10-MHz 60-dB Dynamic-Range 6-dB Variable Gain Amplifier

A thesis submitted to
The Hong Kong University of Science and Technology
in partial fulfillment of the requirements for
the Degree of Master of Philosophy in
Electrical and Electronic Engineering

by

Leonard Hon Yan LEUNG
Department of Electrical and Electronic Engineering
B.Eng. in Electronic Engineering, HKUST
Hong Kong, 1994

August 1996

10-MHz 60-dB Dynamic-Range 6-dB Variable Gain Amplifier

by

Leonard Hon Yan LEUNG

Approved by:

Dr. Jack LAU, Thesis Co-Supervisor

Dr. Bertram E. SHI, Thesis Examination Committee Chairman

Dr. Wing-Hung KI, Thesis Examination Committee Member

Professor S. Simon WONG, Acting Head of the Department

Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology

August 1996

To my parents —

for their constant support and encouragement.

Contents

List of Figures

List of Tables

Acknowledgments

First of all, I would like to express my gratitude to my supervisor, Dr. Aaron W. Buchwald, for his generous support and constructive advises. From the beginning, he had given me a lot of opportunities to broaden my horizon and offered me an excellent MPhil program. I learned a lot from him, and I could not imagine what would happened to my academic path if I have not met him. Now, he is working at Broadcom Corporation. He often shows me his works as design examples and says, "Leo, if I can, you can do it, too." I want to thank him for his confidence in me. His encouragement and tolerance of my weakness make all the ideas in this project possible.

I would also like to thank Dr. Jack Lau, Dr. Wing-Hung Ki, Dr. Bertram Shi and Professor Simon Wong to serve in my Thesis Examination Committee.

I am really lucky that I can work out part of my project at Broadcom Corporation for half a year. All the staffs at Broadcom Corporation are crucial to my project. Their experience, existing works and advises had played an important role in my research. Loke Tan, Joe Laskowski and Agnes Woo have taught me how to do a good and reliable circuit layout. Chi-hung Lin arranged my chip's fabrication with MOSIS. That saved a lot of my working hours.

I am also indebted to the faculties and students at the Department of Electrical and Electronic Engineering at HKUST. I am one of the first class of students. All the faculties have spent much effort in preparing good programs that benefited me. I especially thank Dr. Jack Lau, as he helped me to get into this university, and Dr. Howard Cam Luong, for whom I had been working as TA for three semesters. They have taught me many things personally and professionally. Furthermore, my classmates and my students have asked a lot of interesting and intelligent questions that force me to think and to have a better understanding about the materials in my working area.

It has been a wonderful experience to work in the Analog Research Group in the past few years. The members are hardworking and self-motivated. Paul Chan, Darwin Cheung, Felix Cheung and Kenneth Choi and me studied in the same undergraduate program. I have been influenced greatly by their attitude. My present classmates, Allen Ng, Felix Wong, Darwin Yim, Pang Ho-Yin, Jeffrey Ho, Alice Chan and Wong Man-Kit have also helped me a lot.

I am fortunate to have friends like Thomas Leung, William Tse, Angel Choi, Tony

Au Yeung, Eric Yeung, Chan Hung-Kee, Edmond Kwok and Alva Hon. They are all willing to share their good times and bad times with me. Ray Cheng and Lily Yeung are as close to me as my brother and sister. They would share everything with me unconditionally. I would like to express my deep thanks to all of them here and hope they will always stay with me.

My father, mother and brothers are the greatest support and encouragement to me. Otherwise, I would not have enough strength and confidence to survive all these years. Their devoted love, forgiveness and guidance make a great influence to my life. My elder brother and his wife, Dora, have immigrated to Australia. They often call me up and ask about the progress of my study. Although it has been a pressure to me, I would feel their deep concern. However, it is my greatest regret that my mother can never read this thesis. I believe she is now in heaven and I would always miss her. I really want to share my happiness with her on the day I graduate.

Lastly, I give thanks to my Lord. Hallelujah.

Leung Hon Yan Leonard
HKUST, August 1996

Abstract of the Thesis

This thesis describes a CMOS 10-MHz Variable Gain Amplifier (VGA) with 60-dB dynamic-range. The gain is continuously tunable over a 6-dB range, with a minimum gain of unity and a maximum gain of two. Continuous gain tuning is accomplished by a novel method of interpolating between 10 taps of a feedback resistor, respectively connected to a fully-differential multi-input-stage operational amplifier. Measured results with a 10-MHz input show THD below -60-dB with a 4-V_{pp} differential output signal. Third-order intermodulation distortion, for two equal amplitude tones at 9.5-MHz and 10.5-MHz, remains below -59.17-dBc over all gain settings. The circuit has been fabricated in a 0.8-μm CMOS process (Hewlett Packard CMOS26G), occupies an area of 0.6×1-mm^2 and dissipates 80-mW. The VGA is applicable to 10-bit resolution video systems and digital broadband communication signal-conditioning front-end circuitry.

Chapter 1

Introduction: Overview of CMOS Video-Rate VGAs

In virtually all communications systems the received power is unpredictable. This is especially true in wireless mobile systems, where the distance between transmitter and receiver varies continually. In telephony, the routing of a signal can change for each phone call, thus changing the channel characteristics and attenuation seen at the receiver. Likewise, in cable-TV, the cable lengths are never exactly the same from the head-end to subscriber and cause each set-top-box to receive signals with different amplitudes. To buffer receiver electronics from such changes in signal strength, a Variable Gain Amplifier (VGA), is typically employed in a feedback loop to produce an Automatic Gain Control (AGC) circuit. This keeps the output level of the AGC constant, regardless of the input level. The remainder of the receiver can then be designed for a constant, well-controlled input level.

As the sophistication of broadband communications systems, such as digital-video and personal communication, increases, so too does the required dynamic-range and bandwidth of the analog front-end circuitry, of which the VGA is an integral part. These analog signal-conditioning circuits often limit the performance of the overall system. Therefore, achieving high linearity at ever increasing bandwidths is an essential requirement of all analog front-end circuits.

The goal of this research, is to design and test a CMOS VGA with continuously variable gain, a 60-dB dynamic-range and a bandwidth greater than 10-MHz. The power dissipation from a single 5V supply should be less than 50-mW. Such a VGA will find applications in wireless personal communications, broadband residential communications, digital cable-TV and satellite television, where this demanding dynamic-range specification is required to maintain 10-bit of resolution.

1.1 Background

There are three fundamental ways to design a VGA. The first makes use of a switching network to change between various passive feedback elements, thus producing discrete gain steps. The second method continually varies a parameter of the circuit, either in the feedforward or feedback path, to adjust the gain. The third uses dual signal paths; one path is attenuated and subtracted from the other to achieve a large gain variation. Each method has advantages and disadvantages, which are explained below.

Discrete Gain Steps It is well known that negative feedback can be utilized to reduce distortion, provided linear feedback elements are available. Typically only passive elements meet 10-bit linearity requirements. Therefore, to utilize passive feedback elements in a VGA, they must be selectively switched into the signal path. A simplified schematic of a VGA with N gain settings is shown in Fig. 1.1. Because

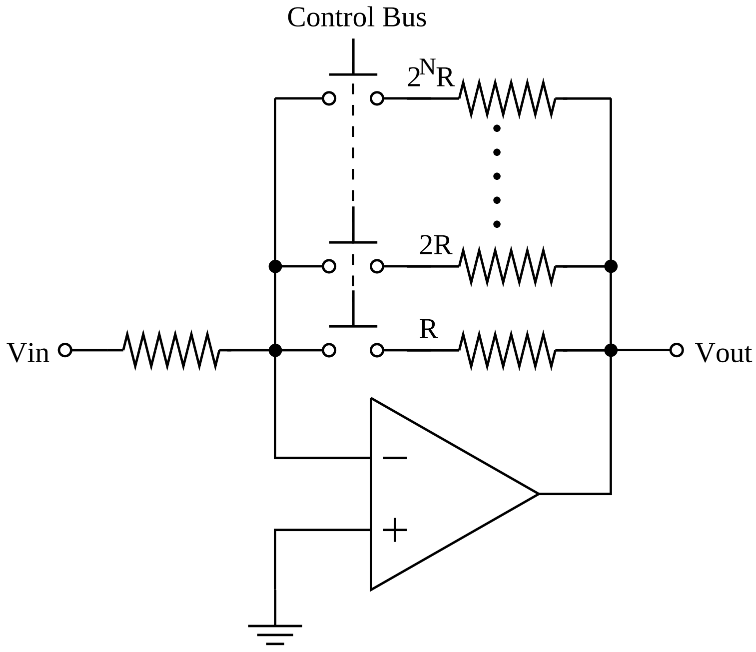


Figure 1.1 Simplified schematic of a VGA using a switched feedback network of passive resistors to achieve discrete gain settings.

such circuits can take full advantage of negative feedback, they are capable of achieving good linearity for large output voltages. However, the gain can only be set at discrete levels, which may be a disadvantage for some types of systems.

Continuous Gain Variation To achieve continuous control of the gain, one or more parameters of the circuit, such as the bias current, or the bias voltage, can be continuously changed. Examples of this method are shown in Fig. 1.2. The bias current of a differential pair (diffpair) can be tuned to change the gain as shown in

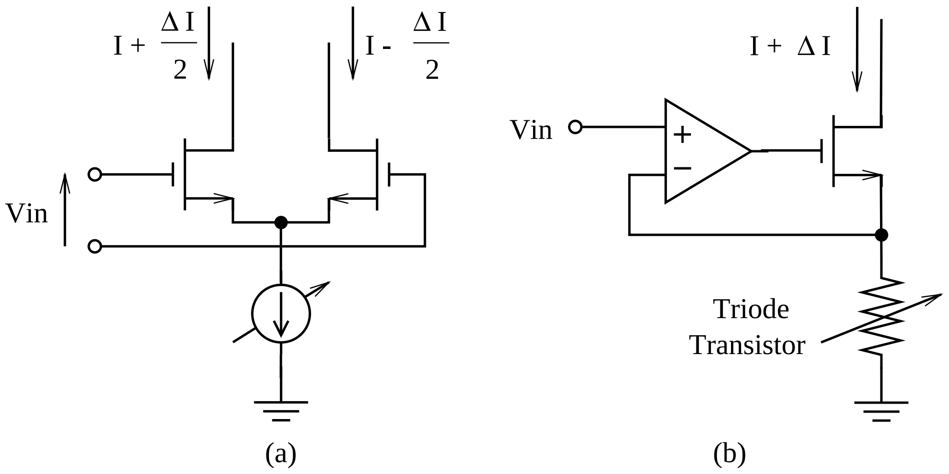


Figure 1.2 Two simple VGAs with continuously variable gain: (a) gain is adjusted by varying the bias current, (b) gain is adjusted by varying the bias voltage of a triode resistor.

Fig. 1.2a, whereas Fig. 1.2b illustrates a circuit which changes the resistance of a gain-setting resistor, which could be constructed from a linearized MOS transistor operating in the triode region. The advantage of the above circuits is that the gain can vary smoothly between extreme values. The disadvantage is that linearity is typically poor. Since the gain is changed by altering the bias point of an inherently nonlinear device, the only way to achieve acceptable distortion performance is to keep the signal levels low, at which point noise limits the dynamic-range. Therefore, circuits of this type are usually limited to small output signals, on the order of 250-mV, for 60-dB dynamic-range in a 10-MHz bandwidth.

Variable Signal Subtraction Another method of achieving continuously variable gain is to use a multi-path circuit, where each path experiences a controlled attenuation before being summed at the output. A common circuit to implement this function is shown in Fig. 1.3. This multiplier circuit has dual signal paths which are summed in opposite phase at the output. The tail current of each diffpair can be adjusted by a tuning voltage V_{ctl}, which determines the relative gain of the positive and negative paths. This circuit is symmetric, such that the gain can be made negative as well as positive. The advantage of this circuit is that the gain can be varied continuously over a large range. The disadvantage is that the noise power doubles due to the dual signal paths, and the linearity is poor for large signals, which limits the dynamic-range.

Transconductance Linearization Techniques A linear tunable transconductor is an essential element in continuous-time filters and there has been a great deal of research in this area. Many of the techniques in CMOS use a feedforward error cancellation technique, which compensates for nonlinearities by subtracting an error of equal magnitude. Some circuits are based on linearizing triode resistors, while others are based on the square-law behavior of a MOSFET in the linear-active region.

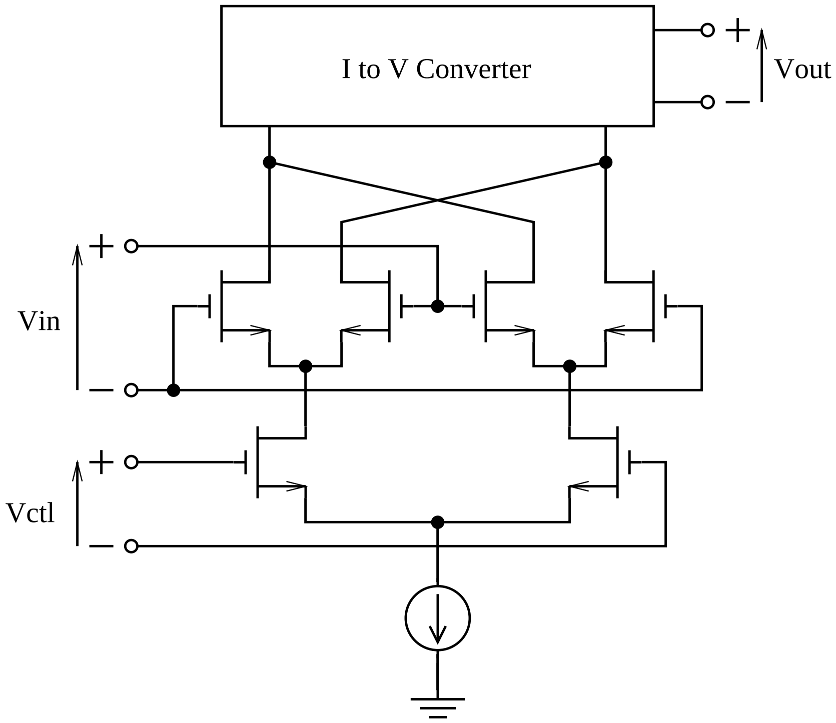


Figure 1.3 A CMOS multiplier circuit used as a VGA.

Although several interesting techniques have been reported [1, 2, 3, 4], virtually all suffer from the following shortcomings:

Matching Accuracy Since linearization is obtained by a feedforward cancellation technique, the achievable linearity is limited by matching accuracy.

Increased Noise Additional elements are needed to generate equal and opposite error signals or to sense errors and subtract them. These additional elements add noise to the circuit.

Reduced Signal Amplitude In the cancellation of errors, a portion of the signal is inevitably subtracted, thus reducing signal amplitude.

Due to the final two short-comings: *increased noise*, and *reduced signal amplitude*, these feedforward linearization techniques have a double penalty in the signal-to-noise ratio (SNR). Although the low-frequency linearity of these circuits may improve, the SNR is degraded, resulting in either minimal, or in some cases, no improvement in dynamic-range.

1.2 Motivation and Challenge

The previous section shows that there are two basic choices to be made in considering VGA performance:

- either get high linearity with a large output voltage swing, but be restricted to discrete gain steps,
- or have a continuously tunable gain range, but be restricted to an output voltage swing of only 100–200-mV.

Although it may be possible, in some cases, to follow a continuously tunable gain VGA with a fixed-gain amplifier, such as the one shown in Fig. 1.4, often the signal amplitude is large at the input and would have to be attenuated to be compliant with this type of VGA. Such attenuation will drop the signal, relative to the noise floor, and will lower the SNR. The goal of this research is to design a VGA which

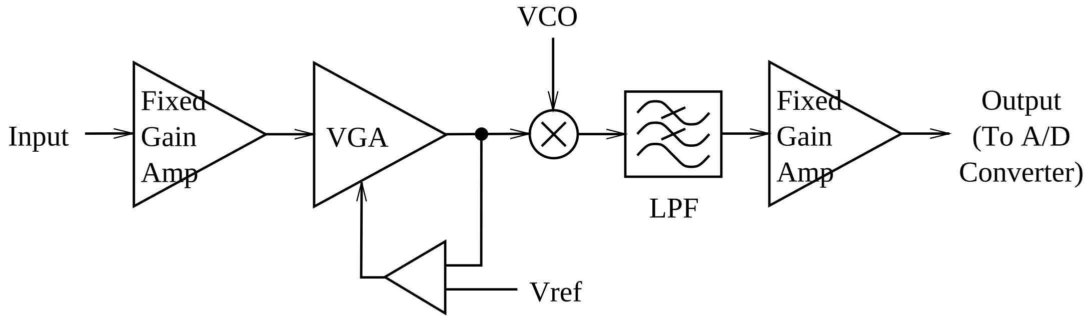


Figure 1.4 Block diagram of an analog front-end IF down conversion receiver with a VGA to maintain a fixed amplitude signal at the input of the mixer.

has the advantages of all the techniques described above without the disadvantages. Therein lies the fundamental challenge of this project: *achieving high linearity for a large voltage swing and still maintaining a continuously variable gain.*

Proposed VGA The approach taken here is to utilize the inherent linearity achievable from negative feedback in a discrete gain-step approach and interpolate between steps to achieve continuous gain tuning. A schematic of this VGA is shown in Fig. 2.7. Before proceeding to the discussion of the circuit design, we will briefly describe how this VGA might be used in application.

1.3 Application Notes

Small Gain Range The proposed VGA has a large output voltage swing of 4-V_{pp} differential and a relatively small gain range of 6-dB. This type of VGA is applicable to ICs where the input signal is on the order of 0.5–1.0-V and the signal amplitude does not vary significantly. Such a VGA could be used to directly drive an Analog-to-Digital Converter (ADC). An example of this front-end configuration is shown in Fig. 1.5. This front-end is typical of digital baseband communication systems

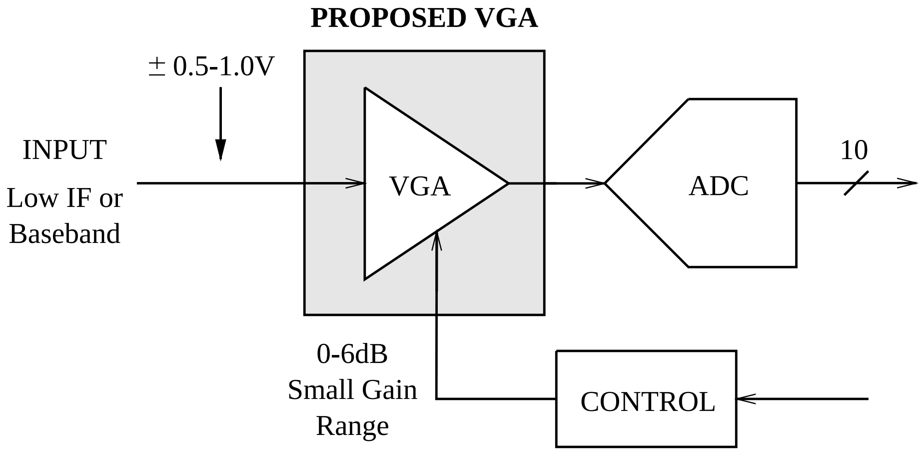


Figure 1.5 Possible front-end using a proposed VGA to drive an ADC.

over short (<1-km) wires. Since the wires are short, the signal strength will not vary more than 6-dB. Therefore, the gain range of the VGA is adequate to cover all expected amplitudes of the incoming signal.

Large Gain Range In a wireless system the signal strength may vary by as much as 40–60-dB. Therefore, a much larger gain-tuning range is needed. It is often advantageous to achieve the gain variations in two steps. This can be understood by considering the down-converting front-end circuit of Fig. 1.6. In this circuit the

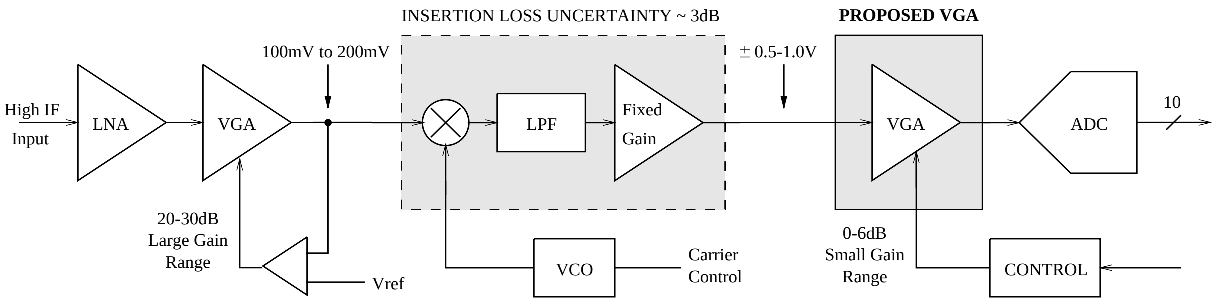


Figure 1.6 Possible front-end which performs frequency down conversion. Variable gain is achieved in two stages: the first stage has a large gain range and stabilizes the input signal to the mixer, the second VGA fine-tunes the gain to optimally load the ADC.

large gain variation is achieved using a VGA immediately after the first-stage Low-Noise Amplifier (LNA). The AGC loop using the first VGA uses local feedback and is not controlled by the digitized data. It is needed to prevent overdriving the mixer, which needs a well-controlled amplitude for optimal linearity.

Advantage of a Dual AGC System A problem that arises in practical systems is that the mixer, and the circuitry following the mixer, such as the lowpass filter and fixed-gain amplifier, will have an uncertain loss, due to parasitic elements. Depending on processing, this loss can vary from 1–6-dB. If a single AGC loop were used at the front-end, the gain of the VGA would need to compensate for the uncertainly in the insertion loss. However, in doing so, it could load the mixer with a signal level that does not give optimal performance. Spurious tones could be generated from the mixers nonlinearities, which would adversely effect the overall dynamic-range of the front-end.

To prevent mixer overloading, a dual AGC system can be used, such as the one shown in Fig. 1.6. The second AGC compensates for the variations in insertion loss, but does not affect the voltage level at the input of the mixer. The second VGA can be adjusted to provide optimal loading for the ADC. It is desirable for the second, fine-tuning VGA, to have a continuous gain range, so that the ADC loading can be set to a precision of 1-LSB (Least Significant Bit). Such a configuration may improve the overall dynamic-range of the front-end, over a single AGC system, by 3–6-dB or 0.5–1.0-bit. The VGA described in this research is ideally suited to this application. In the following chapters the design considerations and measured results of this interpolating VGA will be given.

Chapter 2

Circuit Design: New VGA using Negative Feedback and Interpolation

This VGA circuit is based on a fully-differential multi-input-stage operational amplifier (opamp) with multiple taps on a feedback resistor. In this chapter, the concept of using a multi-input-stage opamp for the purpose of achieving a VGA function, which can smoothly change the voltage gain, will be explained first. Next, the complete architecture and the sub-circuits of the VGA will be discussed. Finally, the control circuit, needed to adjust the voltage gain, will be studied.

