C_{S2}} \tag{3-19}$$

and

$$\omega_{P,Y} = \frac{1}{R_D C_{L,TOT}} \tag{3-20}$$

Therefore, the overall transfer is approximately determined by the following:

$$\frac{V_{out}}{V_{in}}(s) = \frac{\frac{-g_{m1}}{(g_{m2}+g_{mb2})}}{(1+\frac{s}{\omega_{P,A}})(1+\frac{s}{\omega_{P,X}})(1+\frac{s}{\omega_{P,Y}})} \tag{3-21}$$

The exact input-output transfer function is rather difficult to derive and does not add any further insight into the frequency response of the cascode topology, so it is not included in this discussion.

Similar to the simple common source configuration, the approximate 3 dB bandwidth of the cascode topology is calculated using the open circuit time constant technique. There are four capacitors in the cascode configuration; so therefore, there are four open circuit time constants. The first time constant is a result of the gate to source capacitance of M_1 (C_{GS1}), and it is given by the following:

$$\tau_{C_{GS1}} = C_{GS1}R_S \tag{3-22}$$

The second time constant is generated by the gate to drain capacitance of M_1 (C_{GD1}), and is expressed as:

$$\tau_{C_{GD1}} = C_{GD1}\frac{R_{DS}}{2}(1+g_{m1}R_S) \tag{3-23}$$

Next, the equivalent capacitor at node X (C_{S2}) contributes the following time constant:

$$\tau_{C_{S2}} = C_{S2}\frac{R_{DS}}{2} \tag{3-24}$$

And finally, the 4th time constant is determined by the capacitance at output node Y ($C_{L,TOT}$) and it is given by:

$$\tau_{C_{L,TOT}} = C_{L,TOT}\frac{g_m R_{DS}^{\ 2}}{2} \tag{3-25}$$

Thus, the 3-dB bandwidth is the inverted sum of the time constants and is estimated to be:

$$\omega_{3dB} = \frac{1}{C_{GS1}R_S + C_{GD1}\frac{g_m R_{DS}^{\ 2}}{2} + C_{S2}\frac{R_{DS}}{2} + C_{L.TOT}\frac{g_m R_{DS}^{\ 2}}{2}} \tag{3-26}$$

Although the time constants from C_{GD1} and $C_{L,TOT}$ appear to be of the same form, the time constant from $C_{L,TOT}$ will dominate the denominator of the bandwidth expression, because $C_{L,TOT}$ is usually a much larger capacitance than C_{GD1}. Therefore, this topology is preferred over the simple common source topology because it effectively improves the bandwidth and gain by mitigating the Miller Effect. It is important to note that choosing the value of the bias voltage (V_c) on M_2 is not a critical design issue. It just needs to chosen high enough to ensure that M_2 remains in saturation and low enough to guarantee that M_1 also stays in saturation.

The complete differential design for the constant gain circuit including the cascode technique is shown in Figure 3.8.

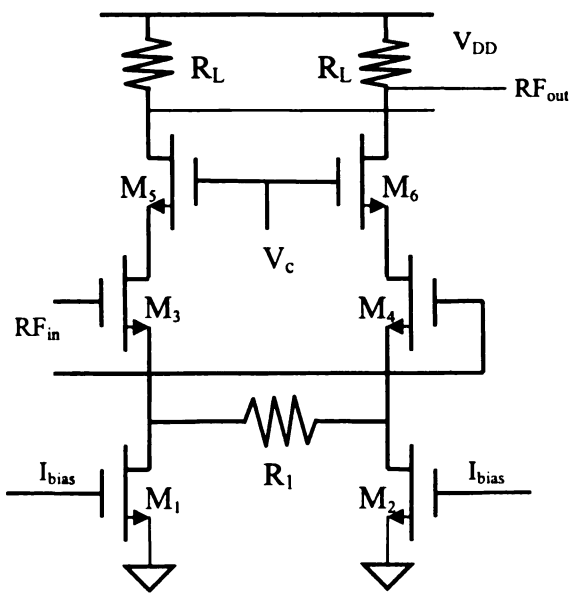


Figure 3.8: Constant Gain Circuit Architecture.

Transistors M_1 and M_2 are current sources whose current is a scalar multiple of the biasing network current of section 3.3. Since high linearity at low gain is so crucial to the VGA, the linearization technique discussed in Chapter 2 which utilizes a degeneration resistor (R_1) is used to maximize the IIP_3. As mentioned earlier, the constant gain stage is also implemented differentially due to the several advantages which differential structures possess over single-ended designs. Some of those advantages include: higher immunity to common mode "environmental noise"; increased maximum achievable voltage swings (larger gain); simpler biasing; and higher linearity. Although there are disadvantages to differential structures, such as: 2x the power consumption versus single-ended topology; increased area; and concerns over transistor matching in layout; the advantages far outweigh the disadvantages, which is why the differential structure is preferred. Transistors M_5 and M_6 are the cascode transistors which are used to resolve the Miller Effect.

3.5 Five Variable Gain Stages

The final piece of the VGA design consists of five variable gain stages in parallel. The primary reason why the five additional gain stages are included in the circuit is to achieve the approximate 30 dB of dynamic range without requiring the 10x variable gain resistors at RF frequencies. The dynamic range is accomplished by using the concept of analog switches to steer the current generated by each stage either towards the load or to another location such as the power supply. The staggering of the gains from the five stages implements a discrete, staircase approach to achieving the required dynamic range.

An added benefit to the multi-stage design is that it is able to adequately satisfy the high gain requirement of the tuner without using large transistor sizes or currents. For single stage designs, high gains are accomplished by using large transistor sizes or increased current biasing or both because the gain of a differential pair is proportional to these quantities. Large transistors are undesirable due to the intrinsic gate to source capacitances which are associated with them. In addition, there is a current limitation due to voltage headroom constraints. The six stage design mitigates this problem by using reasonable sized transistors, and current biasing levels.