2.1 Interpolation Using Dual-Input Operational Amplifier

Most opamps, as shown in Fig. 2.1, consist of three stages: an input-stage, a gain-stage and an output-stage. In general, only the input- and the output-terminals are

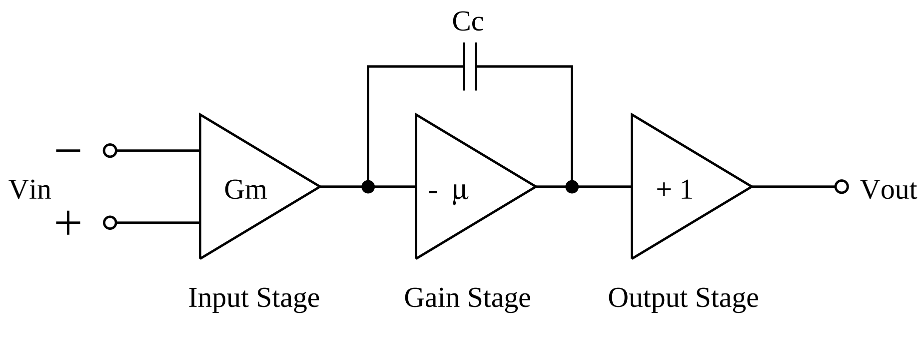


Figure 2.1 Block diagram of a simple operational amplifier.

connected to the external feedback path(s). The closed-loop gain of such an opamp can be altered by adjusting the ratio of feedback elements. In the case of resistive feedback, a gain adjustment may be achieved in two distinct ways: either by sliding the inputs of the opamp on the feedback resistors, which is equivalent to using a mechanical potentiometer, or by switching, or interpolating between pre-existing feedback paths. Our goal is to utilize a multi-input-stage opamp and interpolate between these stages by building an electronic equivalent of a mechanical potentiometer.

Dual-Input Opamp For a feedback amplifier with multiple feedback paths, when any one of the feedback paths are connected, or disconnected, the closed-loop voltage gain will be changed abruptly. However, we would like to be able to change the gain smoothly between these extremes. To illustrate how this can be accomplished, consider an opamp which has two transconductance cells (G_{m} cells) as an input stage, such as the one shown in Fig. 2.2a. This dual-input opamp can be connected with the feedback network depicted in Fig. 2.2b [5]. Each transconductance cell,

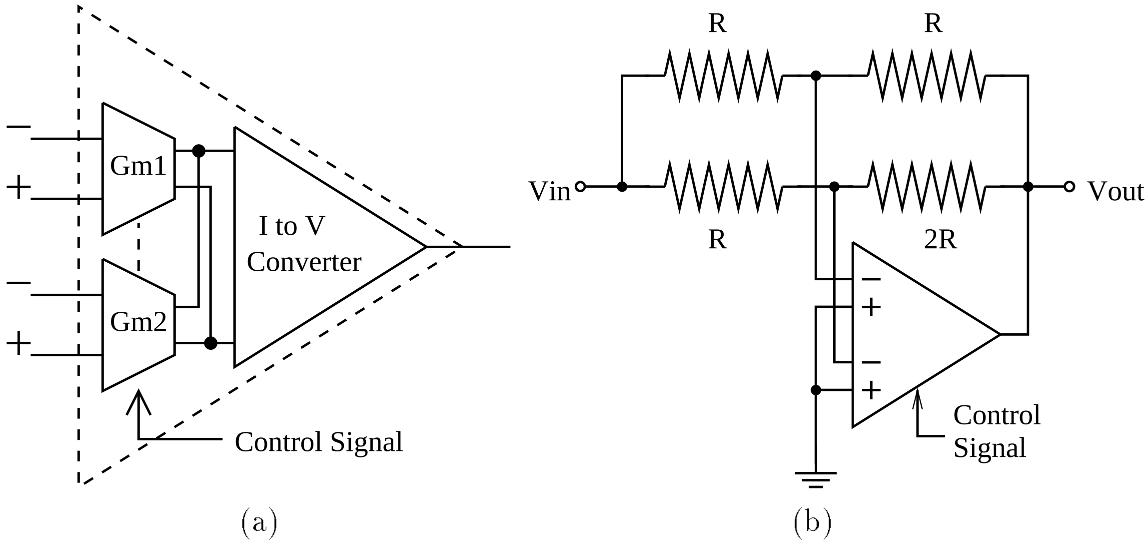


Figure 2.2 A modified operational amplifier which has two transconductance cells as the input stage, (a) internal structure and (b) feedback connection.

G_{m1} and G_{m2}, in combination with the I/V converter constitutes a simple opamp. The control signal in the circuit is used to enable (ON-state) one G_{m} cell and disable (OFF-state) the other. We will consider the pair of simple opamps constructed from G_{m1} followed by I/V and G_{m2}–I/V one at a time. When G_{m1} is ON and G_{m2} is OFF, the overall amplifier looks like a simple opamp with a resistor of value R at the input and a resistor of value R in the feedback path. Therefore, the gain is -1. In the other extreme when G_{m2} is ON and G_{m1} is OFF, the circuit looks like an opamp with an input resistance of R, but the feedback resistance is now $2R$, which gives a gain of -2.

Interpolation Between Input Stages By using a control signal which switches the G_{m} cells either ON or OFF, we obtain a VGA with only two separate gain settings. However, if the control signal is handled properly and provides a smooth transition between the ON- and the OFF-states, the closed-loop gain may take on a continuum of values in the range $[-2, -1]$. For the input stage in Fig. 2.2, each $G_{\rm m}$ cell consists of a current source and a differential pair (diffpair), as shown in Fig. 2.3. The current sources are controlled by the voltage $V_{\rm ctl}$. A simple circuit to provide this control is the multiplier circuit of Fig. 2.4. As $V_{\rm ctl}$ is swept linearly, the bias current is smoothly steered from the first diffpair to the second.

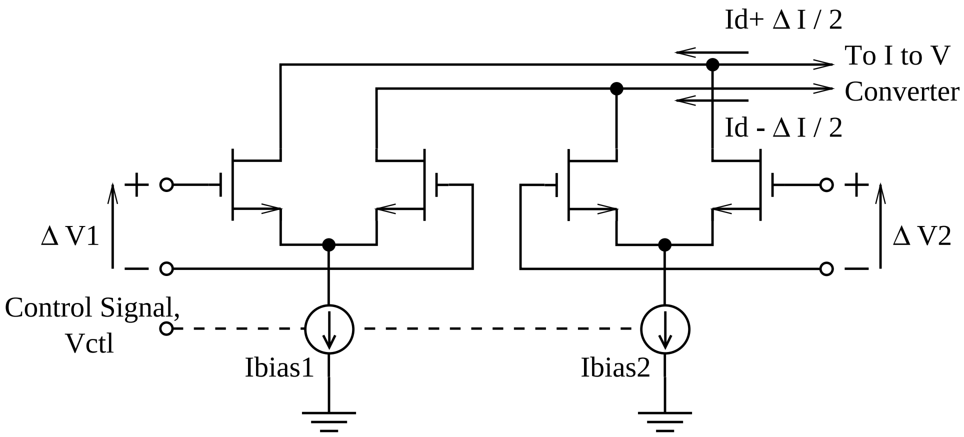


Figure 2.3 The input stage with two transconductance cells.

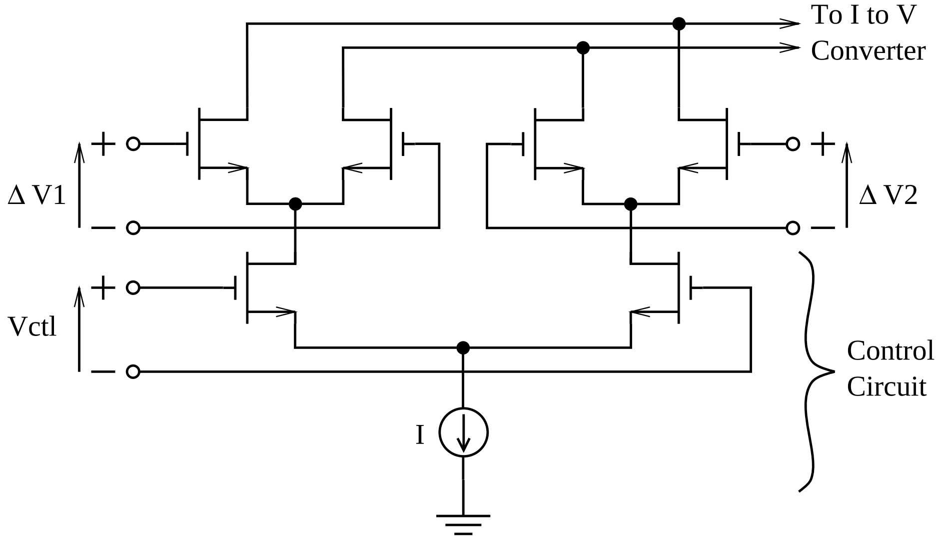


Figure 2.4 Dual-input-stage and its control circuit.

Virtual Ground It is helpful to consider the *projection* diagram of Fig. 2.5 to understand how the dual-input opamp of Fig. 2.2b works as a VGA. When the gain is -1, the virtual ground (pivot-point of the projection diagram) is between the input resistor of value R and the feedback resistor of value R. When the gain is -2, the virtual ground is between the input resistor of value R and the feedback resistor of value $2R$. When the gain is set to a value between -1 and -2, the input voltage

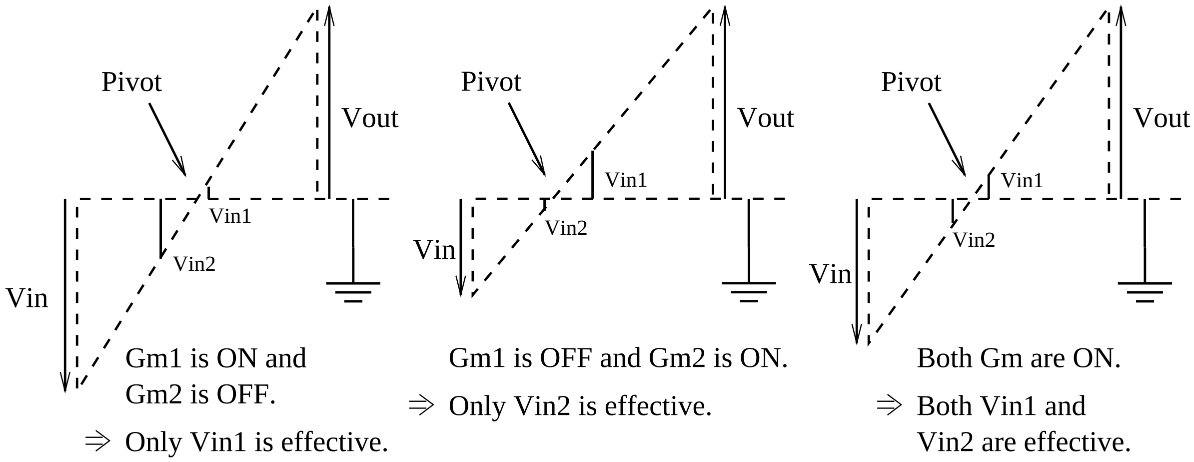


Figure 2.5 The projection diagram for an amplifier with the closed-loop gain from 1 to 2.

of two G_{m} cells will be 180° out of phase. G_{m1} will contribute a positive differential output current, while G_{m2} will contribute a negative differential output current. The effective virtual ground will be somewhere between each of the extremes and each diffpair will see a non-zero input voltage which is sufficient to cause the sum of the currents from G_{m1} and G_{m2} to go to zero.

2.2 Architecture of the VGA

Tuning Range vs. Distortion In the previous section it was shown that a dual-input opamp can be used to construct a VGA. However, it was also shown in Fig. 2.5 that when the VGA is interpolating between extreme gain points, there will be non-zero voltages on the differential inputs. This non-zero voltage will cause a nonlinear current to be generated from each diffpair. The larger the gain separation of the extremes, the larger will be the voltage excursion on the diffpairs. Therefore, the achievable tuning range will be limited by distortion requirements.

In order to change the gain continuously over a larger range, and still maintain acceptable distortion performance, the number of G_{m} cells may be increased such that the input stage can be switched among a set of pre-existed feedback paths. This will keep the gain variation between adjacent taps low, thus reducing distortion. The tuning range can then be increased by increasing the number of taps. A block diagram of this multi-input-stage opamp is shown in Fig. 2.6. The opamp is fully-differential so as to be insensitive to common-mode interference and to attenuate even harmonic distortion.

Modified Feedback Network The feedback network can also be modified; instead of interpolating between a parallel bank of feedback resistors, we simply use multiple taps of a single resistor, as shown in Fig. 2.7. Since the inputs of the

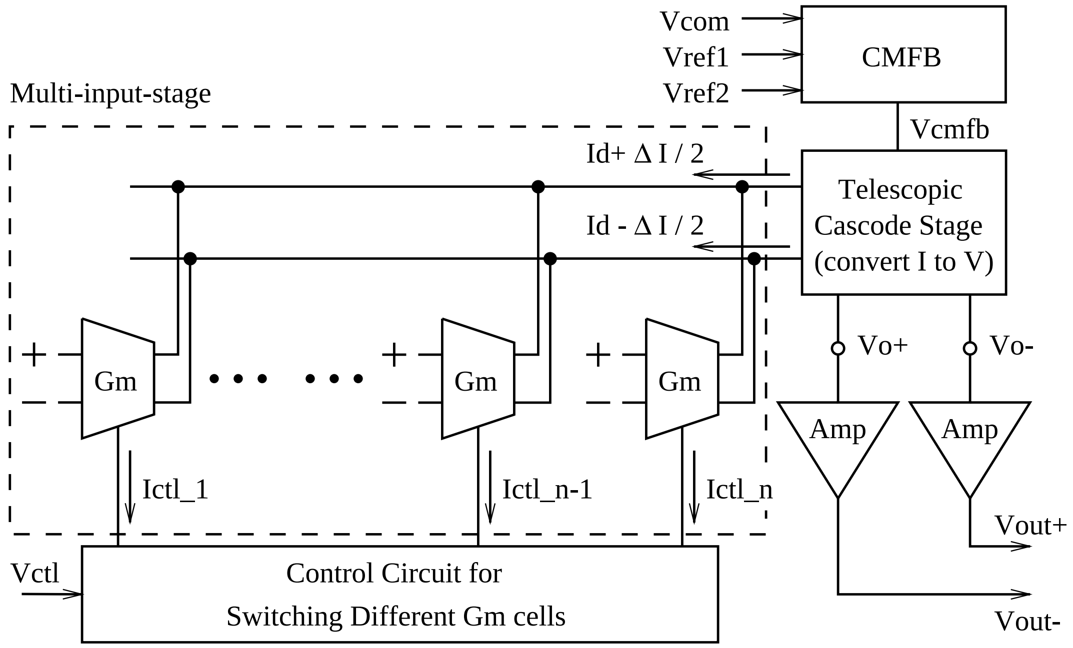


Figure 2.6 Block diagram of an operational amplifier which has a series of transconductance cells as the multi-input-stage.

opamp will be tapped linearly on the same feedback resistor, the gain will be varied monotonically with respect to the control voltage. In the following sections, the sub-circuits and the components depicted in Fig. 2.6 and Fig. 2.7 will be discussed.

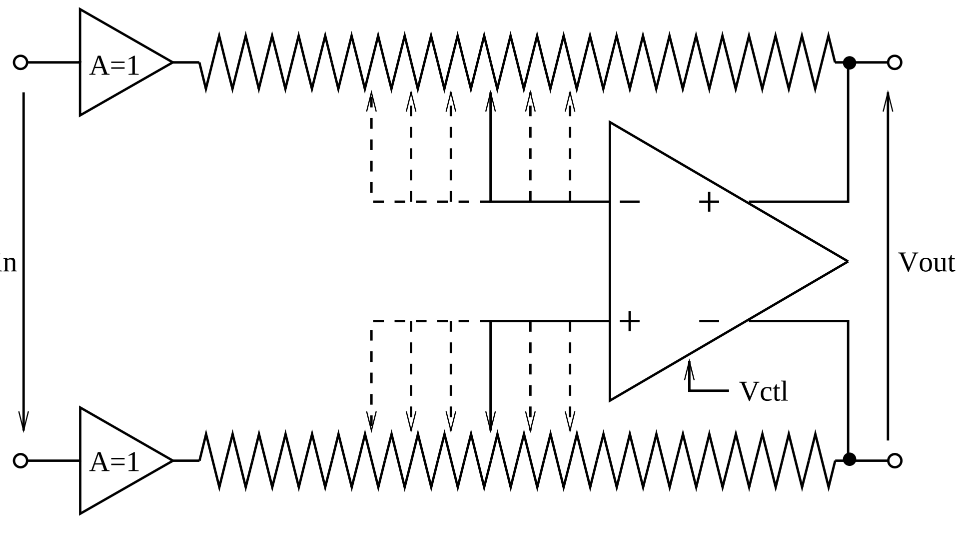


Figure 2.7 Overall architecture of the new VGA.

2.3 Multi-Input-Stage

The multi-input-stage, as shown in Fig. 2.6, comprises a series of G_{m} cells connected in parallel. Each cell consists of a diffpair and a current mirror. The schematic for each G_{m} cell is shown in Fig. 2.8. The magnitude of the transconductance of each G_{m} cell is controlled by a bias current signal $I_{\mathrm{ctl},i}$ which is a function of the control voltage V_{ctl}. The outputs of the G_{m} cells are connected together and fed to the telescopic I/V converter.

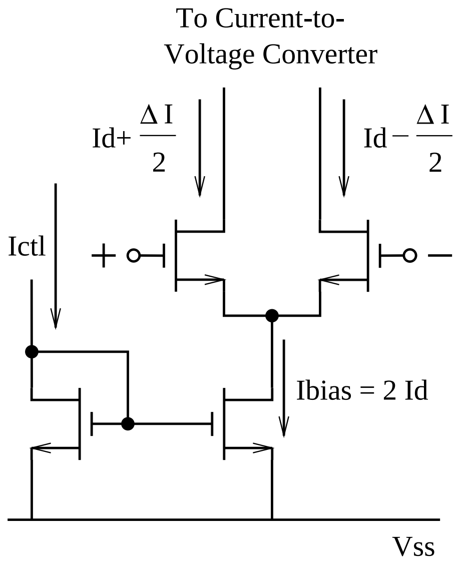


Figure 2.8 Transconductance cell.

Although the currents from N separate G_{m} cells contribute to the total signal current, not all G_{m} cells will be switched on simultaneously. Only consecutive pairs will be operating; all other cells will be switched off by shutting off their bias currents. In general, if there are N G_{m} cells, the overall current entering the telescopic I/V converter can be expressed using vector notation. Making the following definitions,

$g_{\mathrm{m},i}(V_{\mathrm{ctl}})$ = the transconductance of i^{th} cell which is a function of V_{ctl},
V_{ctl} = the control voltage, and
$\Delta V_{\mathrm{in},i}$ = the differential input voltage of i^{th} cell.

The overall transconductance ($\overrightarrow{G_{\mathrm{m}}}$) can be written as

$$\overrightarrow{G_{\mathrm{m}}} = \begin{bmatrix} g_{\mathrm{m},1}(V_{\mathrm{ctl}}) & g_{\mathrm{m},2}(V_{\mathrm{ctl}}) & \cdots & g_{\mathrm{m},n}(V_{\mathrm{ctl}}) \end{bmatrix}. \tag{2.1}$$

The output current (ΔI) is therefore given by

$$\Delta I = \overrightarrow{G_{\mathrm{m}}} \cdot \overrightarrow{\Delta V_{\mathrm{in}}}, \tag{2.2}$$

$$\Delta I = \begin{bmatrix} g_{\mathrm{m},1}(V_{\mathrm{ctl}}) & g_{\mathrm{m},2}(V_{\mathrm{ctl}}) & \cdots & g_{\mathrm{m},n}(V_{\mathrm{ctl}}) \end{bmatrix} \cdot \begin{bmatrix} \Delta V_{\mathrm{in},1} \\ \Delta V_{\mathrm{in},2} \\ \vdots \\ \Delta V_{\mathrm{in},n} \end{bmatrix}. \tag{2.3}$$

In this multi-input-stage, the transconductance of each cell ($g_{m,i}$) can be changed by the control voltage (V_{ctl}), which must produce necessary bias currents to properly control all G_m cells. The details of the control circuit and the method used for converting a voltage signal to current control signals will be elaborated upon further in Section 2.9.

2.4 Telescopic-Cascode Stage

Following the multi-input-stage, the output currents will be collected at the drains of each G_m cell and converted to voltages using a high-gain telescopic-cascode stage of Fig. 2.9. Although this stage has a limited output voltage swing, an output stage, with additional gain, will follow the telescopic-cascode stage so that this limited swing will not limit the output swing of the overall VGA.

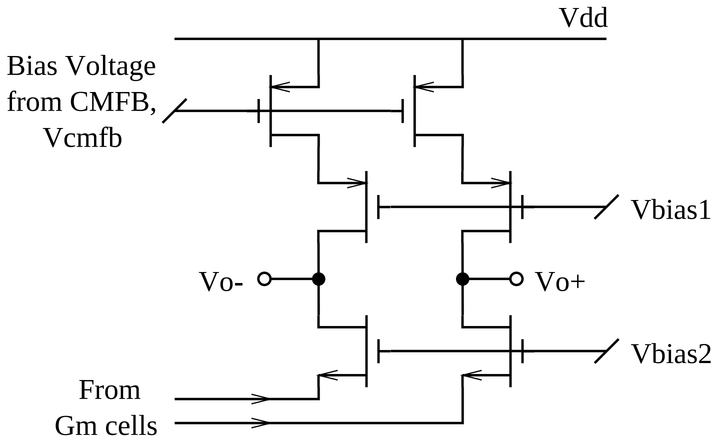


Figure 2.9 Modified telescopic-cascode stage.

2.5 Differential Output Stage

The telescopic-cascode stage has a limited output voltage swing and a high output impedance and therefore cannot be used to directly drive the low resistive load at output. An output stage with DC gain and large current handling capabilities is necessary.

Class-A vs. Class-AB Either a Class-A or Class-AB output stage can deliver sufficient signal current to the load. Each has advantages and disadvantages.

Class-A output stage There exists an output device that is always conducting a bias current larger than the maximum signal, therefore power delivering efficiency is poor. For CMOS technology, a Class-A stage can be constructed from a source follower configuration with a common-source bias current stage, or from a common-source input stage. [6, Fig. 6.10a and Fig. 6.26a]. The first has a low output impedance, but has a DC offset and no DC gain; the latter has DC gain, but a high output impedance.

Class-AB output stage There exists a pair of complementary devices, which deliver current in response to the signal. They are biased at only a small quiescent current when there is no output signal, therefore, standby current dissipation is low and efficiency is increased. In the presence of a signal, if an output signal appears, the pair of complementary output devices will be switched on alternatively. However, complementary devices are active only half of the signal cycle and the *cross-over* between devices is a source of distortion. For CMOS processes, a Class-AB stage can be realized using a pair of source followers with two transistor-connected diodes used for the quiescent biasing [7, Fig. 5.29].

This VGA is designed to operate from a single 5-V power supply; the single-ended output range should be ±1-V with respect to the common-mode potential. Since a source follower requires a level shift of the gate-source voltage ($|V_{gs}| > |V_t| \approx$ 1-V), it is difficult to utilize this circuit configuration with a large-swing output stage. Furthermore, at high frequency operation (e.g., 10-MHz), nonlinearities due to the crossover distortion[1] is not acceptable for a simple Class-AB output stage, and we decided not to use it in this application. As a result, a Class-A common-source output stage as in Fig. 2.10 was selected. Although the power efficiency is poor, the trade-off is that we are able to obtain a large voltage-swing with low harmonic-distortion. In the fully-differential implementation we use two output stages working in anti-parallel. A detailed analysis of the distortion of the output stage will be given in Section 3.1.4 and Section 3.2.

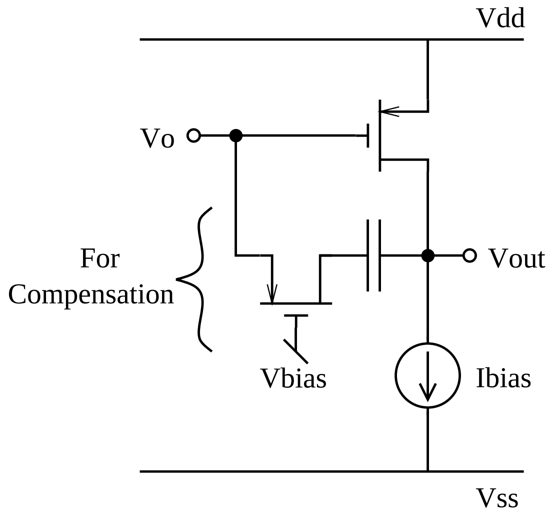


Figure 2.10 Output stage.