This high gain of the multi-stage design is derived from the fact that transconductances which are placed in parallel with one another can be added to yield a larger equivalent transconductance and hence a larger gain. For example, the maximum gain of a six stage amplifier (such as the one used in this VGA) is given by the following:

$$A_v = -(g_{m1} + g_{m2} + g_{m3} + g_{m4} + g_{m5} + g_{m6})R_L \tag{3-27}$$

where $g_{m1,2,3,4,5,6}$ represent the individual transconductances of stages one through six. One drawback to this topology is that it has an increased area, and it wastes power when it is in the low gain states. Figure 3.9 shows a block diagram which illustrates this functionality.

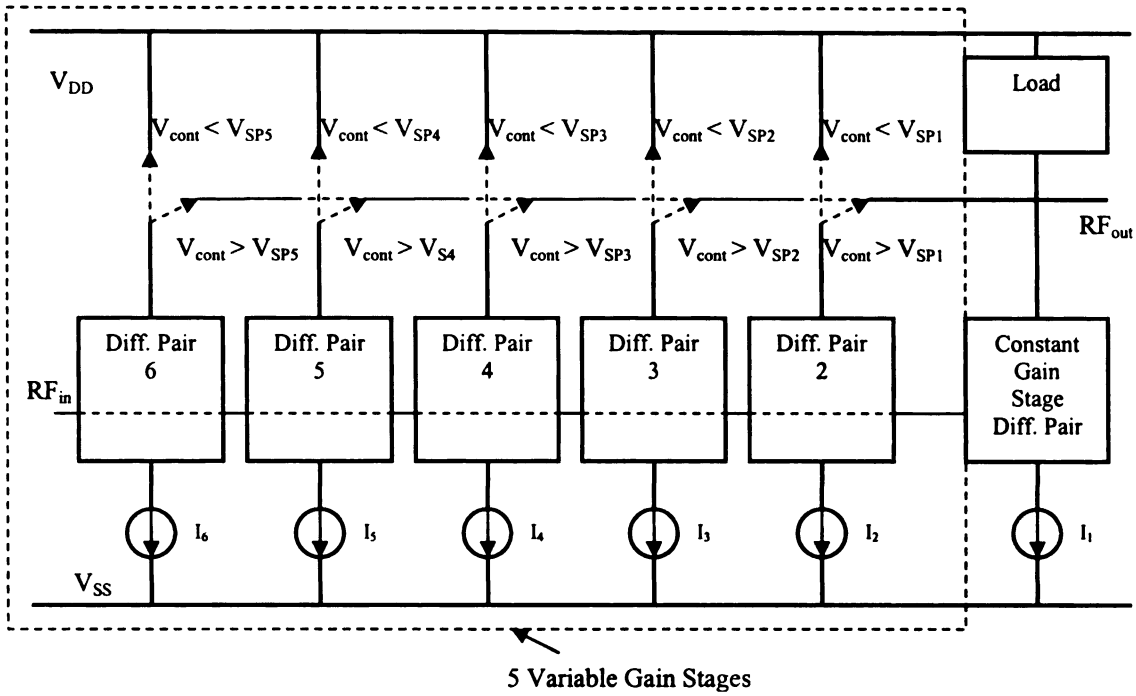


Figure 3.9: Block Diagram of the VGA Functionality.

In Figure 3.9, the current generated by each of the differential pair stages is either steered toward the load or back to the power supply depending on a control voltage V_{cont}. The switching point voltage for each of the analog switches is V_{SPi}. Initially, in the low gain state, only the constant gain stage steers current toward the load and the control voltage remains less than each of the switching point voltages. Therefore, the variable gain stages two through six steer all their current back to the power supply. As the control voltage exceeds the switching point voltages of each stage, current begins to be steered toward the load in a staircase fashion. The switching points $V_{SP1,2,3,4,5}$ are designed so that $V_{SP1} < V_{SP2} < V_{SP3} < V_{SP4} < V_{SP5}$. Therefore, in the high gain state, when V_{cont} exceeds V_{SP5}, all of the current from each stage is completely steered toward the load. In addition, the biasing currents and differential pair transistor sizes are increased with the

increasing stage number. So, the sixth stage is capable of providing the most gain due to the device sizes and biasing current and is termed the "high gain" stage. Figure 3.10 shows an example of one of the variable gain stages in parallel with the constant gain stage.

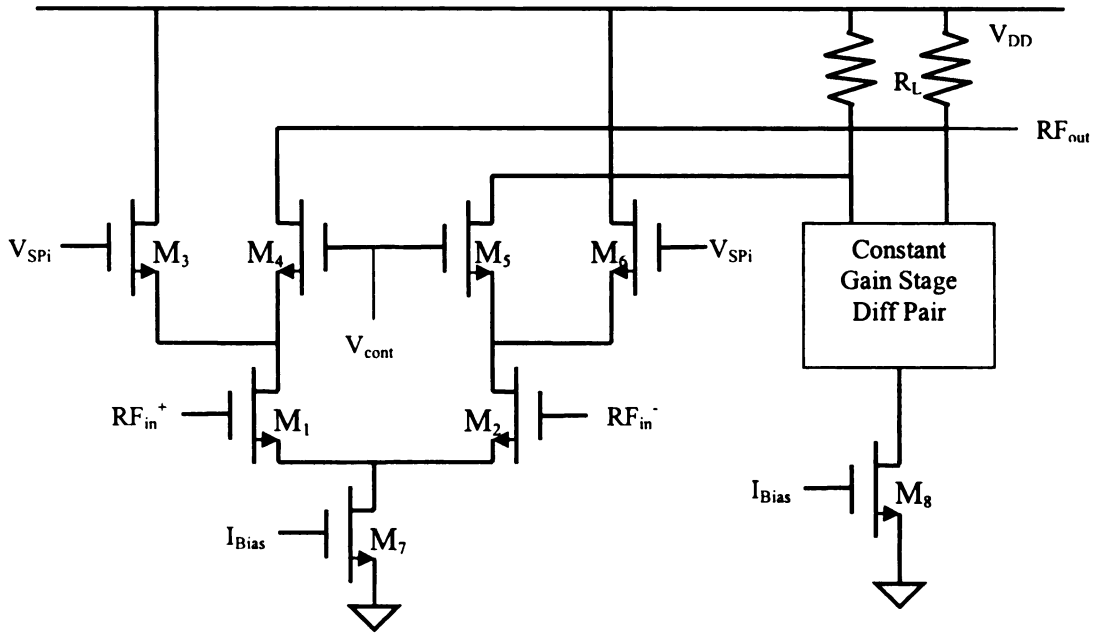


Figure 3.10: Variable Gain Stage.

Transistors M_3 through M_6 are cascode transistors which perform the analog switching function. Similar to the constant gain stage, they also provide shielding from input to output to eliminate the Miller effect. The switching points, V_{SPi} have a uniform separation from one another of 100 mV, with the first switching point (V_{SP1}) at 1.8 V and the last (V_{SP5}) at 2.2 V. The switching points are optimized so that each preceding stage is turned completely on before the next stage is switched in.

3.6 Final VGA Design

Figure 3.9 shows the final RF VGA circuit including a few extra passive components to enhance broadband performance. As the schematic suggests, the first two variable gain stages incorporate degeneration to improve IIP_3 at low gains. The higher order stages do not incorporate any degeneration so that they are capable of providing more gain. In addition, the high gain states are responsible for processing smaller valued RF input signals, so linearity is less of a concern. One of the passives added is a parallel 100 Ω resistor between the positive and negative RF inputs. The resistor is added to force the input impedance of the circuit to approach 75 Ω to improve matching and input return loss. Without the resistor, the input impedance of the VGA is highly capacitive, varies greatly with frequency, and does not provide a good match. A second pair of passives added are the series inductors between the cascode transistors and the load. These inductors extend the bandwidth of the circuit by approximately 200 MHz and increase the gain by approximately 1 dB. It does so by resonating out a portion of the parasitic capacitance which results from the input impedance of the next stage (Mixer) and from the cascode transistors.

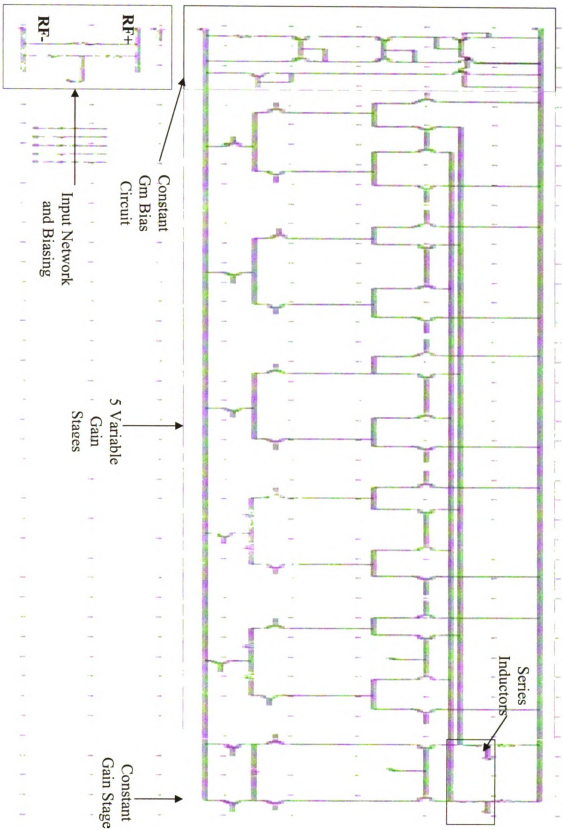


Figure 3.11: Final RF VGA Circuit

3.7 Conclusion

Very few CMOS variable gain amplifier designs that have been published which satisfy the high demands of a satellite tuner. The authors of [4] present one of these designs. This chapter discussed an improved alternative design. The implementation consists of three sub-circuits: the constant g_m bias circuit; constant gain stage; and five variable gain stages. The six stages are placed in parallel to achieve 30 dB of dynamic range and to take advantage of the fact that transconductances add in parallel. This technique allows the circuit to attain high gains while also providing the dynamic range and linearity required of the tuner.

Chapter 4: Simulation Results

4.1 Introduction

This chapter presents the simulation results for the VGA using 0.35 μm CMOS device models in a Spectre RF simulator. The Spectre RF simulator is SPICE based and has many features which make it attractive to RF/analog simulations. In addition, the schematic test bench used to characterize the VGA is illustrated. The results for the various metrics were obtained with the following Analog Artist analysis engines: DC, Transient, S-parameter, and Periodic Steady State (PSS). The simulation data presented includes: Dynamic Range; Frequency Response; Noise Figure; Linearity; Input Return Loss and Stability. Finally, a summary of the simulations results in comparison to the design specifications is tabulated in a design compliance matrix.

4.2 Test Bench

All of the simulations characterizing the performance of the RF VGA were run using the test bench in Figure 4.1.