[1] Here, the crossover distortion relates not only to the dead-band in the transfer characteristic at the actual cross-over, but also to the asymmetry in conduction from one device to another.

2.6 Feedback

It is important for the gain of a VGA to be changed monotonically with respect to the control signal. Otherwise, it is very difficult to design a stable Automatic Gain Control (AGC) circuit that can adjust and lock the gain to maintain a fixed output level. In this implementation, the gain is adjusted by electronically sliding the virtual ground (pivot-point) along the feedback resistor, which insures monotonic gain variations. The feedback resistors for the positive and negative signal paths are partitioned into three parts as shown in Fig. 2.11.

R — The middle part of the resistor. The inputs of the multi-input-stage will be tapped here. Depending on the pivot location, it can be further partitioned into R_1 and R_2, where $R = R_1 + R_2$.

R_{in} — The initial part of the resistor, which is always in the input path and never in the feedback path.

R_{out} — The final part of the resistor. This portion is always in the feedback path and never in the input path.

The total resistance, $R_{\text{total}} = R_{\text{in}} + R + R_{\text{in}}$, is a constant.

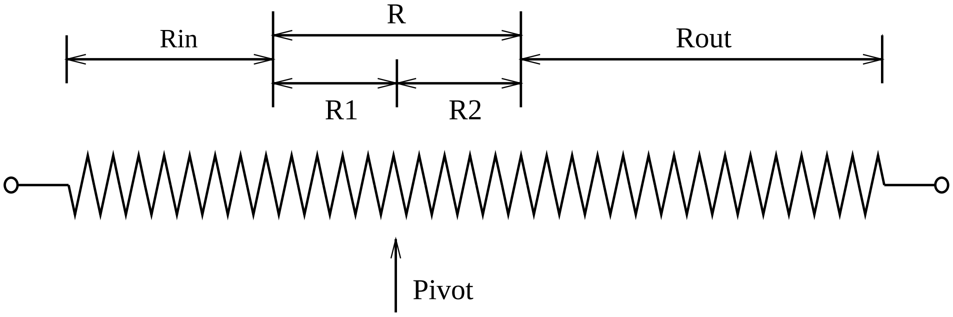


Figure 2.11 Partitioning of a feedback resistor.

Gain vs. Control Voltage If the open-loop gain of the opamp is assumed to be infinite, the closed-loop voltage gain (A_v) is found to be

$$A_v = \frac{R_{\text{out}} + R_2}{R_{\text{in}} + R_1}, \tag{2.4}$$

$$A_{v,\text{max}} = \frac{R_{\text{out}} + R}{R_{\text{in}}}, \tag{2.5}$$

$$A_{v,\text{min}} = \frac{R_{\text{out}}}{R_{\text{in}} + R}. \tag{2.6}$$

Referring to (2.4), it can be seen that the closed-loop gain displays a nonlinear characteristic, when the taps on the feedback resistor are linearly spaced. The rate of change of the gain is low when the gain is low, and high when the gain is high, which is similar to an exponential characteristic. This characteristic is plotted in Fig. 2.12. If a linear gain characteristic is needed, the spacing between taps on the feedback resistor can be chosen to generate an overall linear gain variation vs. control-voltage characteristic. However, for most AGCs only monotonicity is required. So a linear tap spacing was used in this design, which results in the mildly nonlinear gain characteristic of Fig. 2.12.

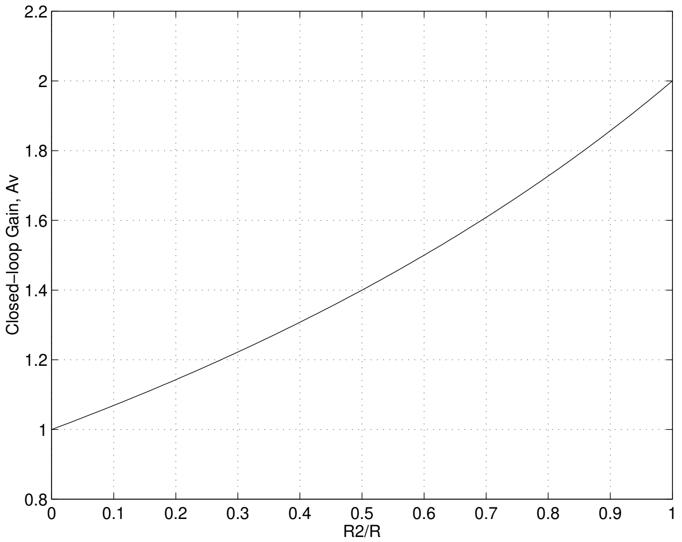


Figure 2.12 Plot of the closed-loop gain A_v when $R_{\text{in}} = 2$, $R = 1$, $R_{\text{out}} = 3$, from (2.4).

2.7 Common-Mode Feedback (CMFB) Circuit

The VGA is fully-differential and requires a Common-Mode Feedback (CMFB) circuit to insure that the outputs remain balanced around a fixed common-mode potential. It is typical in CMFB circuits to monitor the outputs ($V_{\text{out}+}$ and $V_{\text{out}-}$) directly. However, these signals may swing within the full output range. In this circuit, since feedback resistors are used, the scaled-down version of the output signals can be conveniently probed in the middle of the feedback resistors. This is illustrated graphically in Fig. 2.13. The CMFB circuit can take advantage of this reduced differential swing so that the CMFB circuit will operate in a more linear region. This will help to eliminate differential-mode induced common-mode variations. When we consider the location to tap the feedback resistor which feeds to the CMFB circuit, we use the criteria that the swing should be identical for the highest and the lowest closed-loop gain. This criteria is illustrated in Fig. 2.14.

To stabilize the common-mode output level, the CMFB circuit of Fig. 2.15 is used. The circuit monitors the differential signal between nodes ($V_{\text{ref},1}$ and $V_{\text{ref},2}$). If these signals show perfect odd symmetry about the common-mode reference, V_{com}, the

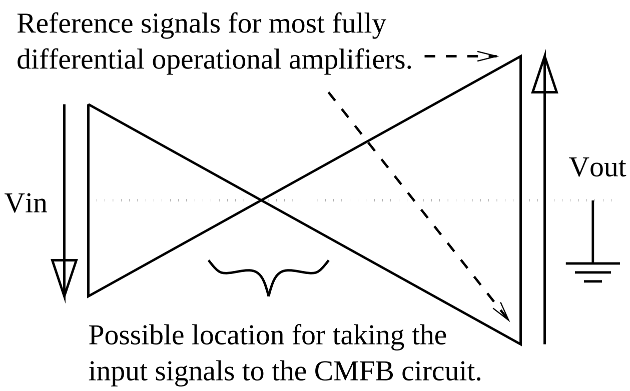


Figure 2.13 Input signals to the CMFB circuit.

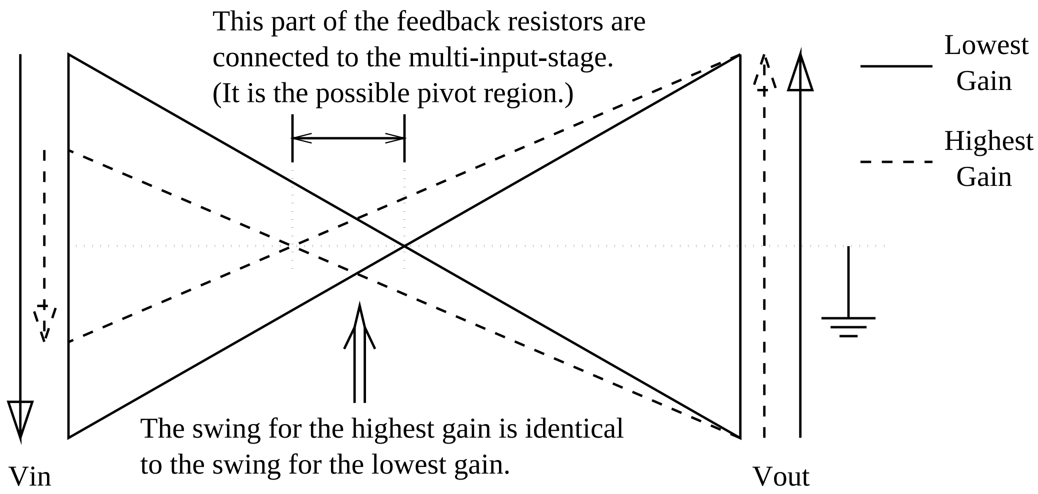


Figure 2.14 Optimal location (⇑) for probing input signals to the CMFB circuit.

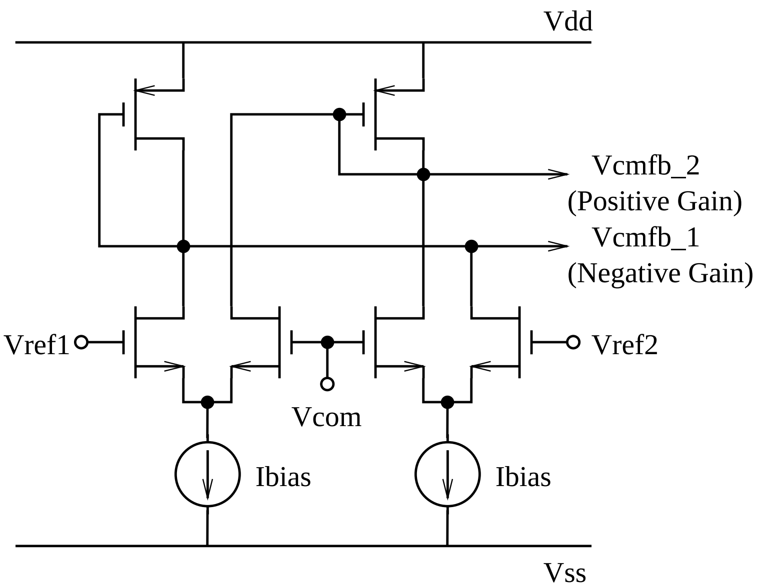


Figure 2.15 CMFB circuit.

cross-coupled connection will insure that the voltage at $V_{\text{cmfb},1}$ and $V_{\text{cmfb},2}$ remain at a fixed DC level. However, if the common-mode signal, $(V_{\text{ref},1} + V_{\text{ref},2})/2$, is greater than V_{com}, more current will be diverted to $V_{\text{cmfb},1}$ and the voltage will drop, while less current will be drawn from $V_{\text{cmfb},2}$ and the voltage will increase.

The negative output of the CMFB circuit, $V_{\text{cmfb},1}$, is used to bias the top p-channel transistors in the telescopic-cascode stage of Fig. 2.9. When the output common-mode level rises, the bias voltage on the gates of the p-channel current sources will reduce, which will cause the common-mode output voltage to reduce, too. This results in a negative feedback loop, which will force the common-mode output voltages to be balanced at the proper level. Care must be taken, however, to make sure that this negative feedback loop is stable and does not oscillate.

2.8 Input Voltage Buffer

Because this VGA uses resistive feedback, the input impedance to the VGA would be determined by the input resistor if a voltage buffer were not used. In order to prevent drawing current from the input signal, dual input voltage buffers are inserted between the input terminals and the feedback resistors. The required performance specifications of the input buffer is tighter than that of the VGA such that the overall circuit performance will not be adversely affected. The specifications of the voltage buffer are listed in Table 2.1.

Table 2.1 Specifications for the input voltage buffer.

Parameter	Expected Performance
Gain	1 (0-dB)
Output Range	4-V_{pp} differential; $\pm$1-V_{p} single-ended
Frequency Range	DC to 10-MHz
Linearity (THD)	Below -65-dB for frequencies up to 10-MHz
Signal-to-Noise (SNR)	$\gg$ 60-dB

The voltage buffer used in this design is a simple opamp with unity feedback, as shown in Fig. 2.16. This circuit is a two-stage opamp which has a Class-A common-source output stage. Its THD is mainly determined by the performance of the output stage. Its common-mode operating range is limited by the input stage. The common-mode range could be expanded by using a rail-to-rail design [8, 9], however, the performance of the simple two-stage opamp is more than adequate for our needs.

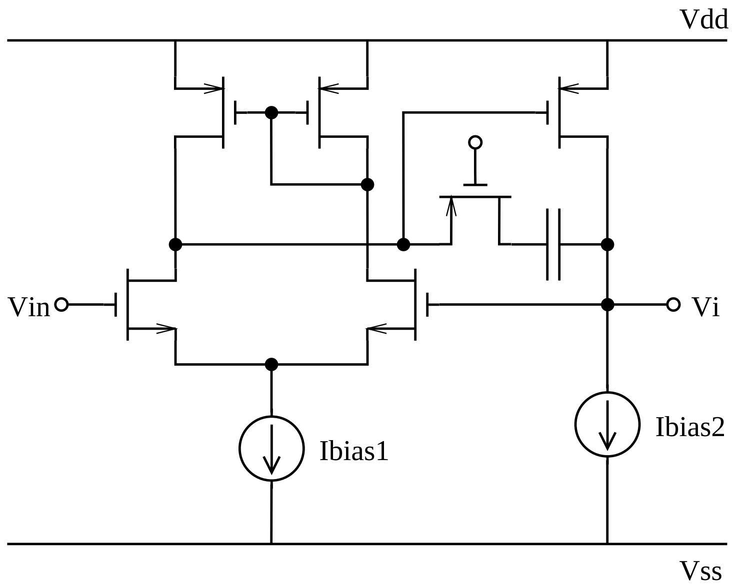


Figure 2.16 Schematic for the input voltage buffer.

2.9 Gain Control Circuit

In the previous sections the sub-circuits needed for the signal-path of the VGA have been described. In this final section, we will describe the gain control circuit, which produces the proper bias currents needed by the $G_{\rm m}$ cells in the multi-input-stage. The following guidelines must be considered in the design of the control circuit.

- For each $G_{\rm m}$ cell, its ON or OFF state is controlled by its reference bias current. This means that the given control voltage signal should be converted into current signals which control the actual switching. We design the bias current control to provide smooth switching between adjacent cells.

- Since interpolation is performed only between consecutive taps, only two $G_{\rm m}$ cells can be switched ON at any given time.

- To illuminate excess distortion it is necessary to turn OFF each $G_{\rm m}$ cell completely, when it is not being used.

The general shape of the bias currents needed, as a function of the control voltage is shown in Fig. 2.17. Each current signal is approximately a bell-shaped transfer curve with respect to $V_{\rm ctl}$. When the bias current is non-zero, its corresponding $G_{\rm m}$ cell will be operating in the active region, and when the current equals zero, the $G_{\rm m}$ cell will be OFF.

To generate the current signals needed, we can take advantage of the transfer characteristics of a diffpair. The V-I transfer curve of a diffpair is plotted in Fig. 2.18b. A bell-shaped V-I transfer curve can be obtained if the shifted versions of I_1 and I_2 are subtracted, for example, $I_1(\Delta V + V) - I_1(\Delta V - V)$ or $I_2(\Delta V - V) - I_2(\Delta V + V)$.

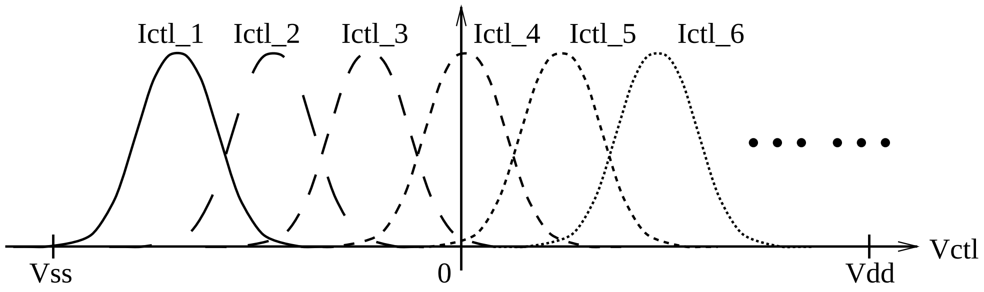


Figure 2.17 Expected current signals from control circuit.

The resulting current is shown in Fig. 2.18c. This is precisely the desired current pulse needed to gradually switch the input G_{m} cells ON and OFF.

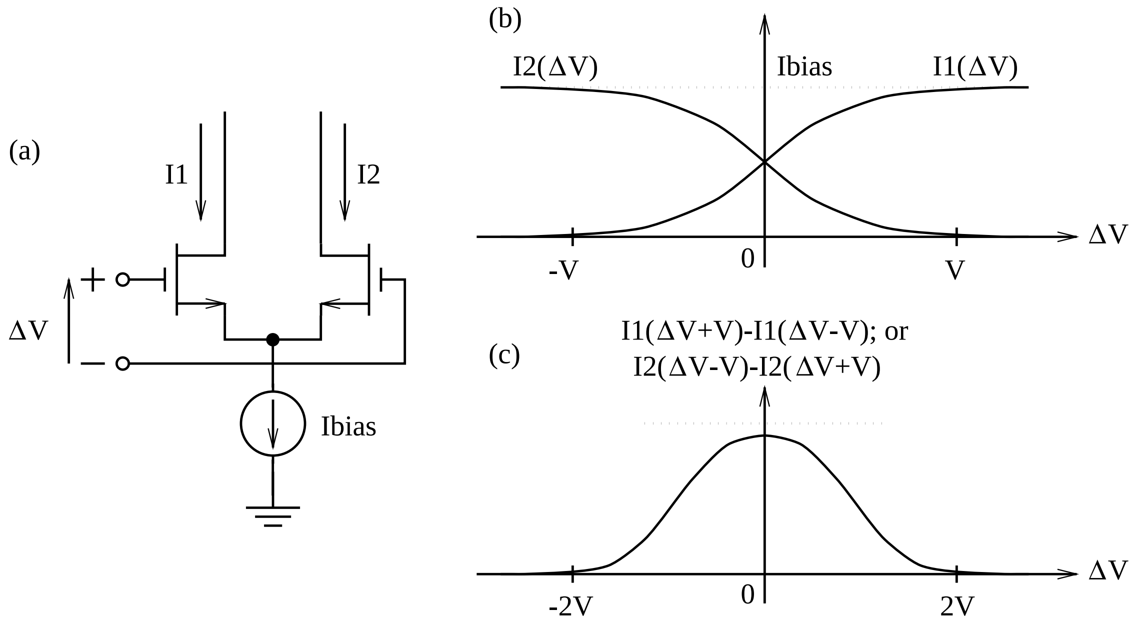


Figure 2.18 The characteristic of a differential pair.

We have shown how to generate one of the desired current pulses, it remains to generate all the pulses with the proper relative spacing. This can be accomplished using the circuit of Fig. 2.19. This circuit is a series of diffpairs; all the positive input terminals are taken from successive taps of a reference resistor, while the negative input terminals are connected to a common control voltage, V_{ctl}. Under these connections, the control signal is compared to different reference potentials and shifted versions of I_1 and I_2 as in Fig. 2.18b can be obtained.

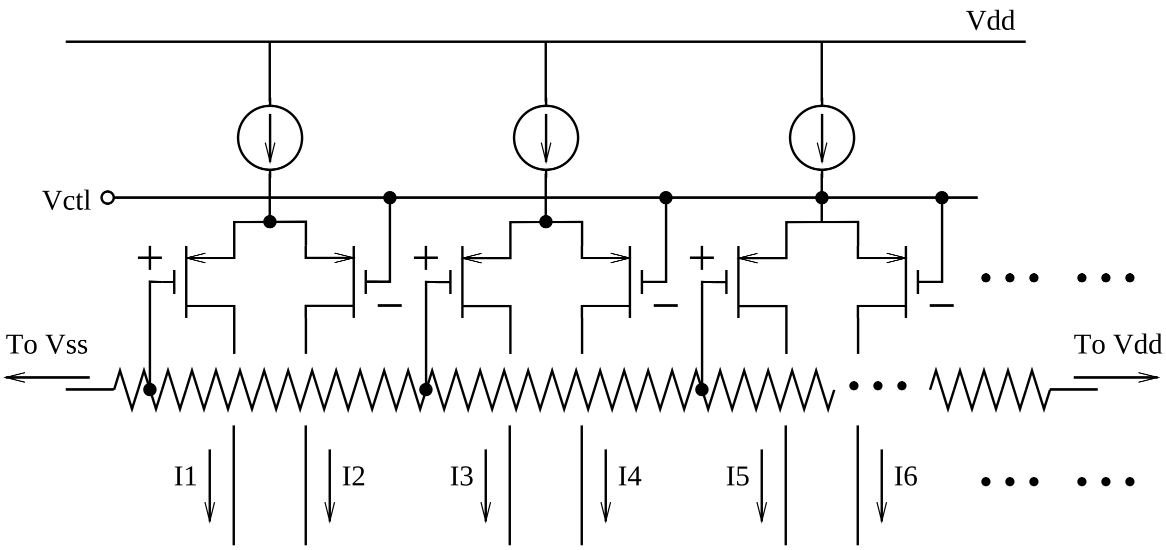


Figure 2.19 The core part of the control circuit.

Generation of Bias Control from Primitive Currents The control current signals for each G_m cell can be determined as follows.

$$\begin{aligned} I_{\mathrm{ctl},1} &= I_2 - I_4, \\ I_{\mathrm{ctl},2} &= I_5 - I_3, \\ I_{\mathrm{ctl},3} &= I_6 - I_8, \\ &\vdots \end{aligned}$$

$$I_{\mathrm{ctl},i} = \begin{cases} I_{2i} - I_{2i+2} & \text{if } i \text{ is odd,} \\ I_{2i+1} - I_{2i-1} & \text{if } i \text{ is even.} \end{cases} \tag{2.7}$$

Current Subtraction Circuit The current subtraction can be achieved in several ways. The circuit that was used here is a simple current mirror, as shown in Fig. 2.20. Two current signals are applied to the input and the output of the mirror. Because of the duplication of the input current on the output side, a difference current ($I_2 - I_4$) will be generated.

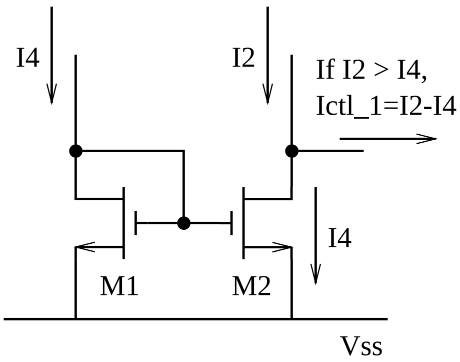


Figure 2.20 A simple current mirror for current differencing.

Offset Current to Insure Bias Shut-Off It is important that the G_m cell is OFF when it is supposed to be OFF. If we used a simple subtraction method, random offsets would result in a residual current in some of the G_m cells, which would leave them slightly ON at the wrong moment. This offset is incorporated into the design by making the maximum value of I_2 smaller than the maximum value of I_4. This will force the voltage on the drain of M_2 low, which will shut-off the bias current being fed to the G_m cell. In general, it can be achieved by assigning the current ratio for biasing the odd and the even diffpairs to be $1 : 1.1$.

Trade-Off in Design The spread, or width, of the bell-shaped V-I transfer curve is related to the aspect ratio (W/L) of the diffpairs. If the aspect ratio is higher, the width of the transfer curve is narrower; and vice versa. Since the transfer curves of Fig. 2.17 should fit within the desired control voltage range, in this case between V_{dd} and V_{ss}, the number of control signals that can be accommodated is limited. (In this design, 10 taps were used.) If more G_m cells are needed, a higher aspect ratio should be assigned to each diffpair in the control circuit. This has the disadvantage of increasing the die area of the control circuit. If this becomes unwieldy, a nested control circuit can be designed based on a nested folding structure [10].