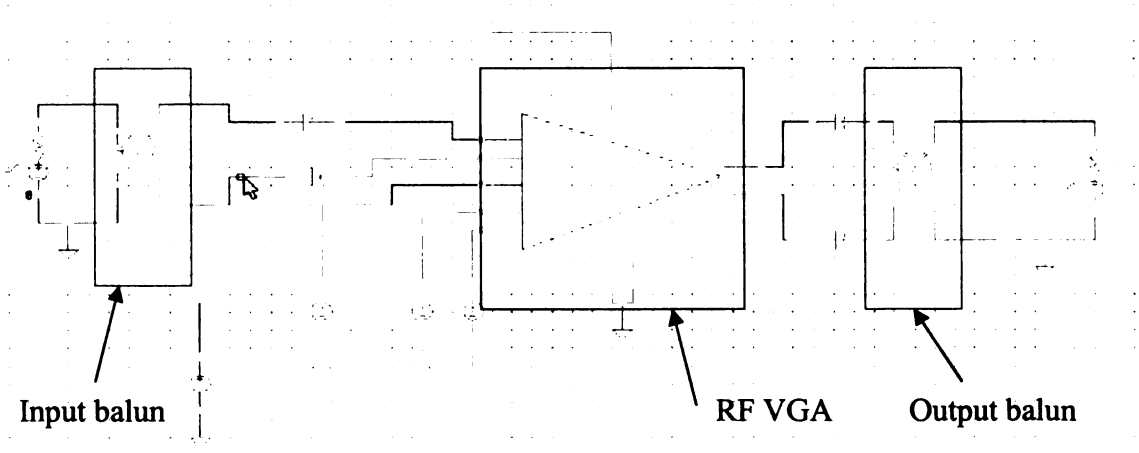


Figure 4.1: RF VGA Test Bench.

In Figure 4.1, a single ended RF signal is used as the input. It is then converted to a differential RF signal via an ideal balun. Next, a pair of capacitors couple the RF input to the VGA to prevent the dc biasing voltage of the VGA from disturbing the RF input. Finally, at the output, a pair of coupling capacitors connect the RF output signal from the VGA to the balun. The capacitors block the DC from the VGA and feed the RF signal through to the mixer, which is the next stage RF front end. In order to measure the differential output of the VGA, an output balun is placed in between the load and the coupling capacitors to generate a single ended output. The load is a 150 Ω resistor used to represent the input impedance of the mixer.

4.3 Dynamic Range

The first and arguably most important metric that was simulated was the dynamic range of the VGA. Figure 4.2 is a graph of the gain at 2 GHz versus the control voltage (Vcon1) and it illustrates the VGAs dynamic range capability. The graph was generated using S-parameter analysis in Cadence by sweeping the control voltage from the minimum value of 1.3V to the maximum value of 2.7V. As the graph indicates, the VGA is capable of generating a dynamic range of -25dB to +8dB. The high end gain meets the requirement at 8dB.

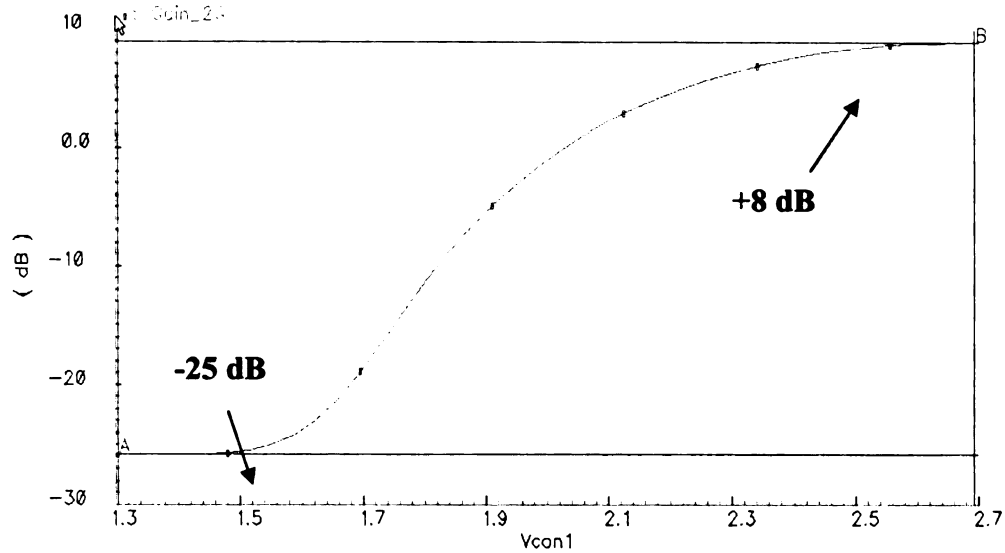


Figure 4.2: Gain versus Control Voltage (Vcon1).

4.4 Frequency Response

The second metric simulated is the frequency response of the VGA. As was stated in previous chapters, the VGA must have constant gain over the entire frequency range. Figure 4.3 shows the frequency response of the VGA prior to the addition of the series inductors. As the graph indicates the capacitive loading at the output node cause the gain to vary greatly over the frequency band of interest. Figure 4.4 is the frequency response of the VGA including the series peaking inductors in the high gain setting. As the plot shows, the frequency response is much improved and there is only approximately .1dB of variation across the entire bandwidth of the circuit.

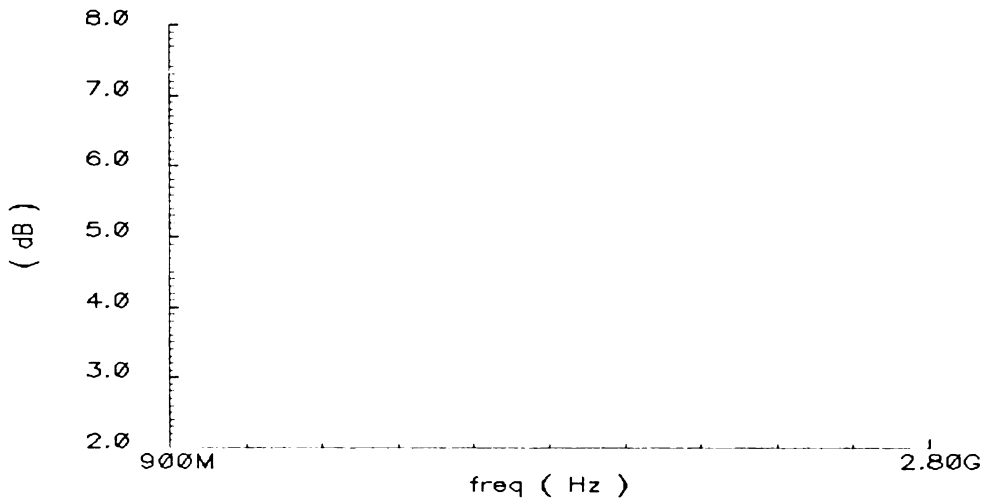


Figure 4.3: High Gain versus Frequency without Series Inductors.