Summary In this chapter, the architecture and the building blocks of this interpolating VGA have been presented. The architecture was chosen because of its superior linearity. In the next chapter we will analyze the nonlinearity of the circuit in detail and optimize the circuit for maximum dynamic-range.

Chapter 3

Design Considerations: Optimization for Maximum Dynamic-Range

In the proposed VGA, the output voltage swing is large (4-V_{pp} differential); therefore the dynamic-range is limited by distortion rather than noise. The distortion is primarily caused by the multi-input-stage and the output stage. Each of these distortion sources will be discussed separately. Before this discussion, we will review some important basics of distortion theory in quasi-linear circuits. This review follows Willingham [11]. The key aspects as they relate to this design are repeated here for the readers convenience.

3.1 Nonlinear Distortion in Quasi-Linear Circuits

In the broadest sense, distortion can be considered to be any change in signal shape caused by the circuit operation, except simple scaling. This could include cross-coupling, power-supply noise, nonlinear group-delay, and a long list of other ailments. This differs from noise, in that distortion is predictable. However, this definition of distortion is too broad for our purpose. We will be only concerned with signal distortion which arises from circuit nonlinearities. In any practical circuit, we cannot find components with perfectly linear characteristics. We only approximate linear behavior, and we use measures, such as, Total Harmonic Distortion (THD) and Intermodulation Distortion (IMD) to indicate how well our quasi-linear circuit approximates ideal linear behavior.

3.1.1 Basic Theory

In general, the transfer function of a single-input-single-output (SISO) quasi-linear system, such as the one depicted in Fig. 3.1 can be expressed as a power series expansion around its operating point, (x_0, y_0),

$$\begin{aligned} y - y_0 = f(x) &= a_1(x-x_0) + a_2(x-x_0)^2 + a_3(x-x_0)^3 + a_4(x-x_0)^4 \\ &\quad + a_5(x-x_0)^5 + \cdots . \end{aligned} \tag{3.1}$$

The first coefficient, a_1, is the familiar incremental *linear* gain of the system. Other

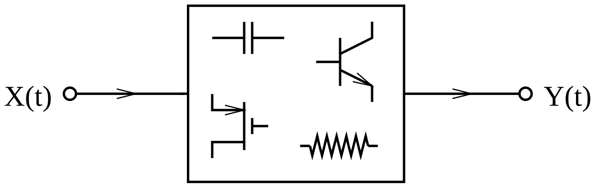


Figure 3.1 All transistor circuit are nonlinear systems.

higher-order coefficients, a_2, a_3, a_4, a_5, ..., characterize the *nonlinear* distortion. If a simple cosine signal ($x = x_0 + \rho_x \cos\omega t$) is applied to the system, the output will be

$$\begin{aligned} y - y_0 &= a_1\rho_x \cos\omega t + a_2(\rho_x\cos\omega t)^2 + a_3(\rho_x\cos\omega t)^3 + a_4(\rho_x\cos\omega t)^4 \\ &\quad + a_5(\rho_x\cos\omega t)^5 + \cdots , \\ y - y_0 &= (\frac{1}{2}a_2\rho_x^2 + \frac{3}{8}a_4\rho_x^4 + \cdots) \\ &\quad + (a_1 + \frac{3}{4}a_3\rho_x^2 + \frac{5}{8}a_5\rho_x^4 + \cdots)\rho_x\cos\omega t \\ &\quad + (\frac{1}{2}a_2 + \frac{1}{2}a_4\rho_x^2 + \cdots)\rho_x^2\cos 2\omega t \\ &\quad + (\frac{1}{4}a_3 + \frac{5}{16}a_5\rho_x^2 + \cdots)\rho_x^3\cos 3\omega t \\ &\quad + (\frac{1}{8}a_4 + \cdots)\rho_x^4\cos 4\omega t \\ &\quad + (\frac{1}{16}a_5 + \cdots)\rho_x^5\cos 5\omega t \\ &\quad + \cdots . \end{aligned} \tag{3.2}$$

The transfer function contains harmonics which are integer multiples of the input (*fundamental*) frequency. The harmonic distortion of the system is defined to be the amplitude ratio of the harmonic content to the fundamental frequency,

$$\mathrm{HD}_i \triangleq \frac{\text{Amplitude of } i^{\text{th}} \text{ harmonic}}{\text{Amplitude of the fundamental}}, \tag{3.3}$$

where i is greater than 1. The total harmonic distortion is defined to be

$$\mathrm{THD} \triangleq \sqrt{\sum_{i=2}^{\infty} \mathrm{HD}_i^2}. \tag{3.4}$$

We can take the simple sum of the powers as in (3.4) and ignore all the cross-products because all the harmonics are mutually orthogonal.[1]

For quasi-linear systems, the power series terms typically converge rapidly and the coefficients above a_3 can be ignored. Further, for small inputs such that the third-order term is significantly smaller than the linear term, $(3/4)a_3\rho_x^2 \ll a_1$, the system series-expansion function (3.2) can be approximated to be

$$\begin{aligned} y - y_0 &= \frac{1}{2}a_2\rho_x^2 + (a_1 + \frac{3}{4}a_3\rho_x^2)\rho_x \cos\omega t + \frac{1}{2}a_2\rho_x^2 \cos 2\omega t \\ &\quad + \frac{1}{4}a_3\rho_x^3 \cos 3\omega t + \cdots, \\ &\approx a_1\rho_x \cos\omega t + \frac{1}{2}a_2\rho_x^2 \cos 2\omega t + \frac{1}{4}a_3\rho_x^3 \cos 3\omega t. \end{aligned} \tag{3.5}$$

When the THD is dominated by second- and third-order harmonic distortion, the circuit is said to be in its *mild distortion* regime.

Fully-Differential Design: Cancellation of Even-Order Distortion For a fully differential system, if a balanced cosine signal, $x_1 = x_0 + \rho_x \cos\omega t$ and $x_2 = x_0 - \rho_x \cos\omega t$, is applied to the system, the output can be written as follows,

$$y' = f(x_1) - f(x_2). \tag{3.6}$$

Recalling (3.1), the power series terms can be separated as,

$$\begin{aligned} y' &= \sum_{i=1}^{\infty} a_i(\rho_x \cos\omega t)^i - \sum_{i=1}^{\infty} a_i(-\rho_x \cos\omega t)^i, \\ &= 2\sum_{j=1}^{\infty} a_{2j-1}(\rho_x \cos\omega t)^{2j-1}. \end{aligned} \tag{3.7}$$

Due to even symmetry, all the even coefficients in (3.7) sum to zero. Referring to (3.2), the coefficient of all the even harmonics are composed of only the even coefficients in the power series expansion of (3.1). Therefore, in a fully-differential system, all even-order harmonics will vanish.

It should be remembered, however, that this property occurs due to cancellation, it is only as reliable as the matching accuracy of the positive and negative signal paths.

---

[1] When two real periodic functions, $v(t)$ and $w(t)$, are orthogonal to each other, their dot-product is equal to zero, $(u, v) = 0$, where

$$(u, v) \triangleq \int_t^{t+T} v(t)w(t)dt = 0.$$

T is the integer multiple of the periods of both $v(t)$ and $w(t)$. However, for any two harmonics, they will satisfy the above condition.

$$(u_m, u_n) = \int_t^{t+\frac{1}{f}} \cos(m \cdot 2\pi ft)\cos(n \cdot 2\pi ft)dt = \begin{cases} 0 & \text{if } m \neq n, \\ \frac{1}{2f} & \text{if } m = n, \end{cases}$$

where m and n are positive integers. Because of the orthogonality, all the cross-product terms will vanish in the integral, and the THD is calculated by the root-mean-square method.

Improvements Relative to a Single-Ended Circuit Since the even harmonics disappear, and the relative amplitude of the odd harmonics, with respect to fundamental, remains fixed, the THD of the fully differential system will be improved in comparison to a single-ended circuit. This improvement is very significant, since most single-ended systems have distortion dominated by the second-order harmonic, and in general, the lower-order nonlinearities dominate, $a_2 > a_3 > a_4 > a_5 > \ldots$.

3.1.2 Distortion in a CMOS Differential Pair: Example Calculation

To see how to use the basic theory to optimize circuit performance, let's consider a simple CMOS differential pair (diffpair), shown in Fig. 3.2. By using Kirchoff's

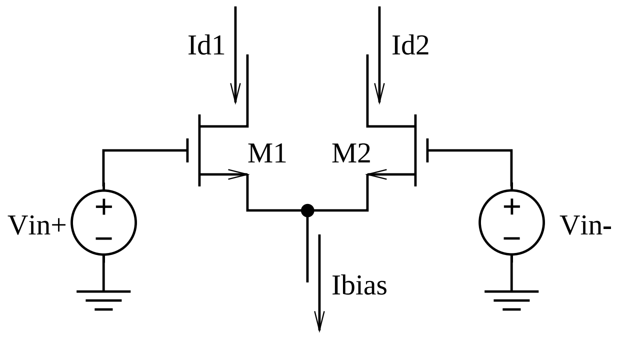


Figure 3.2 CMOS Differential Pair.

Voltage Law (KVL) around the gate-source loops of the coupled-pair,

$$\begin{aligned} V_{\text{in},+} - V_{\text{gs},1} + V_{\text{gs},2} - V_{\text{in},-} &= 0, \\ V_{\text{in},+} - V_{\text{in},-} &= V_{\text{gs},1} - V_{\text{gs},2}, \\ \Delta V_{\text{in}} &= \sqrt{\frac{2I_{\text{d1}}}{K}} - \sqrt{\frac{2I_{\text{d2}}}{K}}, \\ &= \sqrt{\frac{2\overline{I_{\text{d}}}}{K}} \left(\sqrt{1 + \frac{\Delta I}{2\overline{I_{\text{d}}}}} - \sqrt{1 - \frac{\Delta I}{2\overline{I_{\text{d}}}}} \right), \end{aligned} \tag{3.8}$$

we can write the input differential voltage, ΔV_{in}, as a function of the output differential current, ΔI. It is helpful in the analysis to factor out bias quantities and write the signals as a normalized deviation away from the bias point. We can use the following differential notation,

$$\begin{aligned} \Delta V_{\text{in}} &= V_{\text{in},+} - V_{\text{in},-}, \\ I_{\text{d}} &= (K/2)(V_{\text{gs}} - V_{\text{t}})^2, \\ K &= \mu_{\text{n}} C_{\text{ox}}(W/L), \\ \overline{I_{\text{d}}} &= (I_{\text{d1}} + I_{\text{d2}})/2 = I_{\text{bias}}/2, \\ \Delta I &= I_{\text{d1}} - I_{\text{d2}}, \\ I_{\text{d1}} &= \overline{I_{\text{d}}} + \Delta I/2, \\ I_{\text{d2}} &= \overline{I_{\text{d}}} - \Delta I/2. \end{aligned}$$

We can then define a normalized effective-gate-source voltage as the difference between the normalized gate-source voltage ($\overline{V_{\mathrm{gs}}}$) and the threshold voltage (V_{t}),

$$\overline{V_{\mathrm{gt}}} = \overline{V_{\mathrm{gs}}} - V_{\mathrm{t}} = \sqrt{\frac{2\overline{I_{\mathrm{d}}}}{K}} = \sqrt{\frac{I_{\mathrm{bias}}}{K}}. \tag{3.9}$$

The V/I relationship of (3.8) can then be rewritten as

$$\frac{\Delta V_{\mathrm{in}}}{\overline{V_{\mathrm{gt}}}} = \left(\sqrt{1+\frac{\Delta I}{2\overline{I_{\mathrm{d}}}}} - \sqrt{1-\frac{\Delta I}{2\overline{I_{\mathrm{d}}}}}\right). \tag{3.10}$$

To determine the harmonic distortion we can expand the square-root terms into power series,[2]

$$\frac{\Delta V_{\mathrm{in}}}{\overline{V_{\mathrm{gt}}}} = \frac{\Delta I}{2\overline{I_{\mathrm{d}}}} + \frac{1}{8}\left(\frac{\Delta I}{2\overline{I_{\mathrm{d}}}}\right)^3 + \frac{7}{128}\left(\frac{\Delta I}{2\overline{I_{\mathrm{d}}}}\right)^5 + \cdots. \tag{3.11}$$

Anti-Causal Analysis and Harmonic Predistortion The result of (3.11) is not in a suitable form for our purpose. To analyze the harmonic distortion of the system we desire the output (ΔI) to be written as an explicit function of the input (ΔV_{in}). However, it is often more convenient in the analysis to write the input as an explicit expression of the output. This anti-causal expression is generally simpler, especially in the presence of feedback. Willingham explains [11, Chapter 4], how harmonic *pre*-distortion (the amount of distortion needed to be added to the input to generate a perfectly linear output) is related to harmonic distortion (the distortion at the output generated when a perfectly linear input is applied.) Willingham's results are summarized in Appendix A, which we will use here to find the THD of the diffpair. From (A.2), (A.3) and (A.7), (3.11) can be rewritten as,

$$\frac{\Delta I}{2\overline{I_{\mathrm{d}}}} = \frac{\Delta V_{\mathrm{in}}}{\overline{V_{\mathrm{gt}}}} - \frac{1}{8}\left(\frac{\Delta V_{\mathrm{in}}}{\overline{V_{\mathrm{gt}}}}\right)^3 - \frac{1}{128}\left(\frac{\Delta V_{\mathrm{in}}}{\overline{V_{\mathrm{gt}}}}\right)^5 + \cdots. \tag{3.12}$$

This expression inverts the previous series and now expresses the normalized output current as a function of the normalized input voltage. (3.12) implicitly assumes that both transistors are operating in the linear-operating region (saturation region). (If M_1 and M_2 are operating in the triode region, (3.8)–(3.12) cannot be used to describe the relationship between ΔI and ΔV_{in}.)

The distortion of the diffpair contains only odd-harmonics. Since, higher harmonics contribute significantly less distortion than lower harmonics for this circuit,[3] we can

[2] By mean of the Taylor series expansion,

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \frac{7}{256}x^5 + \cdots.$$

[3] Referring to (3.2), if the coefficients beyond a_5 are ignored and all the terms inside the brackets with ρ_x^2, ρ_x^4, ρ_x^6, ... can also be neglected, the function can then be rewritten as

$$y - y_0 \approx a_1\rho_x \cos\omega t + \frac{1}{2}a_2\rho_x^2\cos 2\omega t + \frac{1}{4}a_3\rho_x^3\cos 3\omega t + \frac{1}{8}a_4\rho_x^4\cos 4\omega t + \frac{1}{16}a_5\rho_x^5\cos 5\omega t.$$

restrict our attention to the 3rd- and 5th-order terms, which are as follows,

$$\begin{aligned}
\mathrm{HD}_3 &= \frac{1}{4}\frac{a_3}{a_1}\rho_x^2 = -\frac{1}{32}\rho_x^2, && (3.13)\\
\mathrm{HD}_5 &= \frac{1}{16}\frac{a_5}{a_1}\rho_x^4 = -\frac{1}{2048}\rho_x^4,\\
&= -\frac{1}{2}\mathrm{HD}_3^2. && (3.14)
\end{aligned}$$

The THD of the diffpair is therefore given approximately by

$$\begin{aligned}
\mathrm{THD} &\approx \sqrt{\mathrm{HD}_3^2 + \mathrm{HD}_5^2},\\
&\approx \sqrt{\left(\frac{1}{4}\frac{a_3}{a_1}\rho_x^2\right)^2 + \left(\frac{1}{16}\frac{a_5}{a_1}\rho_x^4\right)^2}, && (3.15)
\end{aligned}$$

which can be simplified to

$$\begin{aligned}
\mathrm{THD} &\approx \mathrm{HD}_3\sqrt{1 + \frac{1}{4}\mathrm{HD}_3^2},\\
&\approx \frac{1}{32}\rho_x^2\sqrt{1 + \frac{1}{4}\left(\frac{1}{32}\rho_x^2\right)^2}. && (3.16)
\end{aligned}$$

Modulation Index: Limits of Analysis Recall that ρ_x is the normalized input amplitude, also known as the *modulation index*. If the differential input is given by $\Delta V_x \cos(\omega t)$, then

$$\rho_x \triangleq \frac{|\Delta V_x|}{\overline{V_{\mathrm{gt}}}} = \frac{|\Delta V_x|}{\sqrt{\frac{I_{\mathrm{bias}}}{\mu_{\mathrm{n}} C_{\mathrm{ox}} (W/L)}}}. \qquad (3.17)$$

The above analysis is valid for $|\rho_x| \leq \sqrt{2}$. When $|\rho_x| > \sqrt{2}$, the circuit enters a strongly nonlinear region and the equations used to derive the above results don't account for this situation.

Numerical Example We can calculate the distortion of the diffpair for a typical sized transistor with $(W/L) = 40$ and $\mu_{\mathrm{n}} C_{\mathrm{ox}} = 163$-$\mu$A/V^2. For a bias tail-current of 200-μA, $\overline{V_{\mathrm{gt}}}$=175-mV in the balanced state when 100-μA flows in each transistor. The distortion of the differential output current of the diffpair is given in Table 3.1 for an input signal of $\Delta V_x \cos(\omega t)$, at various values of ΔV_x. We can see that the THD is dominated by 3rd-order distortion. It is also interesting to note that the modulation index of the diffpair cannot exceed 15% in order to obtain better than −60-dB THD. The key point to stress in this analysis is that *the harmonic distortion is a strong function of the modulation index, ρ_x.*

In the following sections the results and methods discussed here will be used to analyze the distortion performance of the overall VGA. With the results of this distortion analysis, we will then be able to optimize the VGA for maximum dynamic-range.

Table 3.1 Example for the harmonic distortion in a differential pair.

ΔV_x	ρ_x	HD_3	HD_5	THD
25-mV	0.1429	−63.90-dB	−133.83-dB	−63.90-dB
50-mV	0.2857	−51.86-dB	−109.75-dB	−51.86-dB
75-mV	0.4286	−44.82-dB	−95.66-dB	−44.82-dB
100-mV	0.5714	−39.82-dB	−85.67-dB	−39.82-dB
125-mV	0.7143	−35.95-dB	−77.92-dB	−35.95-dB
150-mV	0.8571	−32.78-dB	−71.58-dB	−32.78-dB
175-mV	1.0000	−30.10-dB	−66.23-dB	−30.10-dB

3.1.3 Circuit Analysis — Multi-Input-Stage

The proposed VGA uses a multi-input-stage with 10 transconductance cells (G_{m} cells). However, in the ideal case, only two cells will be active at any given time. Therefore, we can analyze the effect of distortion by considering the simplified circuit of Fig. 3.3, which contains only two input stages. We will consider various cases where G_{m1} and G_{m2} are biased in different states.

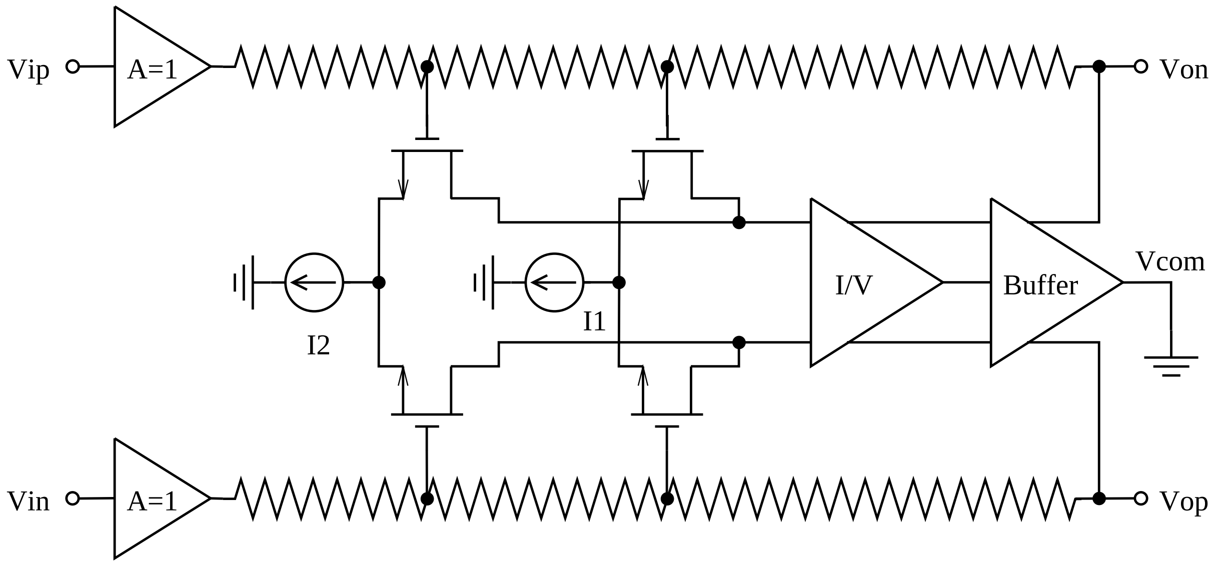


Figure 3.3 Simplified diagram of an interpolating VGA with only two G_{m} input-stages, used to analyze distortion.

Case 1: $\mathbf{G_{m1}}$ is ON, $\mathbf{G_{m2}}$ is OFF

We will first consider the case where G_{m1} is ON and G_{m2} is OFF. The VGA will operate like an ordinary operational amplifier (opamp). If we assume for the moment that the I/V converter and the output buffer are perfectly linear, the distortion at the output will be solely due to the diffpair. The following operating conditions will be used.

Tuning range The gain (A_v) is changed between 1.0 and 1.1.

Open-loop gain The open-loop gain of the opamp is 40 (32-dB); a value which is typical at frequencies of 5–10-MHz.

Voltage swing The output voltage is fixed to be 1-V_p single-ended or 4-V_{pp} differential; fixing the output voltage swing will mimic the operation of the VGA in a closed loop AGC system.

Bias current and Device sizes The total bias tail-current, I_{bias}, to share between the two G_m cells is 600-μA. Each transistor in the diffpairs is identical and has an aspect ratio of $(W/L) = 80$.

Processing data $\mu_n C_{ox} = 167\text{-}\mu\text{A/V}^2$ for a 0.8-μm CMOS process.

Given that the output swing is 1-V_p and that the open-loop gain of the VGA is 40, the input voltage to the active G_m cell must be $\pm$1-V/40, which is $\pm$25-mV single-ended, or $\pm$50-mV differential. To determine the distortion that this voltage swing will generate, we need to calculate the modulation index. The effective gate voltage when half the tail current flows in one of the diffpair transistors is given by

$$\overline{V_{gt}} = \sqrt{\frac{I_{bias}}{\mu_n C_{ox}(W/L)}} = \sqrt{\frac{600\text{-}\mu\text{A}}{167\text{-}\mu\text{A/V}^2(20)}} = 212\text{-mV}. \tag{3.18}$$

The modulation index from (3.17) is therefore

$$\rho_{x1} = \frac{50\text{-mV}}{212\text{-mV}} = 0.236. \tag{3.19}$$

The distortion of the diffpair can be determined directly from (3.13). Assuming that the I/V converter and the output buffers are linear, the VGA will have open-loop harmonics of

$$\mathrm{HD}_3 = \left|-\frac{1}{32}\rho_{x1}^2\right| = -55\text{-dB}, \tag{3.20}$$

$$\mathrm{HD}_5 = \left|-\frac{1}{2}\mathrm{HD}_3^2\right| = -116\text{-dB}, \tag{3.21}$$

$$\mathrm{THD} \approx \mathrm{HD}_3 = -55\text{-dB}. \tag{3.22}$$

Since we are using a feedback system, this open-loop distortion will be reduced by the loop-gain (return-ratio). For a closed-loop gain of unity, the distortion will be reduced by 32-dB, and by 26-dB for a closed-loop gain of two. Therefore,

$$\mathrm{THD}_{closed} = -55 - 32 = -87\text{-dB} \quad \text{for} \quad A_v = 1, \tag{3.23}$$

$$\mathrm{THD}_{closed} = -55 - 26 = -81\text{-dB} \quad \text{for} \quad A_v = 2. \tag{3.24}$$

This distortion performance exceeds our requirements by 21-dB. Since the gain appears to the third-power in the closed-loop third harmonic, this gives approximately (21/3) or 7-dB of margin in the open-loop gain. Therefore, the open-loop gain needs to remain above $32 - 7 = 25$-dB (17.78), for the device sizes and bias quantities used, such that a THD of better than -60-dB is maintained. This means the gain-bandwidth product of the opamp should be greater than ($2 \times 17.78 \times 10$-MHz = 355-MHz). This closed-loop distortion may not be obtainable, because other causes of distortion, not yet considered, could limit the performance.

Case 2: $\mathbf{G_{m1}}$ is OFF, $\mathbf{G_{m2}}$ is ON

The other extreme case where G_{m1} is OFF, and G_{m2} is ON is virtually identical to the previous case. Since the output voltage swing remains fixed, the input swing on the diffpairs must also be the same as before (± 50-mV). Therefore the modulation index and open-loop distortion figures will be the same in above. The closed-loop gain is now 1.1 instead of 1, so the closed-loop distortion will increase by 0.8-dB.

Before proceeding, It will be useful to calculate some quantities to aid in the upcoming analysis. The effective gate voltage and the transconductance of the diffpair in the balanced-state, when 300-μA are flowing through each transistor is given by

$$V_{\mathrm{gt}} = \sqrt{\frac{2I}{\mu_{\mathrm{n}} C_{\mathrm{ox}}(W/L)}} = 212\text{-mV}, \tag{3.25}$$

$$g_{\mathrm{m}} = \frac{I}{V_{\mathrm{gt}}/2} = \frac{300\text{-}\mu\mathrm{A}}{106\text{-mV}} = 2.83\text{-}\mu\mathrm{A/mV}. \tag{3.26}$$

The open-loop gain is the product of this g_{m} and the transresistance R_{m} of the I/V converter. In this analysis we have assumed the product $g_{\mathrm{m}} R_{\mathrm{m}} = 40$. Given that $g_{\mathrm{m}} = 2.83$-mA/V, then R_{m} is simply calculated to be 14.1-kΩ.

Case 3: $\mathbf{G_{m1}}$ is Partially ON, $\mathbf{G_{m2}}$ is Partially ON

The most important case occurs during interpolation, when both G_{m} cells are partially ON. The multi-input-stage has an adverse effect on the distortion when the gain is interpolated between the two extremes. The increased distortion is a result of the virtual ground position moving away from the input of the active diffpair. This situation is illustrated in the projection diagram of Fig. 3.4, which shows the virtual ground approximately halfway between the first and second G_{m} cells. The overall feedback acts to force the sum of the current from the two G_{m} cells to zero, but does not force each individual current to zero, therefore, a potentially large differential voltages of opposite phase can result at the input of the two diffpairs. The magnitude of these voltages can be much larger than the swing at the virtual ground node. Therefore, the modulation index of each diffpair will increase during interpolation. This increases the distortion from each diffpair, however these distortions are out of phase, so the effect of cancellation of distortion also must be considered.

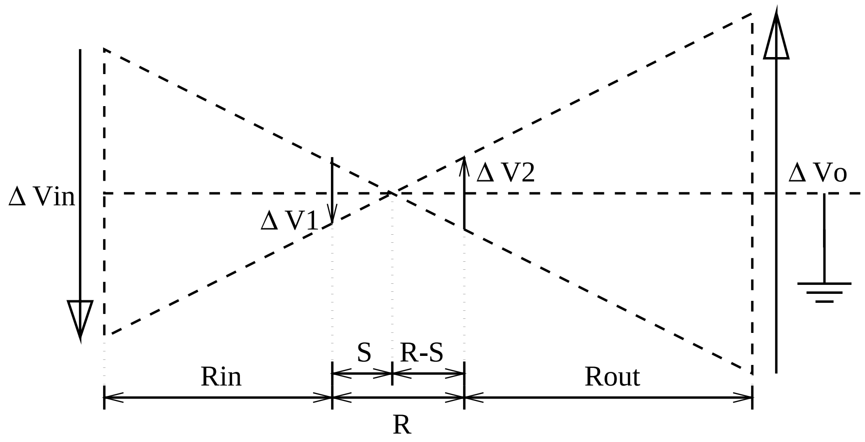


Figure 3.4 Projection diagram of an opamp with two transconductance cells.

Difference in Voltage Swing at the Input of Each $\mathbf{G_m}$ Cell In order to determine the distortion in the presence of interpolation, we need to find the magnitude of the voltage swing that will be present on the inputs of each diffpair. We can easily find the difference in magnitudes of the voltages from the projection diagram. In the case where the lowest gain is desired, the virtual ground will appear at the input to G_{m1}. The voltage difference across R_{out} is simply ΔV_{o}. Since a constant current flows in the feedback resistor, the voltage difference between G_{m2} and G_{m1}, is just given by the ratio of the resistors,

$$\Delta V_2 - \Delta V_1 = \frac{R}{R_{\mathrm{out}}} \Delta V_{\mathrm{o}}. \tag{3.27}$$

Since the VGA's function is to keep the output voltage constant, it is usually convenient to express voltages in terms of the output voltage swing, which should be a constant, known quantity. If we keep the output voltage fixed and repeat the analysis for the minimum gain setting. The virtual ground node is now at G_{m2}; the output voltage appears across the series combination of $R_{\mathrm{out}} + R$. Therefore, the voltage difference seen between G_{m2} and G_{m1} is slightly smaller and given by

$$\Delta V_2 - \Delta V_1 = \frac{R}{R_{\mathrm{out}} + R} \Delta V_{\mathrm{o}}. \tag{3.28}$$

This is similar to the above result of (3.27) when $R << R_{\mathrm{out}}$, which is typically the case.

Expressing $\mathbf{\Delta V_2 - \Delta V_1}$ in Terms of $\mathbf{\Delta A_v}$ It will be more convenient for design iterations if the difference in voltage is expressed in terms of the closed-loop gain increment ΔA_v. For an opamp with infinite open-loop gain, the difference in the closed-loop gain can be expressed as

$$\Delta A_v = \frac{R_{\mathrm{out}} + R}{R_{\mathrm{in}}} - \frac{R_{\mathrm{out}}}{R_{\mathrm{in}} + R}, = \frac{R(R_{\mathrm{in}} + R + R_{\mathrm{out}})}{R_{\mathrm{in}}(R_{\mathrm{in}} + R)}. \tag{3.29}$$

If we operate under the assumption that $R << R_{\text{in}}, R_{\text{out}}$, we see that ΔA_v is given approximately by

$$\Delta A_v \approx R\left[\frac{R_{\text{in}} + R_{\text{out}}}{{R_{\text{in}}}^2}\right]. \tag{3.30}$$

We can then solve for R such that

$$R = R_{\text{in}}\Delta A_v\left[\frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{out}}}\right]. \tag{3.31}$$

Dividing the second term by R_{in} gives

$$R = R_{\text{in}}\Delta A_v\left[\frac{1}{1 + R_{\text{out}}/R_{\text{in}}}\right]. \tag{3.32}$$

When this value of R is substituted into (3.27), we find that

$$\Delta V_2 - \Delta V_1 = \Delta V_{\text{o}}\Delta A_v\left[\frac{R_{\text{in}}}{R_{\text{out}}}\right]\left[\frac{1}{1 + R_{\text{out}}/R_{\text{in}}}\right]. \tag{3.33}$$

For the assumption of $R << R_{\text{in}}, R_{\text{out}}$, the closed-loop gain is just $A_v = -R_{\text{out}}/R_{\text{in}}$, therefore, we can write the above results in terms of ΔV_{o},

$$\Delta V_2 - \Delta V_1 = \Delta V_{\text{o}}\left[\frac{\Delta A_v}{|A_v|}\right]\left[\frac{1}{1 + |A_v|}\right]. \tag{3.34}$$

This result shows that the difference in voltage at the two G_{m} stages is proportional to the output voltage, ΔV_{o}, and the fractional gain change, $\Delta A_v/|A_v|$. It is also inversely proportional to $(1 + |A_v|)$. The reason for this is that for a fixed R, the amount of the change in gain, will depend on the size of the original gain; the smaller the original gain (smaller R_{out}), the larger the change in incremental gain, since R is a larger fraction of R_{out}.

Since this voltage difference will be used throughout the forthcoming analysis, it is convenient to define a symbol ΔV_A for this difference,

$$\Delta V_A \triangleq \Delta V_{\text{o}}\left[\frac{\Delta A_v}{|A_v|}\right]\left[\frac{1}{1 + |A_v|}\right]. \tag{3.35}$$

where the subscript A reminds us that the difference is a function of the closed-loop gain deviation.

Actual Voltage Swing at the Input of Each $\mathbf{G_m}$ Cell The above relationship constrains the difference of the input voltages to each diffpair. We need an additional constraint to find the actual magnitude of each individual voltage swing. We will use the constraint imposed by the negative feedback. In this first-order analysis we will assume an infinite open-loop gain to facilitate the calculations. We can then relax that constraint and consider a finite gain opamp.

If the opamp has infinite gain, the feedback will force the sum of the currents from the two G_{m} cells to zero.

$$\Delta I = g_{\text{m}1}\Delta V_1 + g_{\text{m}2}\Delta V_2 = 0. \tag{3.36}$$

This gives an added constraint to the input voltages. We can write ΔV_1 as

$$\Delta V_1 = -\Delta V_2 \frac{g_{\mathrm{m}2}}{g_{\mathrm{m}1}}. \tag{3.37}$$

Combining this with the constraint that $\Delta V_2 - \Delta V_1 = \Delta V_A$ we find that the input voltage are given by the simple lever-arm relationship,

$$\begin{aligned} \Delta V_1 &= \frac{-g_{\mathrm{m}2}}{g_{\mathrm{m}1} + g_{\mathrm{m}2}} \Delta V_A, & (3.38) \\ \Delta V_2 &= \frac{g_{\mathrm{m}1}}{g_{\mathrm{m}1} + g_{\mathrm{m}2}} \Delta V_A. & (3.39) \end{aligned}$$

Distortion as a Function of Bias Current The gain of the VGA is controlled by changing the bias current of each G_{m} cell. Therefore, it is desirable to find a relationship between the distortion and the bias current. In the following analysis it will be assumed that we have a fixed tail current that can be steered between stages 1 and 2. Therefore the total current is a constant and the individual tail current can be expressed in terms of the current-steering parameter x.

$$\left.\begin{aligned} I_1 &= (1-x) I_{\mathrm{bias}} \\ I_2 &= x I_{\mathrm{bias}} \end{aligned}\right\} \quad \text{for} \quad 0 \leq x \leq 1. \tag{3.40}$$

The output current of each diffpair can be written as a Taylor series expansion such that

$$\frac{\Delta I_1}{I_1} = f\left(\frac{\Delta V_1}{V_{\mathrm{gt}1}}\right), \tag{3.41}$$

and the total current ΔI is just the sum of the currents from each G_{m} cell;

$$\Delta I = \Delta I_1 + \Delta I_2 = I_1 f\left(\frac{\Delta V_1}{V_{\mathrm{gt}1}}\right) + I_2 f\left(\frac{\Delta V_2}{V_{\mathrm{gt}2}}\right). \tag{3.42}$$

Writing this in terms of I_{bias} and x, we get

$$\frac{\Delta I}{I_{\mathrm{bias}}} = (1-x) f\left(\frac{\Delta V_1}{V_{\mathrm{gt}1}}\right) + x f\left(\frac{\Delta V_2}{V_{\mathrm{gt}2}}\right). \tag{3.43}$$

Expressing V_{gt} and ΔV in Terms of x The effective gate voltage V_{gt} for each diffpair varies as the square-root of the bias current. If the maximum V_{gt} is defined as $V_{\mathrm{gt}0}$ which is the effective gate voltage when half the tail current I_{bias} flows in one of the diffpair transistors, such that

$$V_{\mathrm{gt}0} \triangleq \sqrt{\frac{I_{\mathrm{bias}}}{\mu_{\mathrm{n}} C_{\mathrm{ox}}(W/L)}}, \tag{3.44}$$

then $V_{\mathrm{gt}1}$ and $V_{\mathrm{gt}2}$ can be written as

$$\begin{aligned} V_{\mathrm{gt}1} &= \sqrt{\frac{I_1}{\mu_{\mathrm{n}} C_{\mathrm{ox}}(W/L)}} = \sqrt{1-x}\, V_{\mathrm{gt}0}, & (3.45) \\ V_{\mathrm{gt}2} &= \sqrt{\frac{I_2}{\mu_{\mathrm{n}} C_{\mathrm{ox}}(W/L)}} = \sqrt{x}\, V_{\mathrm{gt}0}. & (3.46) \end{aligned}$$

Finally, we can write the input voltage swing in terms of x. For a simple square-law model, g_{m} is proportional to the square-root of the bias current. Therefore,

$$\Delta V_1 = \frac{-g_{\mathrm{m2}}}{g_{\mathrm{m1}}+g_{\mathrm{m2}}}\Delta V_A = \frac{-\sqrt{x}}{\sqrt{1-x}+\sqrt{x}}\Delta V_A, \tag{3.47}$$

$$\Delta V_2 = \frac{g_{\mathrm{m1}}}{g_{\mathrm{m1}}+g_{\mathrm{m2}}}\Delta V_A = \frac{\sqrt{1-x}}{\sqrt{1-x}+\sqrt{x}}\Delta V_A. \tag{3.48}$$

Distortion as a Function of Design Variables With the above expressions we can now write the normalized voltage swings in terms of design quantities (V_{o}, A_v, ΔA_v, V_{gt0}, and x);

$$\frac{\Delta V_1}{V_{\mathrm{gt1}}} = \frac{-\sqrt{x}/\sqrt{1-x}}{\sqrt{1-x}+\sqrt{x}}\left(\frac{\Delta V_A}{V_{\mathrm{gt0}}}\right) = -\gamma_1(x)\left(\frac{\Delta V_A}{V_{\mathrm{gt0}}}\right), \tag{3.49}$$

$$\frac{\Delta V_2}{V_{\mathrm{gt2}}} = \frac{\sqrt{1-x}/\sqrt{x}}{\sqrt{1-x}+\sqrt{x}}\left(\frac{\Delta V_A}{V_{\mathrm{gt0}}}\right) = \gamma_2(x)\left(\frac{\Delta V_A}{V_{\mathrm{gt0}}}\right), \tag{3.50}$$

where the coefficients $\gamma_1(x)$ and $\gamma_2(x)$ have been defined as

$$\gamma_1(x) \triangleq \frac{\sqrt{x}/\sqrt{1-x}}{\sqrt{1-x}+\sqrt{x}}, \tag{3.51}$$

$$\gamma_2(x) \triangleq \frac{\sqrt{1-x}/\sqrt{x}}{\sqrt{1-x}+\sqrt{x}}. \tag{3.52}$$

The total output current from the two G_{m} cells is then given by

$$\frac{\Delta I}{I_{\mathrm{bias}}} = (1-x)f\left(-\gamma_1(x)\frac{\Delta V_A}{V_{\mathrm{gt0}}}\right) + xf\left(\gamma_2(x)\frac{\Delta V_A}{V_{\mathrm{gt0}}}\right). \tag{3.53}$$

Taylor Series: Linear Term We can now consider the Taylor series coefficient of $f(\cdot)$. The linear term is simply

$$\mathrm{Term}_{\mathrm{1st}} = (1-x)(-\gamma_1(x))\frac{\Delta V_A}{V_{\mathrm{gt0}}} + x\gamma_2(x))\frac{\Delta V_A}{V_{\mathrm{gt0}}}, \tag{3.54}$$

$$= -\frac{\sqrt{1-x}\sqrt{x}}{\sqrt{1-x}+\sqrt{x}}\frac{\Delta V_A}{V_{\mathrm{gt0}}} + \frac{\sqrt{1-x}\sqrt{x}}{\sqrt{1-x}+\sqrt{x}}\frac{\Delta V_A}{V_{\mathrm{gt0}}}, \tag{3.55}$$

$$\equiv 0. \tag{3.56}$$

The fact that the linear term vanishes is a result of our assumption that the open-loop gain is infinite. In reality, the open-loop gain $G_{\mathrm{m}}R_{\mathrm{m}}$ is approximately 32-dB at the frequencies of interest. Therefore there must be an output current of

$$\Delta I = \frac{\Delta V_{\mathrm{o}}}{R_{\mathrm{m}}}, \tag{3.57}$$

and normalizing by the bias current

$$\frac{\Delta I}{I_{\mathrm{bias}}} = \frac{\Delta V_{\mathrm{o}}}{I_{\mathrm{bias}}R_{\mathrm{m}}}. \tag{3.58}$$

This equation gives use of the magnitude of the linear term. However, we can rearrange this magnitude so that it is expressed in terms of design variables. We can multiply and divide by V_{gt0}

$$\frac{\Delta I}{I_{\mathrm{bias}}} = \frac{\Delta V_{\mathrm{o}}}{V_{\mathrm{gt0}}} \frac{1}{R_{\mathrm{m}}} \left[\frac{V_{\mathrm{gt0}}/2}{I_{\mathrm{bias}}/2} \right]. \tag{3.59}$$

We recognize the final fraction as the transconductance of a transistor biased in the balanced state. Therefore, we can write

$$\frac{\Delta I}{I_{\mathrm{bias}}} = \frac{\Delta V_{\mathrm{o}}}{V_{\mathrm{gt0}}} \frac{1}{g_{\mathrm{m}} R_{\mathrm{m}}} = \frac{\Delta V_{\mathrm{o}}}{V_{\mathrm{gt0}}} \frac{1}{A_f}, \tag{3.60}$$

where A_f is the open-loop gain of the opamp at the frequency of interest.

Taylor Series: 3rd-Order Term The third-order term of the Taylor series is given by

$$\mathrm{Term}_{\mathrm{3rd}} = -\frac{1}{8} \left[(1-x)\gamma_1(x)^3 - x\gamma_2(x)^3 \right] \left[\frac{\Delta V_A}{V_{\mathrm{gt0}}} \right]^3. \tag{3.61}$$

The contributions to the third-order distortion are plotted in Fig. 3.5. The two terms $(1-x)\gamma_1(x)^3$ and $x\gamma_2(x)^3$ are shown as a function of the bias current steering parameter x. Notice that the modulation index changes rapidly near the edges.

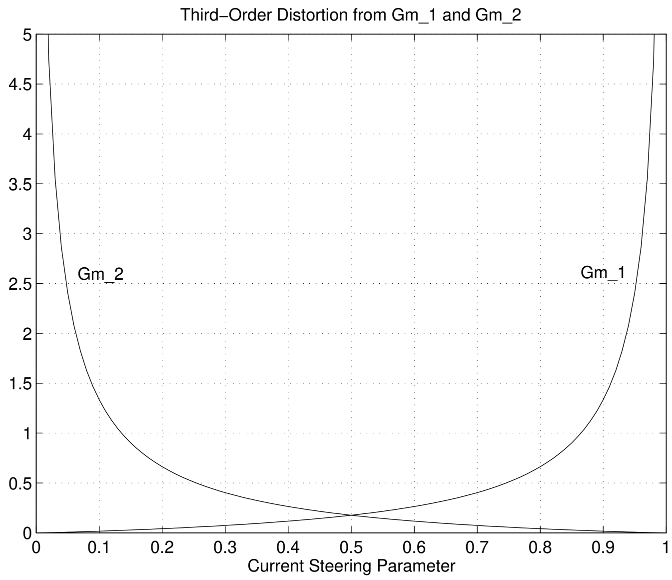


Figure 3.5 Third-order normalized distortion contribution from individual cells G_{m1} and G_{m2}.

Since the two diffpairs have voltages that are out of phase, there will be a cancellation of the third-order term. The combined third-order distortion coefficient for both G_{m} cells in operation simultaneously is shown in Fig. 3.6, which plots $\Gamma(x) = [(1-x)\gamma_1(x)^3 - x\gamma_2(x)^3]$ as a function of x. The absolute value of the same curve is plotted in Fig. 3.7 where the scale is in decibels. This represents the relative change in distortion as a function of x.

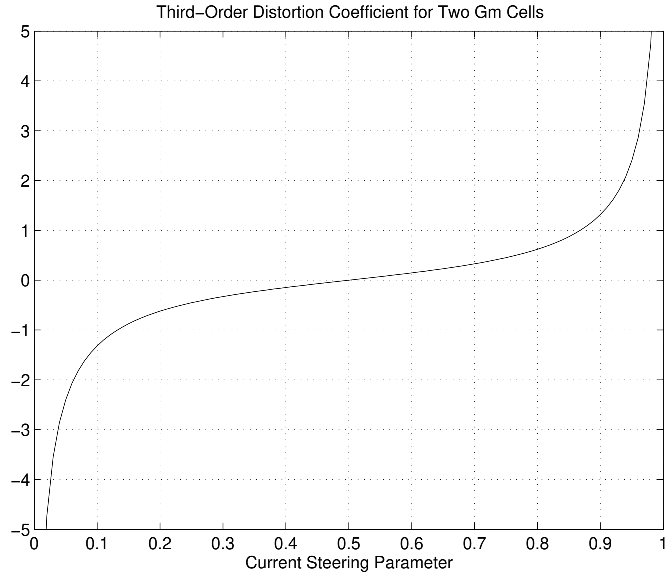


Figure 3.6 Third-order normalized distortion contribution from combined cells G_{m1} and G_{m2}.

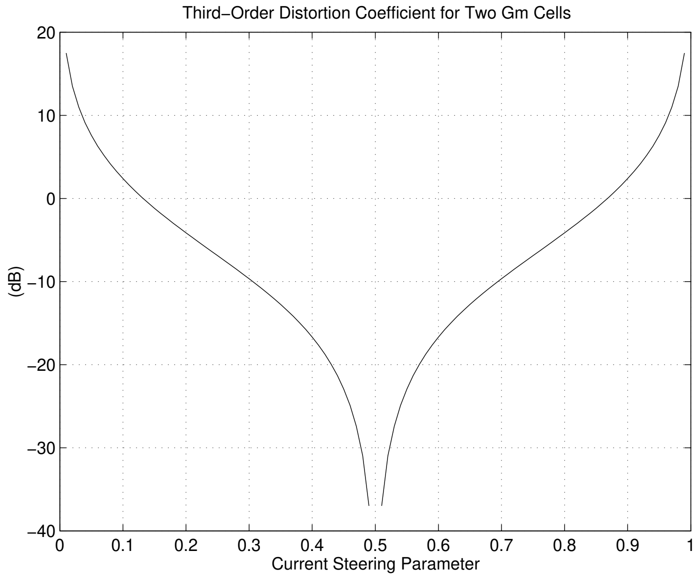


Figure 3.7 Third-order normalized distortion contribution in decibels from combined cells G_{m1} and G_{m2}.

Increased Distortion at End-Points The interesting things to note from Fig. 3.6 is that a perfect cancellation occurs at the balance points. In this case the two diff-pairs are exactly canceling each others distortion. The second point is that the distortion increases rapidly toward the end points. This phenomenon can be understood by considering the situation illustrated in Fig. 3.8. The second G_m cell has two-thirds of the bias current, while the first has one-third of the bias current. The overall distortion increases in two ways as the current between the diffpairs becomes more heavily unbalanced.