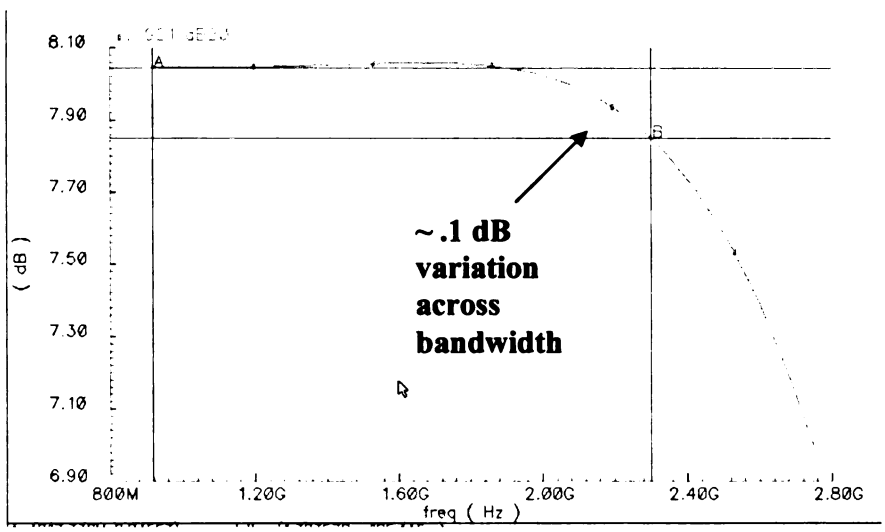


Figure 4.4: High Gain versus Frequency Plot.

In addition, Figure 4.5 shows the flat frequency response of several gain settings between the high and low states.

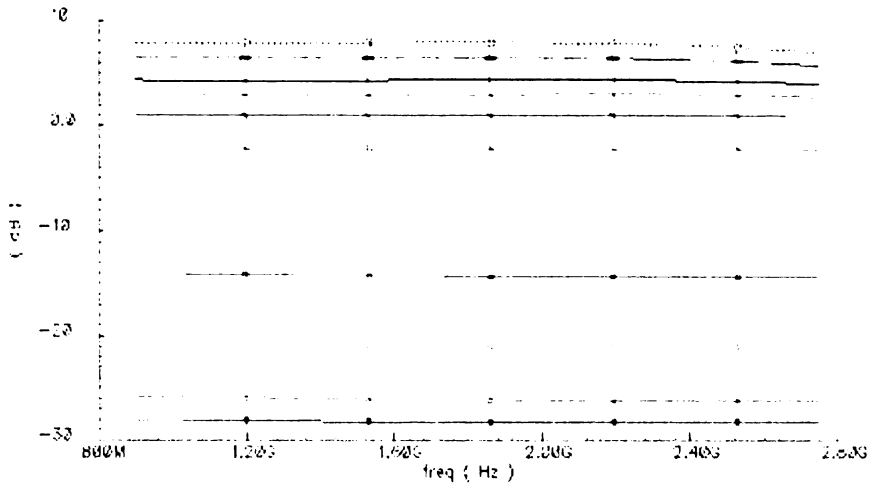


Figure 4.5: Gain versus Frequency and Control Voltage (Vcon1) Plot.

4.5 Noise Figure (NF)

Noise Figure is the next VGA characteristic evaluated. As Chapter 2 discussed, noise limits overall receiver sensitivity. The high gain state of the VGA is responsible for amplifying the weakest RF input signals, and therefore, it determines the minimum noise figure for the circuit. Figure 4.6 is a graph of the VGAs noise figure versus frequency in the high gain state. As the figure illustrates, the VGA yields a noise figure between 5.9 and 6.8 over the frequency range of interest.

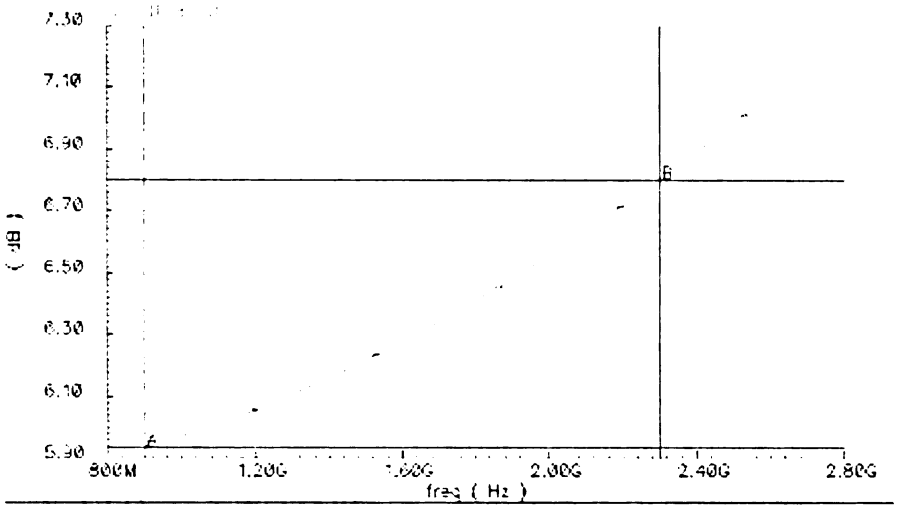


Figure 4.6: Noise Figure versus Frequency at High Gain Plot.

4.6 Linearity (IIP_3)

The next performance metric of interest for the VGA is IIP_3. IIP_3 is limited by large input signals in the low gain state of the VGA. It can be simulated in a number of ways with Spectre RF; however, the most accurate results are generated using a two-tone test and the periodic steady state (PSS) engine. Figure 4.7 plots the power of the fundamental and third order output tones versus input power. The PSS engine then extrapolates the graphs to determine the IIP_3 for the low gain state. As the figure illustrates, the IIP_3 at low gain is 14.98 dBm.