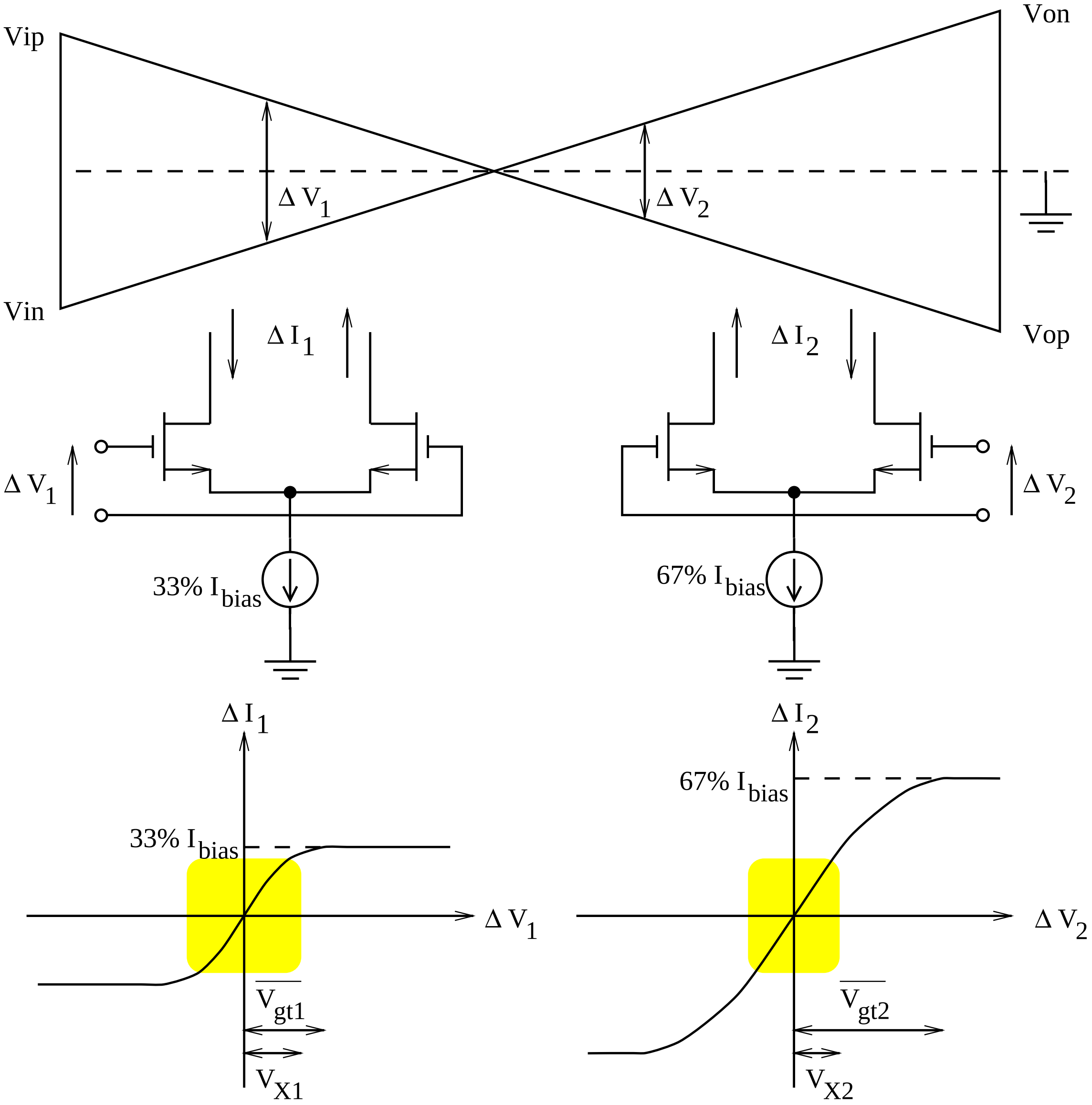


Figure 3.8 Changes in modulation index as a function of bias current.

Higher Voltage Swing As the current is increased in one diffpair the feedback tends to place the virtual-ground node closer to that diffpair so the voltage swing will reduce. Conversely, the diffpair with less current will see a larger voltage swing.

Lower Linear Voltage Range The diffpair with a larger current will have a large effective gate voltage V_{gt}, whereas the diffpair with less current will have a reduced V_{gt}, thus reducing its linear-range of operation.

These two effects cause the modulation index of the diffpair with less current to rise rapidly. The signal amplitude is increasing while the linear range of the diffpair is getting smaller. Despite the fact that the bias current is getting smaller in the more heavily distorted diffpair, so that the gain is reduced to the output, the modulation index is increasing much faster so that the current generated from the diffpair is heavily distorted and the end result is an increase in total distortion.

Distortion Expression for Interpolation

We can now use the results of the previous sections to get a first-order estimate of the THD as a function of the tuning current. We know that the 3rd-order harmonic distortion is obtained by taking one-fourth of the ratio of the 3rd-order term to the linear term. If we define the current-steering function $\Gamma(x)$ as,

$$\Gamma(x) \triangleq \left[(1-x)\gamma_1(x)^3 - x\gamma_2(x)^3\right]. \tag{3.62}$$

Then, the open-loop third-order harmonic distortion is given by

$$\mathrm{HD}_3 = \frac{-\frac{1}{32}\left[\Gamma(x)\right]\left[\frac{\Delta V_A}{V_{\mathrm{gt0}}}\right]^3}{\left[\frac{\Delta V_o}{V_{\mathrm{gt0}}}\right]\frac{1}{A_f}}. \tag{3.63}$$

Recalling (3.35), this can be simplified to

$$\mathrm{HD}_3 = -\frac{1}{32}\left[\Gamma(x)\right]\left[\frac{\Delta V_o}{V_{\mathrm{gt0}}}\right]^2\left[\left(\frac{\Delta A_v}{|A_v|}\right)\left(\frac{1}{1+|A_v|}\right)\right]^3 A_f. \tag{3.64}$$

This expression can be further simplified if we normalize the output voltage by the open-loop gain, to obtain the voltage at the virtual-ground node.

$$\mathrm{HD}_3 = -\frac{1}{32}\left[\frac{\Delta V_o/A_f}{V_{\mathrm{gt0}}}\right]^2\left[\Gamma(x)\right]\left[\left(\frac{\Delta A_v}{|A_v|}\right)\left(\frac{1}{1+|A_v|}\right)\right]^3 A_f^3. \tag{3.65}$$

We recognize the first two terms as the harmonic distortion of the open-loop VGA when only one diffpair is operational, which we can define as $\overline{\mathrm{HD}_3}$, such that

$$\boxed{\overline{\mathrm{HD}_3} \triangleq -\frac{1}{32}\left[\frac{\Delta V_o/A_f}{V_{\mathrm{gt0}}}\right]^2.} \tag{3.66}$$

The open-loop distortion of the VGA with interpolation is then given by

$$\mathrm{HD}_3 = \overline{\mathrm{HD}_3}\left[\Gamma(x)\right]\left[\left(\frac{\Delta A_v}{|A_v|}\right)\left(\frac{1}{1+|A_v|}\right)\right]^3 A_f^3. \tag{3.67}$$

The closed loop distortion will be reduced by the loop gain which is approximately equal to $A_f/|A_v|$. Therefore, we finally can write the third-order harmonic distortion of the VGA as

$$\boxed{\mathrm{HD}_{3,\mathrm{closed}} = \overline{\mathrm{HD}_3}\,[\Gamma(x)]\left[\left(\frac{\Delta A_v}{|A_v|}\right)\left(\frac{1}{1+|A_v|}\right)\right]^3 A_f^2|A_v|,} \tag{3.68}$$

where we recall that A_v is the closed-loop gain of the VGA, A_f is the open-loop gain, $\overline{\mathrm{HD}_3}$ is the distortion in open-loop when there is no interpolation, and $\Gamma(x)$ is plotted in Fig. 3.7.

Numerical Example

We can evaluate the distortion as a function of x given in (3.68) for $\Delta A_v = 0.1$, $A_v = -1$, $A_f = 40$, $\Delta V_o = \pm 2$-V, and V_{gt0}=212-mV for (W/L)=80. We found from (3.20) that under these conditions $\overline{\mathrm{HD}_3} = -55$-dB.

$$\left|A_f^2 A_v\right|\left[\left(\frac{\Delta A_v}{A_v}\right)\left(\frac{1}{1+|A_v|}\right)\right]^3 = 0.2 = -14\text{-dB}. \tag{3.69}$$

The total distortion will be

$$\begin{aligned} \mathrm{HD}_{3,\mathrm{closed}} &= -55 - 14 + 20\log[\Gamma(x)]. && (3.70)\\ &= -69 + 20\log[\Gamma(x)]. && (3.71) \end{aligned}$$

Therefore, $\Gamma(x)$ must remain below 9-dB to achieve -60-dB THD over all gain settings. HD_3 for this numerical example is plotted in Fig. 3.9 as a percentage scale and in decibels. Since THD is dominated be HD_3, these plots also represent an accurate estimate of the THD for the circuit parameters chosen. We can see that the THD remains below -60-dB for about 90% of the interpolation range. To check our results, the above equation is compared to the simulation results of Fig. 3.10. The simulation results show very close agreement to the calculated curve of Fig. 3.9. Although the distortion performance exceeds 60-dB dynamic-range for most of the gain settings, there is still a sharp increase in distortion at the edges that can degrade the results. To improve performance, we can change the bias quantities and devices sizes to increase V_{gt}. However, this soon becomes unwieldy, requiring either a huge increase in current, or causing a severe penalty in bandwidth and stability. Therefore, a different, more efficient, method is needed to gain further improvements. The method that is used here is to take advantage of further cancellation.

Cancellation Techniques Using Three Active $\mathrm{G_m}$ Cells

In the previous analysis we have considered only two diffpairs active at any given time. However, it is possible to make three diffpairs active. If this is done, the two diffpairs on the ends will have large distortions, but their distortions will tend to

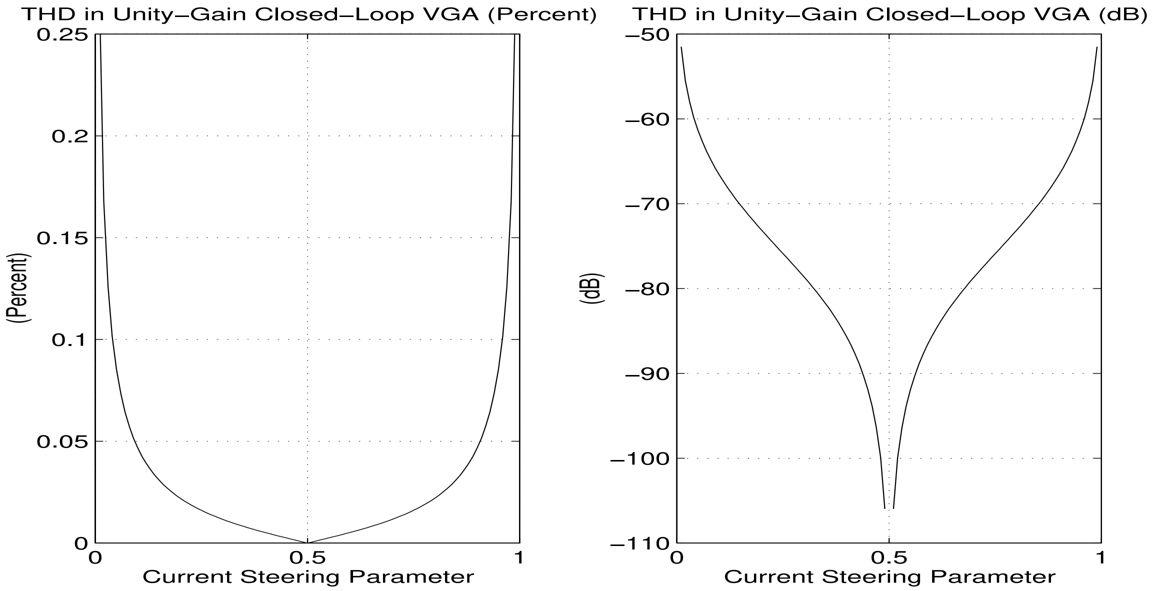


Figure 3.9 Calculated THD for interpolation between two transconductance cells: Linear scale, and dB scale.

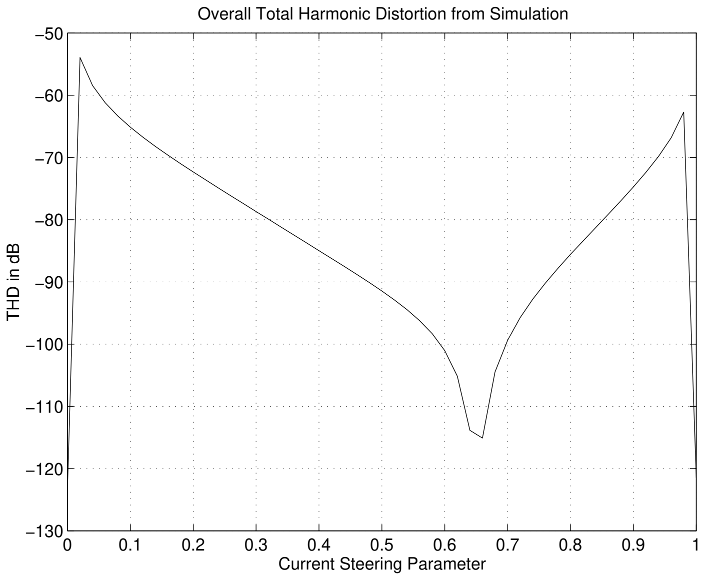


Figure 3.10 Simulated THD in dB for interpolation between two transconductance cells.

cancel. In order to have the maximum control on the voltage gain and the possible lowest THD, the first and the last $G_{\rm m}$ cell are not in use. The overall distortion for this case as a function of the control voltage is predicted in Fig. 3.11a. The exact calculation of this effect will not be given. Rather we will turn to simulations to optimize the result and use the above analysis as a guide for which changes in the design will be most effective in reducing distortion.

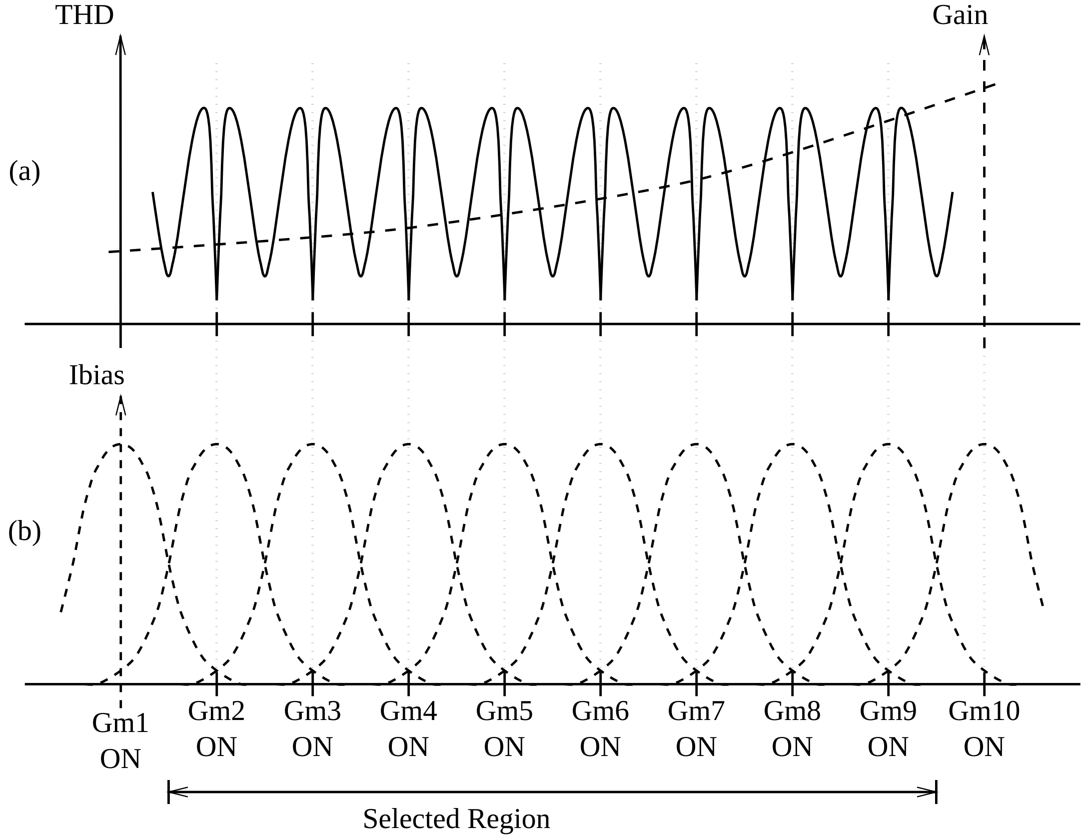


Figure 3.11 Predicted THD for the VGA with a multi-input-stage.

3.1.4 Circuit Analysis — Output Stage

The output stage used in the VGA is a differential Class-A common-source configuration. If the output transistor is a perfect square-law device, ideally, it would only give rise to second-order harmonic distortion, provided that all the output current is diverted to the linear load resistor. We can refer to Fig. 2.10 with a loading resistor, $R_{\rm load}$, connected to ground. The following equations describe the output voltage.

$$\begin{aligned} V_{\rm out} &= R_{\rm load} \times (I_{\rm d} - I_{\rm bias}), \\ &= R_{\rm load} \times \left[(K/2)(V_{\rm dd} - V_{\rm o} - V_{\rm t})^2 - I_{\rm bias} \right]. \end{aligned} \tag{3.72}$$

If $V_{\rm dd} - V_{\rm o} = V_{\rm gt} + V_{\rm t} + V_{\rm m}\cos\omega t$ and $V_{\rm gt} > V_{\rm m}$,

$$\begin{aligned} V_{\rm out} &= R_{\rm load} \times \left[(K/2)(V_{\rm gt} + V_{\rm m}\cos\omega t)^2 - I_{\rm bias}\right], \\ &= R_{\rm load}\frac{K}{2}V_{\rm gt}^2 \times \left\{\left[1 + \frac{1}{2}\left(\frac{V_{\rm m}}{V_{\rm gt}}\right)^2 - \frac{I_{\rm bias}}{\frac{K}{2}V_{\rm gt}^2}\right] + 2\frac{V_{\rm m}}{V_{\rm gt}}\cos\omega t \right. \\ &\left. + \frac{1}{2}\left(\frac{V_{\rm m}}{V_{\rm gt}}\right)^2 \cos 2\omega t\right\}. \end{aligned} \tag{3.73}$$

If the system is designed differentially, the second harmonic distortion will be equal and of opposite phase for the positive and negative signal path. Therefore, in differential mode, the second harmonic will disappear and the output stage becomes a perfectly linear system. This means that, in term of THD, the Class-A common-source configuration is a good choice for the output stage.

In actuality we cannot obtain perfect cancellation and there will be distortion from the output stage. There are several assumptions that mislead us into drawing the previous conclusions. These assumptions are as follows:

- The square law of the CMOS transistor is a mathematical approximation. The actual transistor will not behave as an ideal square-law device and some odd-order distortion will be generated.
- The cancellation of the second-order distortion relies on perfect matching. In a real circuit there will be random mismatches so a fraction of the even-order distortion terms will appear in the differential output.
- In Fig. 2.10, the ideal current source is used. In a practical situation, $I_{\rm bias}$ will vary with $V_{\rm out}$. Therefore, signal current will be lost in the output impedance of the output stage, and this current will be nonlinear.

If the channel modulation effect of the transistor and the practical current source are considered, (3.72) should be modified to be

$$V_{\rm out} = R_{\rm load} \times \left\{\frac{K}{2}(V_{\rm dd} - V_{\rm o} - V_{\rm t})^2\left[1 + \lambda(V_{\rm dd} - V_{\rm out})\right] - I_{\rm bias} \times f(V_{\rm out})\right\}. \tag{3.74}$$

Now the situation is different; all the higher-order harmonics will not disappear.

As we have seen the distortion is highly sensitive to the modulation index, which is proportional to the open-loop gain. Therefore, the choice of compensation capacitor, which will alter the loop gain as a function of frequency, will have a large impact on the distortion. This topic will be considered at the end of Section 3.2, where we discuss other issues related to frequency compensation, such as closed-loop stability.

3.2 Frequency Compensation

Generally, the topology of a CMOS opamp is similar to its BJT counterpart; but its g_{m} value is an order of magnitude lower. This makes the zero of a Miller compensation capacitor at a frequency 10 times lower, and also the location of the second pole, provided the parasitic capacitances are equal in both cases. Moving the zero closer to the passband causes a phase-lag in the opamp and can result in instabilities. We can consider the following two compensation techniques for a common-source stage, as shown in Fig. 3.12 and Fig. 3.13. Both will split the first two dominant poles apart if C_{c} is large enough.

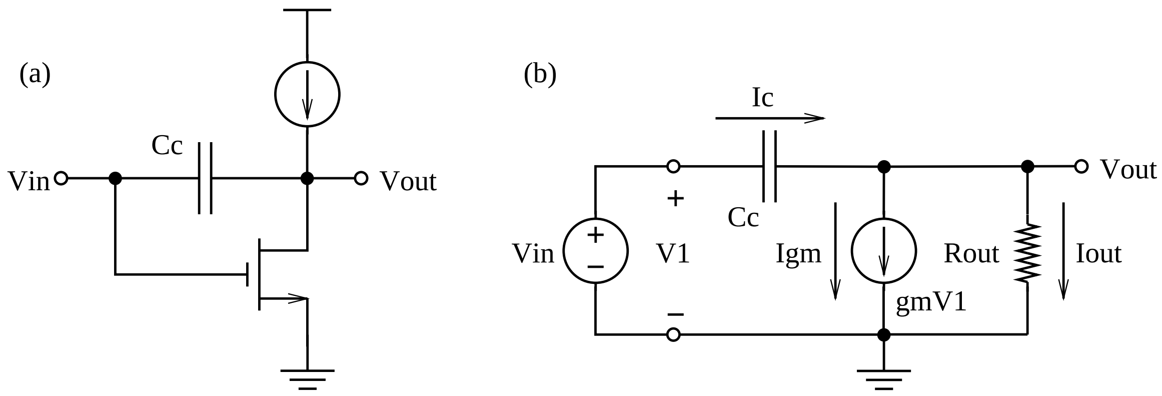


Figure 3.12 Compensation by a capacitor.

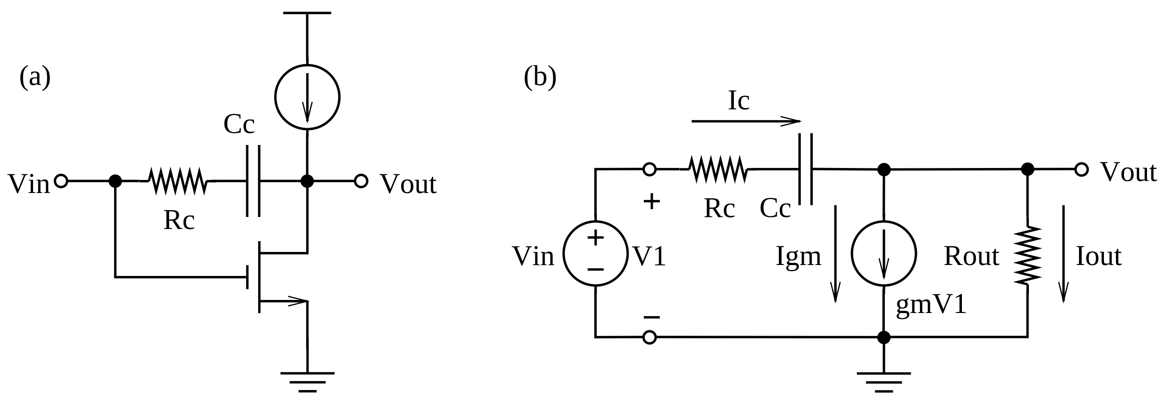


Figure 3.13 Compensation by a capacitor and a resistor.