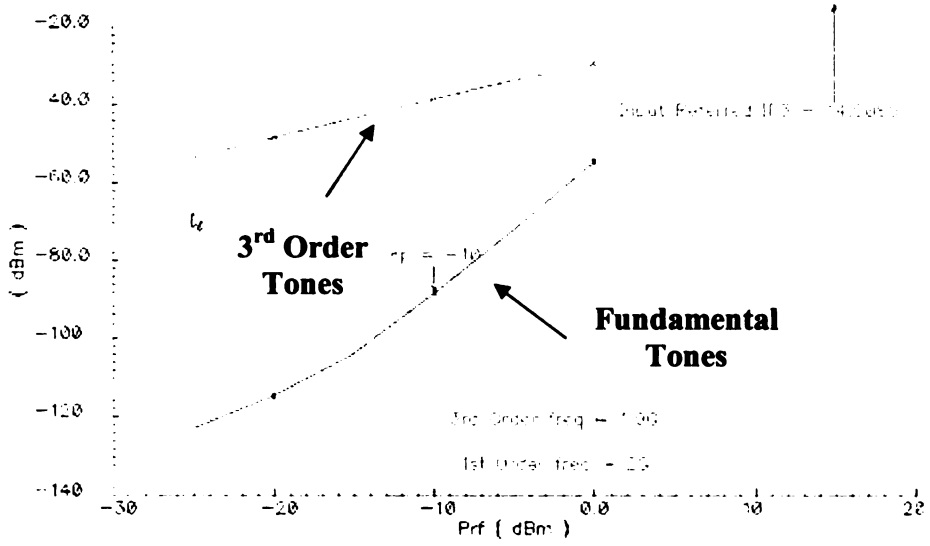


Figure 4.7: PSS Plot of IIP_3 at Low Gain.

In a similar fashion, the IIP_3 was plotted at several different gain settings, and Figure 4.8 graphs the resulting IIP_3 versus gain. The bump at mid-gain occurs as a result of the distortion tones of several of the middle stages being combined. This magnitude of this bump is determined by the breakpoints of the control circuit. For this VGA, the IIP_3 performance exceeds the requirements and the bump actually further improves the margin at mid-gain.

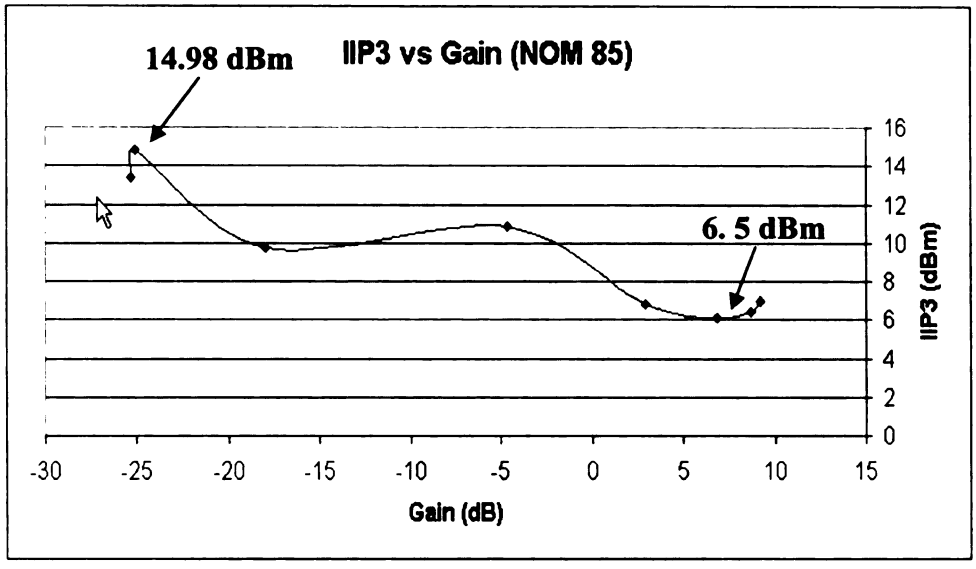


Figure 4.8: Graph of IIP_3 versus Gain.

4.7 Input Return Loss (IRL)

The input return loss performance of the VGA is the next topic. As discussed in Chapter 2, the input return loss quantifies the match of the VGA input impedance to the characteristic impedance of 75 Ω. The initial VGA input return loss simulations are shown in Figure 4.9 and Figure 4.10. Figure 4.9 is the Smith Chart of the input impedance versus frequency and Figure 4.10 plots the S-parameters on a dB-scale versus frequency. The initial VGA design did not include any matching on the RF input. The Smith Chart of Figure 4.9 shows that the input impedance of the VGA is highly capacitive and very poorly matched to 75 Ω. This observation is further quantized by the S_{11} graph of Figure 4.10 because the input return loss is only on the order of -1 to -3 dB. Recall from the VGA specifications that the input return loss should be on the order of -10 dB.

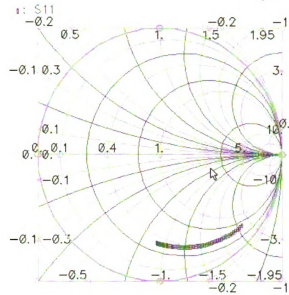


Figure 4.9: Smith Chart for S_{11} without parallel resistor

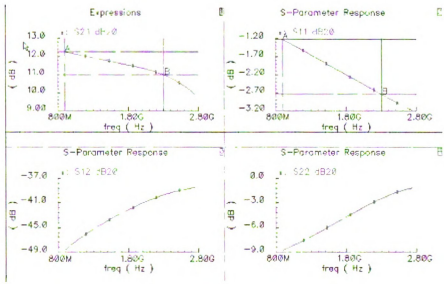


Figure 4.10: S-Parameters without Parallel Resistor.

Broadband matching generally requires a network of several large inductors and capacitors, both of which are too large to implement on-chip. As an alternative, a parallel resistor is implemented at the input on-chip. The resistor degrades the noise figure

slightly, but with it the VGA input return loss is reduced to -8 dB at the high end of the band. The results are illustrated in Figure 4.11 and Figure 4.12. Figure 4.11 is the Smith Chart of the S_{11} and Figure 4.12 is a linear plot of all the S-parameters of the improved design. As the Smith chart indicates, the input impedance is still capacitive; however, it varies less with frequency, and is a better match to 75 Ω than the original design. Also, the input return loss of Figure 4.12 is much improved and on the order of -10 dB for the entire frequency range.

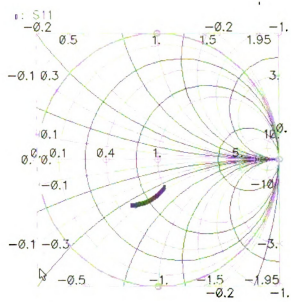


Figure 4.11: Smith Chart for S_{11} with Parallel Resistor.

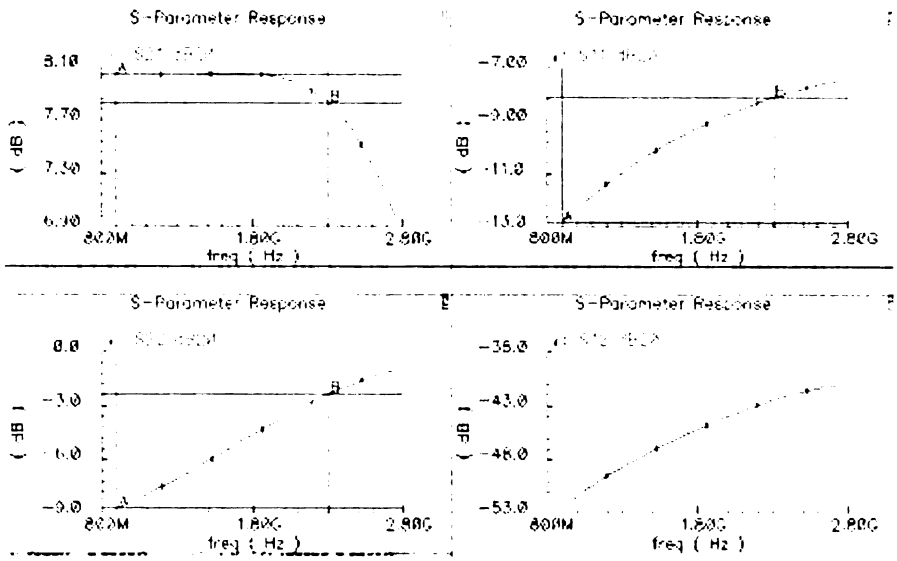


Figure 4.12: S-Parameters with Parallel Resistor.

4.8 Stability

Finally, the last simulation of importance is stability. While all of the other performance metrics are important to the circuits overall functionality, the circuit must be unconditionally stable as well. In the presence of feedback paths from the output to the input, a circuit can become unstable for certain combinations of source and load impedances. The two quantities which are often used to characterize the stability of a circuit are K_f and B_{1f}. A circuit is said to be unconditionally stable if K_f is greater than one over all frequencies, and if B_{1f} is non-negative for all frequencies. Figure 4.13 shows these two quantities, and as the figure illustrates, both of the stability requirements are satisfied. Therefore, the VGA design is unconditionally stable.

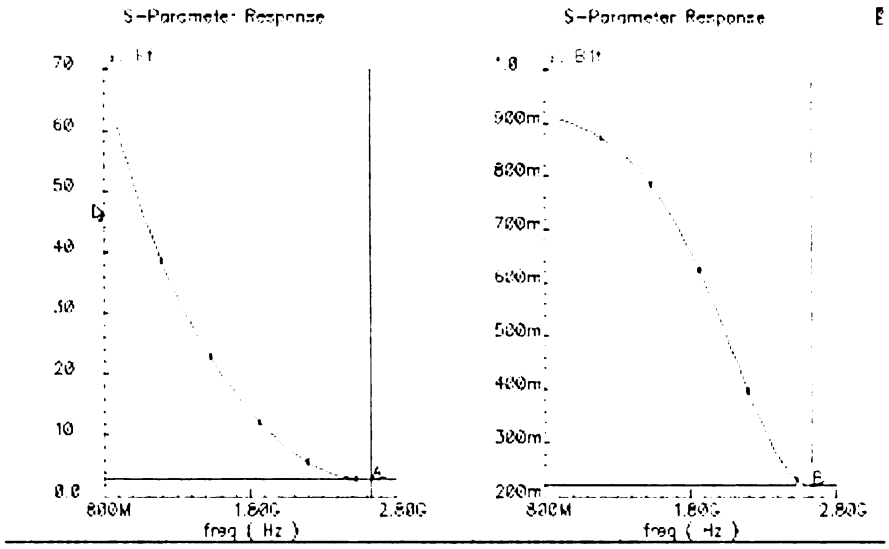


Figure 4.13: K_f and B_{1f} Stability Checks.

4.9 Conclusion and Design Compliance Matrix

This chapter presented the Cadence simulation results of the RF VGA presented in this thesis. The design compliance matrix shown in Table 4.1 summarizes the performance of the RF VGA in comparison to the requirements for Satellite tuners. As the table illustrates, the VGA presented in this thesis complies favorably with all of the specifications for a typical satellite RF front end.