We can find the location of the transmission zero. Referring to Fig. 3.12, if $v_{\mathrm{out}} = 0$ or $i_{\mathrm{out}} = 0$,

$$
\begin{aligned}
\frac{v_{\mathrm{in}}}{1/s_{\mathrm{z}} C_{\mathrm{c}}} &= g_{\mathrm{m}} \cdot v_{\mathrm{in}}, \\
s_{\mathrm{z}} &= \frac{g_{\mathrm{m}}}{C_{\mathrm{c}}}. \qquad (3.75)
\end{aligned}
$$

The location of the zero is fixed. We cannot move it easily. Whereas, in Fig. 3.13, if $v_{\mathrm{out}} = 0$ or $i_{\mathrm{out}} = 0$,

$$\begin{aligned} \frac{v_{\mathrm{in}}}{R_{\mathrm{c}} + 1/s'_{\mathrm{z}}C_{\mathrm{c}}} &= g_{\mathrm{m}} \cdot v_{\mathrm{in}}, \\ \frac{1}{s'_{\mathrm{z}}C_{\mathrm{c}}} &= \frac{1}{g_{\mathrm{m}}} - R_{\mathrm{c}}, \\ s'_{\mathrm{z}} &= \frac{g_{\mathrm{m}}/C_{\mathrm{c}}}{1 - g_{\mathrm{m}}R_{\mathrm{c}}}. \end{aligned} \tag{3.76}$$

If $R_{\mathrm{c}} \rightarrow (1/g_{\mathrm{m}})$ is chosen, $s'_{\mathrm{z}} \rightarrow \infty$. Therefore inserting a resistor in series with the compensation capacitor can move the zero to infinity and out of our way. However, this movement of the zero does not come for free. The second-pole moves to a lower frequency. Therefore, an optimal value of R_{c} exists to get the highest bandwidth for a predefined phase-margin.

In order to have the maximum control of the zero's location, the second method of compensation of Fig. 3.13 is employed in this VGA. The resistor is replaced by a transistor operated in the triode region; its conductance can be controlled by adjusting the gate potential. In the schematic of Fig. 3.13, the swing of V_{in} is always smaller than that of V_{out}. The source of the transistor ties to V_{in}, therefore, the defined conductance with respect to gate is fairly stable for all output voltages.

The size of the compensation capacitor is very large in order to introduce the dominate pole. Also, this capacitor will restrict the maximum rate of change of the output signal. This rate is called the *slew rate* and is given by

$$\frac{dV_{C_{\mathrm{c}}}}{dt} = \frac{I_{C_{\mathrm{c}}}}{C_{\mathrm{c}}}. \tag{3.77}$$

At high frequencies and large output swing, $dV_{C_{\mathrm{c}}}/dt$ will be large, and higher $I_{C_{\mathrm{c}}}$ is required to charge the capacitor. Since $I_{C_{\mathrm{c}}} = g_{\mathrm{m}} \times \Delta V_{\mathrm{in}}$ where g_{m} and ΔV_{in} are the transconductance and the differential input voltage of previous (transconductance) stage, a large $I_{C_{\mathrm{c}}}$ implies a large ΔV_{in}. If we require that the original value of V_{o} remains fixed to achieve a desired voltage gain, then the previous (transconductance) stage will be operated at a larger input signal thus increasing the modulation index and increasing the nonlinearities. This aspect must be considered carefully in the design of a high-fidelity voltage gain amplifier.

Very large distortions can exist during a slew condition. This will, however, only be critical when the rate of change of the output signal is close to the slew rate, which is equivalent to the signal frequency being close to the full power bandwidth.[4] The case of slewing is a highly nonlinear phenomenon and was not considered in the analysis of THD in Section 3.1.4.

[4] The full power bandwidth (FPBW) is defined to be

$$\mathrm{FPBW} = \frac{\text{Slew Rate}}{2\pi V_{\max}}.$$

3.3 Noise Performance

Gray and Meyer [7] state the fundamental limitations due to noise: "The existence of noise is basically due to the fact that electrical charge is not continuous but is carried in discrete amounts equal to the electron charge, and thus noise is associated with fundamental process in the integrated-circuit devices." Different types of noise in CMOS devices includes shot noise, thermal noise, flicker noise, and burst noise [7, 12].

Simple Noise Relationships The noise analysis in both hand calculation and simulation assumes that each noise source is statistically independent from the other noise sources in the circuit. The total output noise is the root-mean-square (rms) of the individual noise source contributions. According to basic probability theories [13], if an event has a probability density function *pdf*(f), and mean and variance of μ_x and σ_x^2, respectively, an experiment repeated N times will leave the mean unchanged, μ_x, but the variance will be reduced by a factor of N or the new variance is σ_x^2/N, and the new standard deviation is reduced by $\sqrt{N}$. This shows that the noise is reduced by averaging. A filter will average an electronic signal. The smaller the bandwidth, the greater the averaging, so we would expect a relationship that the rms value of the noise is proportional to the square-root of the bandwidth.

The above theories can be used to calculate the noise in a transistor as a random fluctuation of the expected signal. Noise averaging also works in the spatial domain. A transistor with an aspect ratio of W/L and a bias current of I_{bias} has a mean-square noise amplitude of $\overline{\mathcal{N}^2} = \sigma_x^2$. If the width of the transistor is scaled by a factor of N, and the current remains unchanged, the squared noise amplitude is expected to be $\overline{\mathcal{N}'^2} = \sigma_x^2/N$. But the price paid is a bigger transistor and therefore a reduction of bandwidth.

Generally, it would be a complicated issue for making a hand calculation of the noise performance of a medium size or a large size analog circuit. However, the noise of the circuit is usually dominated by only a few transistors, which see the largest gain to the output. These transistors can be found at the input stage. In an opamp, the noise (from the device itself and from coupling) in the early stages would have a larger contribution as it would be further amplified by the latter stages (see Fig. 3.14). Therefore, special cares must be taken in both design and layout.

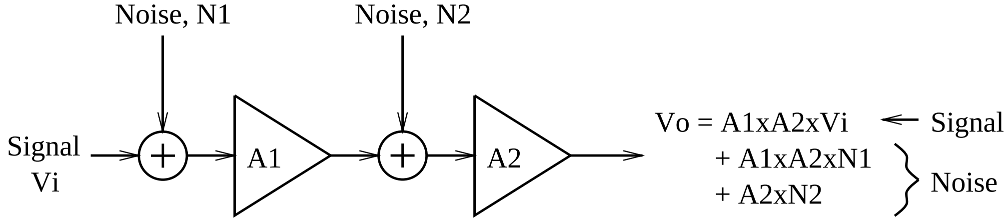


Figure 3.14 Noise in an operational amplifier.

In the proposed VGA, a series of transconductance cells are connected in parallel and form the multi-input-stage. Noise from each cell may dominate the noise that appears at the output. But, since their output is current and the noise (thermal and flicker) will increase with drain current, the cell with low bias current can be neglected. The noise in the multi-input-stage is dominated by a few cells.

The white-noise in a simple CMOS transistor is given as a power-spectral density of

$$S(f) = \frac{2}{3}\frac{4kT}{g_{\mathrm{m}}} \qquad (\mathrm{V}^2/\mathrm{Hz}). \tag{3.78}$$

The rms noise in a given bandwidth is $\sqrt{S(f)\Delta f}$. For two transistors in a diff-pair with equal current the rms value of the input noise is increased by $\sqrt{2}$. An approximate value of noise in a bandwidth of 10-MHz for a g_{m} of 4-mA/V is

$$S(f)\Delta f = \frac{2}{3}4kT(250\text{-}\Omega)(10\text{-MHz}) = 2.76 \times 10^{-11} \qquad (\mathrm{V}^2). \tag{3.79}$$

Therefore, V_{noise} is the square-root of $S(f)\Delta f$, and it is equal to 5.25-μV. The noise of a diffpair is $\sqrt{2}$ greater or

$$V_{\mathrm{noise,diff}} = 7.43\text{-}\mu\mathrm{V}. \tag{3.80}$$

This noise is significantly smaller than we need to meet 60-dB dynamic-range. For a signal of strength 1-V we only require 1-mV of noise. Although this is a rough estimate and does not include other sources of noise in the VGA, it justifies our statement that *the dynamic-range is dominated by the distortion and not the noise.*

Despite the fact that the noise calculations are straightforward, most designers will simulate the circuit to find out the equivalent noise. In HSPICE, the noise calculation are done in conjunction with the ac statement, which is a function of the DC operating points. Therefore, when the operating points of the circuit are drifts, the noise figure may be different. We used HSPICE simulations as a final check of the noise performance.

3.4 Determination of Values

Now that the major constraints of the circuit have been discussed in the previous sections, we can explain how to select the component values for the VGA.

3.4.1 Power Dissipation

In terms of power, it is mainly determined by the output stages or the total resistance used in the feedback resistors. When the single-ended output is 2-V_{pp} and the loading resistance is 2-kΩ, the output current is 500-μA. If the modulation index is

1/3, the bias current for output stage is 1.5-mA. As there are four output stages, the total current is 6-mA.

For a simple transistor (see Fig. 3.12), the small signal parameters equations are

$$g_{\mathrm{m}} = \sqrt{2I_{\mathrm{d}}\mu_{\mathrm{n}}C_{\mathrm{ox}}\frac{W}{L}}, \tag{3.81}$$

$$r_{\mathrm{o}} = \frac{L}{\lambda' I_{\mathrm{d}}}, \tag{3.82}$$

$$A_{v,\mathrm{open}} = g_{\mathrm{m}} \cdot r_{\mathrm{o}} = \frac{1}{\lambda'}\sqrt{\frac{2\mu_{\mathrm{n}}C_{\mathrm{ox}}WL}{I_{\mathrm{d}}}}, \tag{3.83}$$

$$\mathrm{BW} = \frac{1}{r_{\mathrm{o}}C_{\mathrm{total}}} = \frac{\lambda' I_{\mathrm{d}}}{LC_{\mathrm{total}}}, \tag{3.84}$$

$$\mathrm{GBW} = A_{v,\mathrm{open}} \cdot \mathrm{BW} = \frac{g_{\mathrm{m}}}{C_{\mathrm{total}}} = \frac{\sqrt{2I_{\mathrm{d}}\mu_{\mathrm{n}}C_{\mathrm{ox}}\frac{W}{L}}}{C_{\mathrm{total}}}, \tag{3.85}$$

where C_{total} is the total capacitance at the output node. In order to have enough gain-bandwidth product and good linearity, the bias currents of most transistors range from 100-μA to 1-mA. The rest of sub-circuits consume 5-mA, totally. As a result, the total power consumption is 55-mW. (However, there is a mistake found in estimating the size of a resistor which will be discussed in Section 5.1.2. All the bias currents are therefore boosted by 50% so the actual power consumption is around 78-mW.)

3.4.2 Aspect Ratios of CMOS Devices

The aspect ratio is (W/L) of a MOS transistor. Table 3.2 shows the aspect ratios of the transistors in each sub-circuit.

In general, L should be the minimum drawn length such that the bandwidth (or the gain-bandwidth product) can be maximized. When the circuit is operating at high frequencies, the open-loop gain $A_{v,\mathrm{open}}(s)$ is related to the gain-bandwidth product (GBW) rather than the low-frequency open-loop gain ($A_{v,\mathrm{open}}$) stated in (3.83); and a large open-loop gain will improve the circuit performance in various aspects, [14, Chapter 1]. However, other issues make the choice of non-minimum L attractive. The output impedance and the device matching improve with larger L. In this design, as obtaining gain at 10-MHz can be achieved with non-minimum L, devices with gate-lengths larger than 0.8μm were used to increase DC gain and improve matching.

The width W will also affect the characteristics of a transistor. A larger W will give a larger g_{m} (or a larger $A_{v,\mathrm{open}}$); while the effective gate-source voltage V_{gt} in (3.9) will be smaller such that the circuit will have a reduced linear input range. However, the increase of W will also increase all parasitic capacitances associated with the width (e.g., C_{gs} and C_{gd}). This means that there is an upper limit of the gain-bandwidth product. If W is made too wide, the signal bandwidth may be greater than the maximum operating frequency of the VGA and the advantages from applying negative feedback will be lost.

Table 3.2 The selected aspect ratio of the transistors in each sub-circuit.

Sub-Circuit Name	Location	Aspect Ratio (in μm)
Bias Circuit	$I_{\text{bias}} = 100$-μA	$(10/1.4)_{\text{NMOS}} \times 4$ $(20/1.4)_{\text{PMOS}} \times 4$
	$I_{\text{bias}} = 150$-μA	$(10/1.4)_{\text{NMOS}} \times 6$
Input Voltage Buffer	Current Mirror	$(20/1.4)_{\text{PMOS}} \times 4$
	Differential Pair	$(10/1.4)_{\text{NMOS}} \times 4$
	Current Source	$(10/1.4)_{\text{NMOS}} \times 32$
	Output Stage	$(10/1)_{\text{NMOS}} \times 40$ $(20/1)_{\text{PMOS}} \times 40$
	Compensation C_c	2.0-pF
	Compensation R_c	$(20/1)_{\text{PMOS}} \times 3$
Multi-Input-Stage (There are 10 G_m cells.)	Differential Pair	$(15/1.4)_{\text{NMOS}} \times 8$
	Current Mirror	$(10/1.4)_{\text{NMOS,Master}} \times 4$ $(10/1.4)_{\text{NMOS,Slave}} \times 44$
Telescopic-cascode Stage	Upper Stack	$(20/1.4)_{\text{PMOS}} \times 20$
	Middle Stack	$(20/1)_{\text{PMOS}} \times 20$
	Lower Stack	$(10/1)_{\text{NMOS}} \times 20$
Differential Output Stage	Transistor	$(10/1)_{\text{NMOS}} \times 40$ $(20/1)_{\text{PMOS}} \times 40$
	Compensation C_c	2.5-pF
	Compensation R_c	$(20/1)_{\text{PMOS}} \times 2$
Feedback Resistor	Each Total Resistance	4.68-kΩ
CMFB Circuit	Diode	$(20/1.4)_{\text{PMOS}} \times 4$
	Differential Pair	$(10/2)_{\text{NMOS}} \times 4$
	Current Source	$(10/1.4)_{\text{NMOS}} \times 4$
Control Circuit (It contains 11 diffpairs.)	Current Source	$(20/1.4)_{\text{PMOS,Odd}} \times 4$ $(22/1.4)_{\text{PMOS,Even}} \times 4$
	Differential Pair	$(30/1)_{\text{PMOS}} \times 8$
	Current Mirror	$(10/1.4)_{\text{NMOS}} \times 4$

Chapter 4

Layout Considerations: Physical Representation of a Circuit

After the initial design has been chosen and preliminary optimization performed, the design must be converted to a physical layout. The netlist for pre-simulation may be relatively simple. When it is transformed into a physical representation, a lot of parasitic components and unwanted secondary effects are introduced, and the circuit performance will be degraded. In order to minimize degradations and achieve all of the expected performance, proper layout techniques should be employed and proper layout precautions should be taken. After the circuit is laid out, it is essential to extract the parasitics and perform post-layout simulations to reoptimize the circuit. This chapter will talk about various aspects related to layout and the technique for improving the circuit performance through proper biasing.

4.1 Differences between Simulation and Layout

In simulation, some of the parasitic components may not have been considered. For example, if the branches from different sub-circuits are connected to one single node, the wiring resistance will be neglected and the node capacitances may not be accurately accounted for. Moreover, process gradients, that occur during fabrication, may not be given serious consideration unless problems are anticipated.

In this section, the differences between ideal and post-layout simulations are discussed. The following are a primary source of errors.

Device Models Fabrication is a random process. The model of a device is just a mathematical approximation of its physical characteristics, with parameters obtained from statistical averaging. Most models cannot exactly reproduce each device's performance at different bias conditions. Although simulations are used to predict the general behavior of the circuit, the actual performance will deviate from the model profile.

Boundary Dependent Etching When patterns are drawn on the wafer for positive etching, material on the unwanted area are often over-etched. It turns out that the resulting pattern is smaller, or thinner than the drawn size. This would affect the effective W/L ratio of a transistor and the actual resistance of a resistor.

Furthermore, the undercut is boundary dependent. Closely spaced patterns will experience a different undercutting mechanism than patterns spaced farther apart. This is illustrated in Fig. 4.1. If necessary, this effect can be

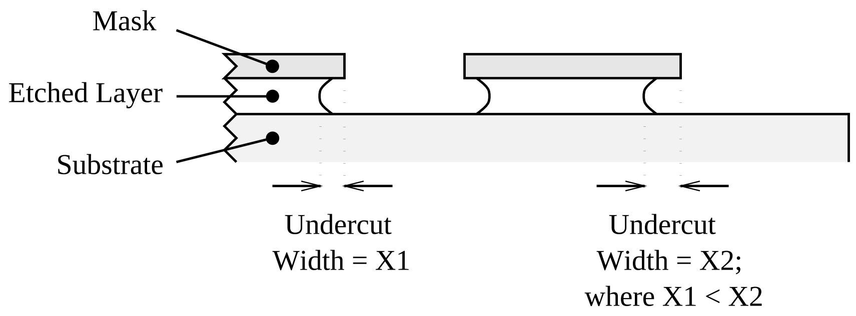


Figure 4.1 Boundary dependent etching problem.

considered in simulation; the floor plan of the layout should be given. Nevertheless, perfect prediction is almost impossible, but Monte Carlo simulations can prove very useful.

Parasitic Components Parasitic resistances and capacitances exist in all routing wires and their values vary significantly with respect to temperature and processing. Such parasitics would degrade circuit performance. In pre-simulation, they are usually neglected. Even if some of them would be considered in post-simulation, their values are rarely accurate.

In general, many potential problems can be minimized by a careful layout. For a fully-differential circuit, the layout should be symmetric such that both sides would suffer the same degradations; then some unwanted effects will appear as a common-mode signal and may be canceled out.

There are many other factors that affect circuit performance. These are often not very substantial for digital systems; but they may have a great influence on the performance of analog circuits. Therefore, the analog designers must be particularly careful in doing physical layouts, and precautions should always be taken to suppress these parasitic effects.

4.2 Physical Layout

Capacitors, resistors and transistors are the basic components in a CMOS analog circuit. This section will discuss some of their layout considerations. Bonding pads and packaging will be studied in the next section. For a more detailed discussion of layout issues, guidelines for a good layout have been discussed in [15, Chapter 16] and [16, Chapter 11], and will not be discussed here.

4.2.1 Layout of Capacitor

A capacitor is formed between two overlapping plates. The oxide thickness between the plates will define the capacitance per unit area. The highest capacitance per unit area is always found between the polysilicon and the active area as the gate oxide thickness is always the thinnest. However, if this capacitor is used, the bottom layer is always biased at V_{dd} or V_{ss}.[1] In addition, the capacitance is slightly voltage dependent, and the bottom-plate will pick up noise from the substrate easily, like an antenna. This capacitor is suitable, however, when one of its terminals is connected to the ac ground, and high linearity is not required.

In general, there are two types of capacitors: grounded, and floating. The grounded capacitor has one terminal connected to the ac ground. The floating capacitor is connected between two floating nodes of a circuit, such as the compensation capacitor of an opamp.

Floating Capacitor Since the capacitor is obtained by overlapping two layers, it actually contains three capacitances, as depicted in Fig. 4.2. C_1 is the designed

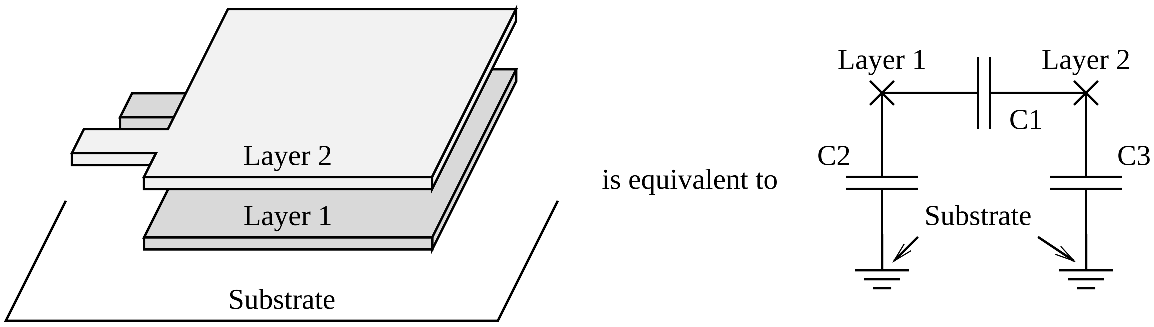


Figure 4.2 A physical capacitor by overlapping two plates.

capacitance. C_2 and C_3 are parasitic capacitances. Because Layer 1 is closer to the substrate, C_2 is larger than C_3 ($C_2 > C_3$). Sometimes, if the distance between Layer 2 and the substrate is far enough, C_3 may be negligible.

If multiple layers are available, a capacitor can be formed by sandwiching, as shown in Fig. 4.3 and the area required will be smaller. In a floating capacitor the parasitic capacitance to the substrate, C_2, can be a problem. It is not uncommon for the parasitic, C_2, to be similar in magnitude to the total capacitance, C_1, in which case circuit performance may be adversely affected. As the first polysilicon (POLY) layer is the closest to the substrate, the parasitic capacitance of the POLY layer to the

[1] This capacitor can easily be obtained from a transistor layout with large W and large L. Both drain and source connect to V_{dd} for PMOS or V_{ss} for NMOS.

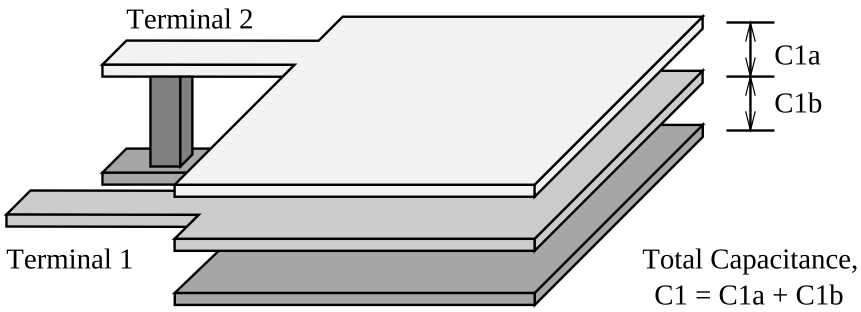


Figure 4.3 A capacitor by sandwiching.

substrate will be greater than for any of the subsequent metal layers. If the substrate parasitic needs to be kept small, the first POLY layer should not be involved in the capacitor sandwich. As a result, the capacitor will be formed by only metal layers and will require a larger area for the same capacitance.

Precaution The designer needs to be aware of the substrate parasitic of a floating capacitor. For example, for a Miller compensation capacitor, the terminal with C_2 should connected to the output of the gain stage such that noise and unwanted signals coupled from C_2 will not be amplified. Furthermore, the output terminal generally has better driving capability, so the bandwidth is not reduced as much.

4.2.2 Layout of Resistor

Integrated resistors are obtained by using suitable resistive layers, such as: P- or N-well, N-diffusion, P-diffusion and polysilicon. In practice, we do not get access to a pure resistive element. There will always be parasitic capacitance coupling with other sub-circuits and the supplies. Therefore, when we lay out a resistor, we should try to prevent noise and unwanted signals from coupling into the resistor.

Comparing all the available choices of resistive layers, the POLY layer is often used as a signal-path resistor, especially when the circuit is operating at high frequencies. The POLY layer has the lowest coupling capacitance and the least interaction with the substrate. This layer also results in the most linear resistor because it does not suffer from width modulation, as a function of voltage, as do diffused resistors.

Layout Techniques & Precautions One of precautions in eliminating noise coupling into the resistors is to put a substrate ring around the resistor, such that the substrate potential is localized. (Sometimes, a well is laid out beneath the resistor such that the resistor is further isolated from other circuits. However, the effectiveness depends on how clean the supplies are.)

Process Variations The resistivity of any resistive layers will have a variation of approximately ±30%. The required resistance and the actual resistance may not match well. If an accurate resistor is needed, advanced techniques (such as laser trimming) should be employed. It is better to design circuit such that performance depends on the relative resistance ratios, rather than exact values. Two layout examples of resistor are shown in Fig. 4.4. The following list some layout precautions.