Design Compliance Matrix		
	Requirement	**Simulation**
Dynamic Range	>30 dB	33dB
NF @ Max Gain	6 dB	6 dB
IIP3 @ Min Gain	8 dBm	14.9 dBm
Input Return Loss	>10dB	9-13 dB
Current	22 mA	22.5 mA

Table 4.1: Design Compliance Matrix

Chapter 5: Conclusion and Possible Future Work

5.1 Conclusion

Satellite TV has become a highly attractive alternative to standard cable TV due to its improved performance and reduced cost. A satellite TV system is composed of several highly complex functional blocks. The tuner is one of the most important blocks of the system and it consists of the variable gain amplifier (VGA), mixer, phase locked loop (PLL) and base-band filters. One of the functions that the tuner, and more specifically the VGA, must realize is that it must amplify or attenuate an RF input signal to provide a constant amplitude output signal for the analog to digital converter and demodulator IC. Traditionally, VGAs designed for satellite applications have been fabricated in GaAs, Bipolar, and BiCMOS technologies because of their high speed advantages. However, all of the digital content of the satellite system is implemented in low cost, high volume CMOS technologies. Therefore, from both cost and integration points of view, CMOS becomes a very attractive technology for the satellite tuner as well.

Implementing the VGA design in CMOS presented many challenges to maintaining the necessary performance of the tuner. Some of the challenges included: overcoming CMOS technology limitations; meeting the high dynamic range and broadband frequency response requirements of the system; optimizing specification trade-offs and making the design fully monolithic. More specifically, the VGA had to meet the following design specifications:

- Must be broadband and cover the frequency range of 925 MHz to 2.175 GHz
- A Dynamic Range of -25 dB to ~ 7dB
- Noise Figure with a 6 to 7 dB roof at maximum gain
- Linearity: 8 dBm input-referred third order intercept (IIP_3) floor at low gain
- Input Return Loss of ~ 10dB or greater
- Vcc current drain of ~ 20 mA or less
- Feed a 150 Ω differential load at output (Mixer)

The proposed VGA was simulated using Spectre RF and benchmarked against the satellite tuner requirements listed above. A comparison between the requirements and the simulation results revealed that the design successfully meets the performance specifications of a Satellite TV tuner VGA.

5.2 Possible Future Work

There a number of areas for future work included layout, fabrication, and lab characterization. One of the most challenging aspects to any RF/analog design is translating the schematic into an optimized layout with minimum parasitics. Excessive parasitic capacitance can contribute to gain roll off, poor input return loss, and even degraded IIP_3. Therefore, careful consideration would be required in order to assure proper functionality in generating the layout of the VGA. Secondly, the design would need to be fabricated in the Jazz 0.35um process, since it was optimized and simulated with the Jazz design kit. Unlike digital design, RF/Analog circuits are often optimized

for a particular technology and are not directly portable to different foundries without significant redesign effort and layout from scratch. Lastly, the VGA would have to be setup for evaluation either in packaged form or in probable breakout cells for characterization on an RF wafer probe station. In packed form, the VGAs would require MLF leadless packages with minimum ground and lead inductance due to the RF nature of the design as well as an evaluation PC board. For wafer probe, an RF wafer probe station would be required with two 50Ω RF probe needles, one for the input and one for the output. Once the proper layout, fabrication, and packaging of the VGA were completed the next step would be to test the circuit in laboratory to verify the model correlation.

Another possible area for future work would be to investigate converting the 0.35μm design to a 0.18μm technology to take advantage of the higher transconductances characteristic of smaller feature sizes. The increased transconductances could possibly increase the gain in the high state, and therefore increase the dynamic range of the circuit. However, a direct conversion to the 0.18um technology might not be possible due to reduced power supply voltages and headroom considerations.

BIBLIOGRAPHY

[1] A. Kwentus, P.Pai, S. Jaffe, R. Gomez, S. Tsai, T. Kwan, H. Hung, Y. Shin, V. Hue, D. Cheung, R. Khan, C. Ward, M. Ku, K. Choi, J. Searle, K. Bult, K. Cameron, J. Demas, C. Reames, and H. Samueli, "A Single-Chip Universal Digital Satellite Receiver with 480-MHz IF Input," IEEE J. Solid-State Circuits, vol. 34, no. 11, pp. 1634-1646, Nov. 1999.

[2] B. Kim, S. Kim, T. Lee, J. Kim, Y. Kim, M. Jeong, K. Kim, S. Kim, S. Park, and B. Ko, "A CMOS Single-Chip Direct Conversion Satellite Receiver for Digital Broadcasting System," 2002 Symposium on VLSI Circuits Digest of Technical Papers, pp. 238-241.

[3] J. Rijns, "CMOS Low-Distortion High Frequency Variable-Gain Amplifier," IEEE J. Solid-State Circuits, vol. 31., no.7, July 1996.

[4] B.Kim, H. Yoon, Y.Cho, J. Lee, S. Kim, T. Lee, J. Lim, M. Jeong, B. Kim, S. Kim, S. Park, B. Ko, S. Yoon, I. Jung, Y. Oh, and Y. Kim, "A 9dBm IIP3 Direct-Conversion Satellite Broadband Tuner-Demodulator SOC," 2003 IEEE International Solid-State Circuits Conference.

[5] K. Koh, Y. Youn, and H. Yu, "A Power-Efficient, Low-Distortion Variable Gain Amplifier Consisting of Coupled Differential Pairs," IEEE Journal of Solid-State Circuits, vol. 32, no. 12, pp. 2105-2110, Dec. 1997.

[6] B. Razavi, *RF Microelectronics,* Upper Saddle River, NJ: Prentice Hall PTR, 1998.

[7] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press 1998.

[8] B. Razavi, *Design of Analog CMOS Integrated Circuits,* New York, New York: McGraw-Hill, 2001.

[9] P. Gray, P. Hurst, S. Lewis, and R. Meyer, *Analysis and Design of Analog Integrated Circuits,* Wiley, 2001.

[10] D. Cheng, *Field and Wave Electromagnetics, Addison-Wesley, 1989.*

[11] D. Johns and K. Martin, *Analog Integrated Circuit Design*, Wiley, 1997.