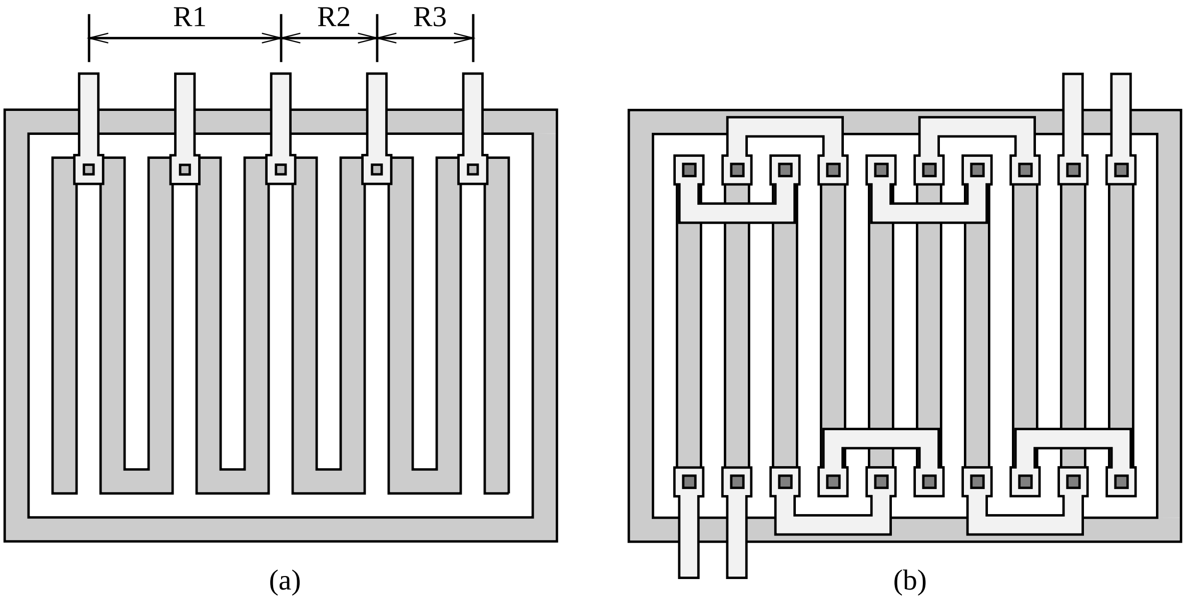


Figure 4.4 Resistor layout examples.

- In order to achieve matched boundary etching, a dummy ring should be put around the resistor and it should be grounded.
- For a series of resistors with an integer ratio (e.g., $R_1 : R_2 : R_3 = 2:1:1$), the layout in Fig. 4.4a is suggested since they are constructed by an unit resistor. (If a unit resistor is not used, resistors may depend on their relative lengths and good matching is difficult to be achieved.)
- For two matched resistors, the layout in Fig. 4.4b should be used. For high frequency applications, the separation distance of the resistive strips should be increased, or dummy grounded resistors should be added in between every two resistive strips, to reduce signal coupling.
- If possible, the width of the resistive strip should be kept uniform. This requirement should be enforced. Since etching deviations are a strong function of width, matching any two devices of different widths in not advisable. Since the ohmic contacts will be located on the resistor for making connections, their dimension will define the minimum width of each strip.

4.2.3 Layout of Transistors

The layout of a transistor consists of two overlapping rectangles, one is from the polysilicon layer and the other is from the active/diffusion area. The overlapping area defines the aspect ratio of the transistor. CMOS transistor layouts are symmetrical; there is no distinction between source and drain. In NMOS, the terminal with the higher potential is defined to be the "Drain" and the other one is the "Source"; and vice versa for PMOS. Simulation tools will generate the correct device model accordingly. It is not necessary to define "Drain" and "Source" explicitly in the netlist.

Composite Transistor In most analog applications, the aspect ratio of each transistor is fairly high. In such a case, each transistor should be separated into several equal parts and connected in parallel. A composite transistor is then formed. However, the new effective W/L ratio will be different from that of a single transistor. In simulation, the netlist should be generated according to the actual layout. For example, a transistor with $W = 40$ and $L = 1$ might be more easily drawn as 4 transistors in parallel, all of width, $W = 10$.

Layout Precautions & Matching Considerations The use of the unit transistor concept can improve the matching accuracy. An example for the two current sources with the current ratio of 1 : 1.5 is shown in Fig. 4.5. The scaling of the

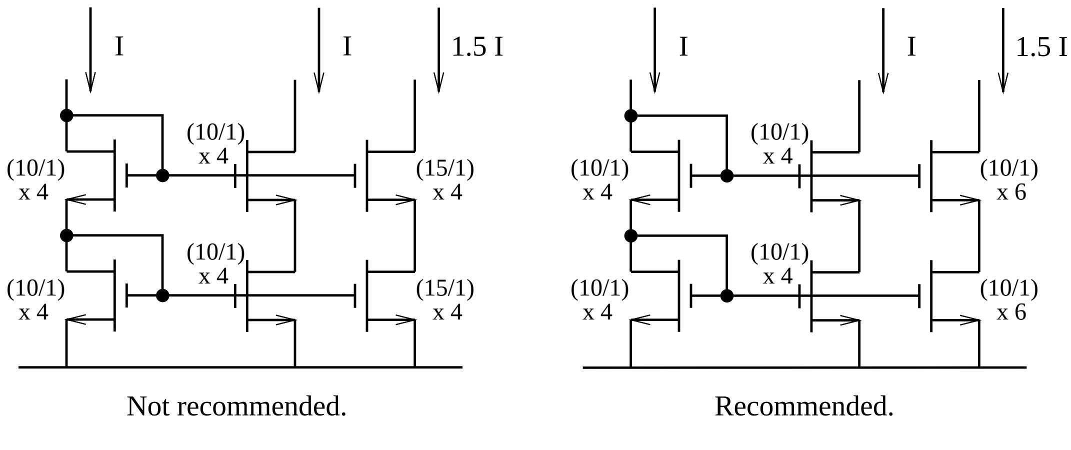


Figure 4.5 Current sources with the current ratio of 1 : 1.5.

aspect ratio directly to achieve this current ratio would not be as effective, since transistor models vary with width. If the aspect ratio is scaled up by 50%, it does not mean the current will be scaled by 50%, also. A better alternative to maintain good matching is to increase the number of the unit transistors. If the number of the unit transistors is increased from 4 to 6, it is equal to one and half of the reference current sources working in parallel or the current is exactly scaled up by 50%. Methods exist for minimizing transistor area, parasitic capacitance and resistance [17]. This methods should be applied as deemed necessary.

4.2.4 Bonding Pads and Packaging

Bonding pad and packaging are important components in an integrated circuit, as all connections of signal and power must be transmitted through them.

Packaging All the bonding wires and the bonding pins in a package will have resistance, inductance and capacitance. If a die connects to a package improperly, the input and the output signals may oscillate. Sometimes, signal cross-coupling can happen. Good pin assignments will minimize all these unwanted effects.

Bonding Pads External circuitry will present a high loading capacitance to the I/O pins, and the I/O pins should be assigned to have enough diving capability. For digital circuits, we have standard rules for designing I/O circuitry [18, P. 357–368], but, unfortunately, there are no such rules for analog circuits. Only diodes and a resistor (see Fig. 4.6) are used for electrostatic discharge (ESD) protection in the prototype VGA.

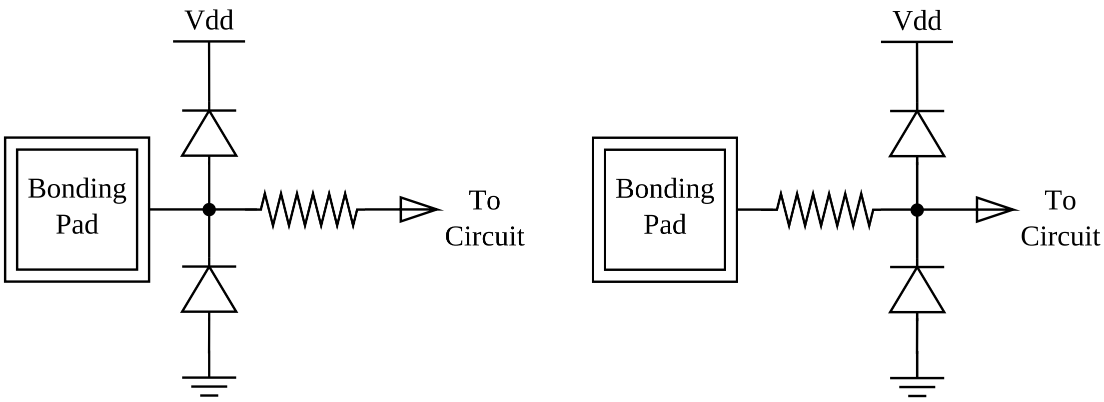


Figure 4.6 Analog pads with diodes and resistor for ESD protection.

Connections to the bonding pads and packages should be considered in all designs. Precautions must be taken in designing circuits with large loads. If the input terminal of a circuit draws a large current, a voltage buffer (e.g., the one suggested in Section 2.8) might be used. Conversely, an output terminal should have enough driving capability to drive external resistors and/or capacitors.

For tapping out a voltage signal, the connection between the bonding pad and the node in the circuit should have minimum wire width. For simplicity, the RC constant of a wire is always a constant. If the wire width is doubled, the resistance will be reduced by half, but the capacitance will be scaled up by 2. In some cases, the degradation due to capacitance is larger than that due to resistance and then the node capacitance should be kept minimum.

For tapping out a current signal, attention must be paid to the design rules for the current density because of the electromigration [19, pp. 104].

The connection between the bonding pad and V_{dd} or V_{ss} of a circuit should be wide enough such that the connection resistance can be minimized and the supply currents will not cause a large voltage drop. Otherwise, this could result in the signal coupling through the power line.

4.3 Bias Circuit

Biasing is very important in most analog circuits. It controls and readjusts the overall circuit performance.

Considerations of Voltage Drop Circuits may be biased using either voltage or current. However, when a bias voltage is connected from the bias circuit to a sub-circuit, there could be voltage drop due to connection resistance; the performance of the sub-circuit may be different from what was designed. This means that using voltage for biasing may be problematic.

It is generally preferred to transport bias quantities as currents and then reconstruct the bias voltage locally. For current biasing, even if a voltage drop appears across the connection, the operation points of a sub-circuit would not be changed as bias voltages are generated locally. At the same time, this local generation method alleviates some problems due to process gradients.

However, currents cannot be *broadcast*. One current is needed for each bias line. Whereas, a voltage can be broadcast and only one line can bias several circuits. This is a trade-off between reliability and power.

In this design, currents are used for biasing. Techniques for handling supply variation and process gradients [20] have not been implemented because they may give rise to uncertainties in debugging if the circuit performance is out of expectations.

A Flexible Approach for Circuit Biasing The master section of the bias circuit is shown in Fig. 4.7. Other cascode current sources are connected in parallel and generate bias currents for all sub-circuits. In a prototype, if the connections between the master bias circuit and the sub-circuits are also connected to the bonding pads, all bias currents are accessible. Additional amounts of bias current can be injected or subtracted. This technique is depicted in Fig. 4.8.

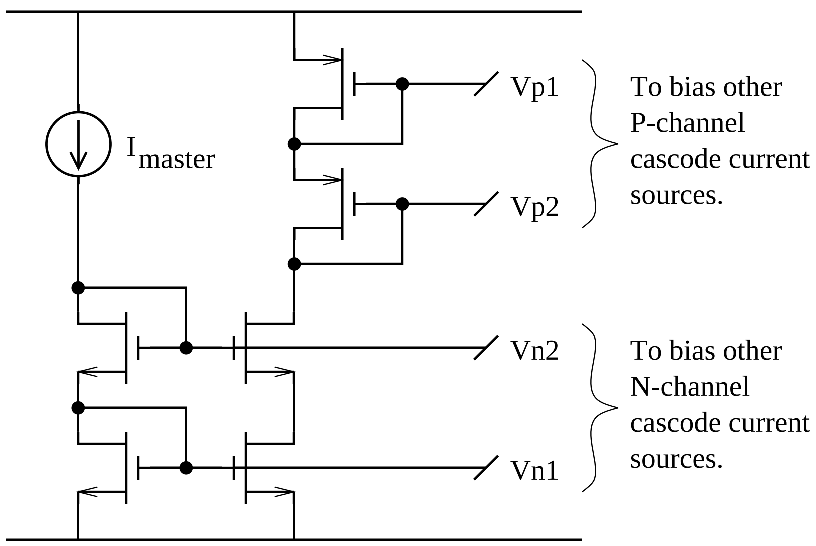


Figure 4.7 The master section of the bias circuit.

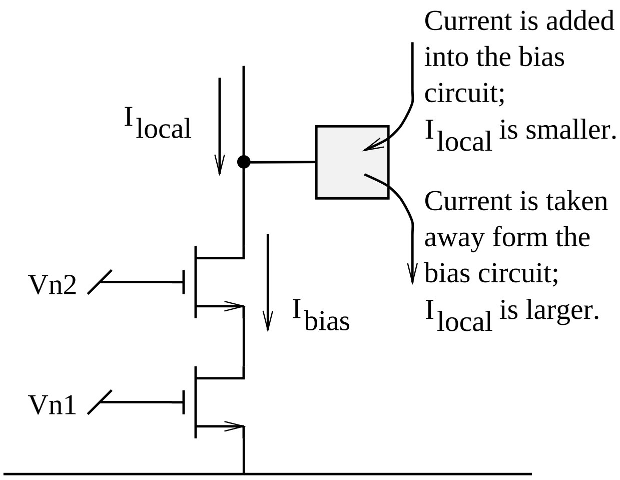


Figure 4.8 Technique for getting access a local bias current.

4.4 Floor Plan

Techniques in realizing a good layout have been discussed briefly in Section 4.2. However, when different sub-circuits are merged together, many other issues should be considered. The floor plan must be revised many times such that space is used effectively and circuit performance would not be seriously degraded due to parasitic capacitance, resistance and signal coupling.

The VGA is fabricated in a 0.8-μm CMOS process (Hewlett Packard CMOS26G) through MOSIS. The active die area is $0.6 \times 1\text{-mm}^2$. The chip is packaged in a 40-pin DIP package. The circuit layout is generated by the GDT Designer tool set which is a product of Mentor Graphics Corporation. All the layouts are described in a programming language. Once the floor plan or sub-circuits are changed, the layout can be regenerated by adjusting some parameters. The floor plan and the die photo of the VGA are shown in Fig. 4.9 and Fig. 4.10, respectively.

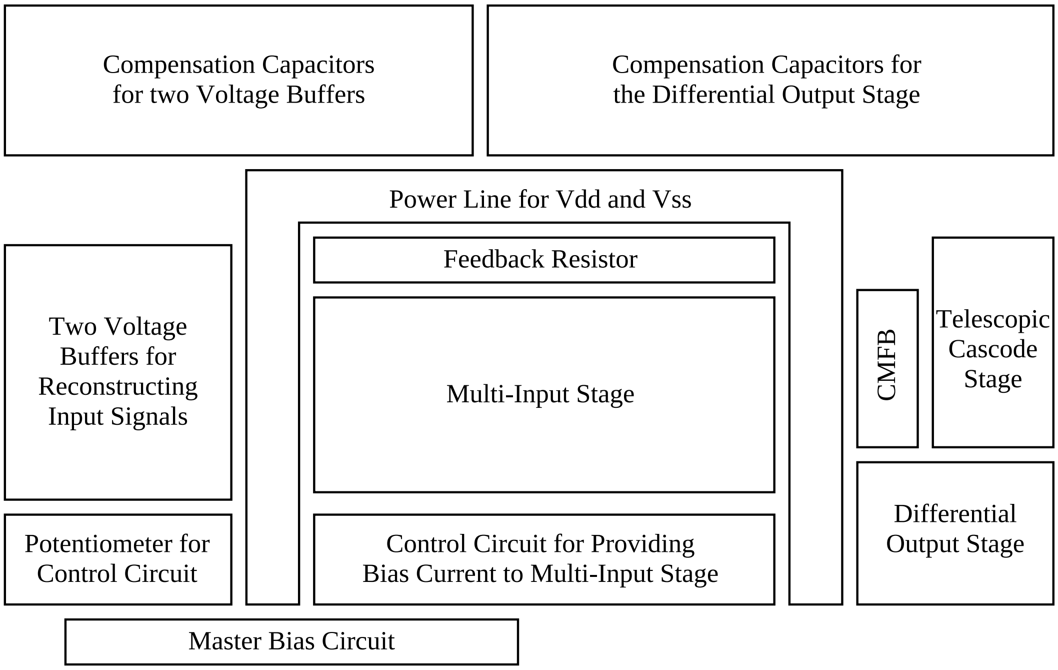


Figure 4.9 Floor plan of the VGA.

Other Precautions for this Prototype VGA In terms of minimizing signal coupling, the following precautions were taken in the layout.

Power Input The power is applied to the bias circuit first. The control circuit is located next to the bias circuit. This arrangement can maintain a very clean power supply for these two circuits. It is very important because the biasing or the circuit operations are initiated by them. If their output signals are not clean enough, noise will be propagated and the overall circuit performance will be affected.

Output Stage All the output stages in this VGA are the Class-A common-source configurations. This minimizes the current transients at the power lines of the output stage, thus reducing the signal coupling from the output stage to the others sub-circuit through the power lines.

Current Handling Capability of Power Lines Among the three available layers of metal, the top metal layer (MET3) is thickest in this CMOS process or it has the best current-handling capability. It was therefore used to route the power lines.

Substrate Biasing Lastly, each composite transistors has its own substrate ring for localization. It would minimize the unwanted signals coupled from its adjacent transistors through the substrate. However, this increases area and may reduce matching accuracy as well.

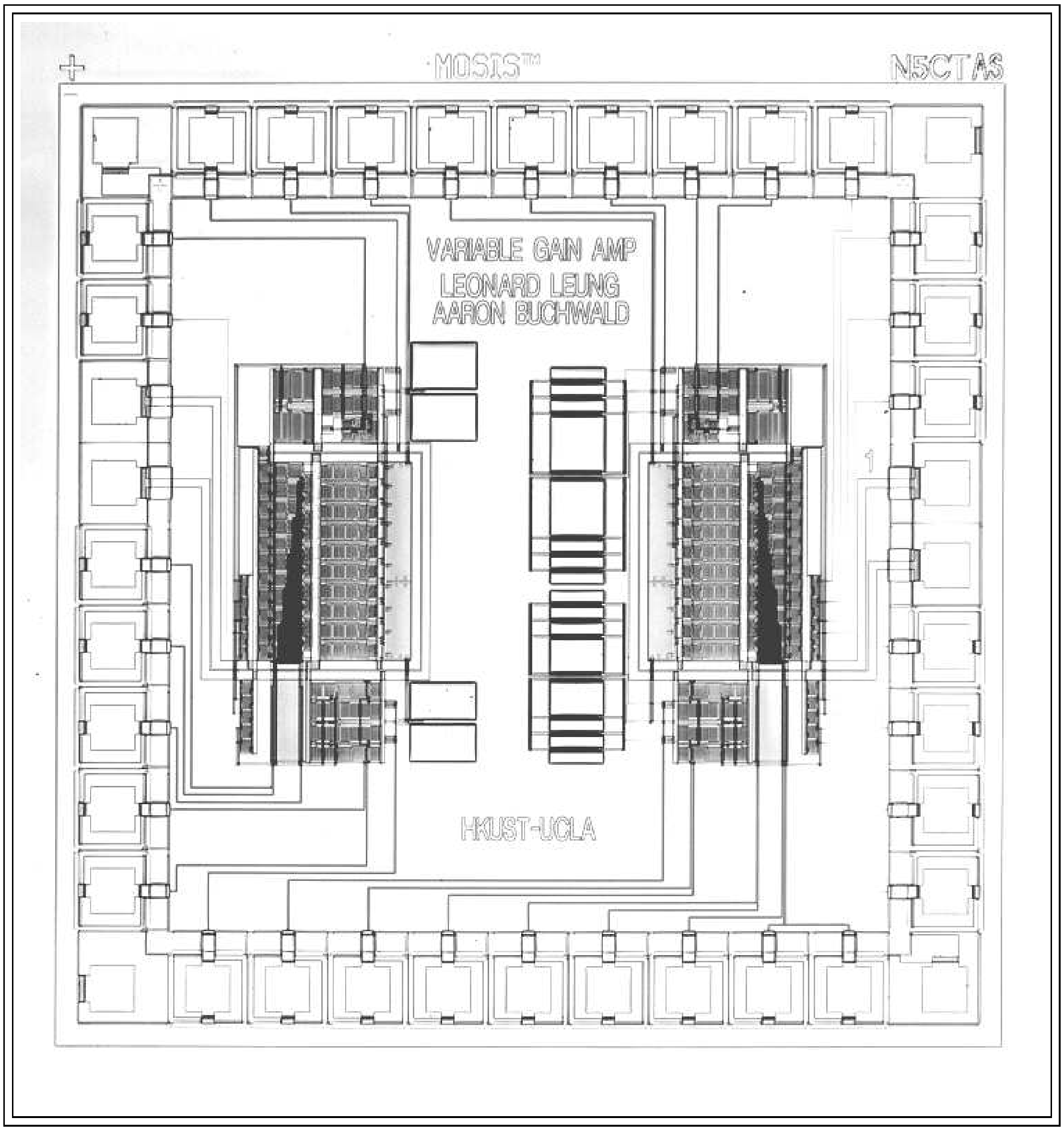


Figure 4.10 Die photo of the VGA.

Test Circuits on the Die It is very helpful to layout diagnostic circuits on the die if possible. Referring to the die photo in Fig. 4.10, two similar circuits are fabricated on the same die; their compensation capacitors are different. For the circuit on the right, each compensation capacitor consists of a series of capacitors connected in parallel. Their capacitances can be adjusted by laser cutting. All outputs from its bias circuit are connected to bonding pads. This prototype is used for debugging and finding the optimal compensation capacitance value. For the circuit on the left, the compensation capacitors are optimum in term of the THD performance, which were found by simulation. To minimize pin count and remain in a small package, the outputs of the bias circuit have not been connected to any bonding pads in this duplicate circuit.

Chapter 5

Simulation and Testing: Verification of VGA Performance

In this chapter, the simulation results will be presented and compared to the measured results of a prototype VGA fabricated in a 0.8-μm CMOS process. The novel idea of this research is the continuous gain tuning, which is accomplished by interpolation between 10 taps of a feedback resistor, respectively connected to a multi-input-stage operational amplifier (opamp). Of primary concern is the distortion; the VGA must be designed to achieve a dynamic-range greater than 60-dB. It will be shown that the VGA meets, or exceeds the desired specifications. In the closing section, ideas for improvements to the design are outlined.

5.1 Design and Simulation

5.1.1 Approach

Optimization of the Multi-Input-Stage In practice, each sub-circuit will contribute harmonic distortion. In order to study the multi-input-stage only, all other sub-circuits are replaced by ideal components. Ideal signal sources are connected to the inputs of the VGA, which drive two feedback resistors directly. Two ideal current-dependent voltage sources (transresistors) are used as the output stage. The schematic of the opamp with a non-ideal multi-input-stage is shown in Fig. 5.1.

The opamp in Fig. 2.2, which has two transconductance cells (G_m cells) as an input stage and a differential feedback path, is used to prove the general idea described in Section 2.1. Next, the number of G_m cells will be increased from two to four for investigating the relationship between THD and the supplied bias currents.

The simulation results show that the THD is lowest if only two consecutive G_m cells are switched on with an equal amount of bias current. The THD rises when either one of these two cells has a very low bias current, as was discussed in detail in

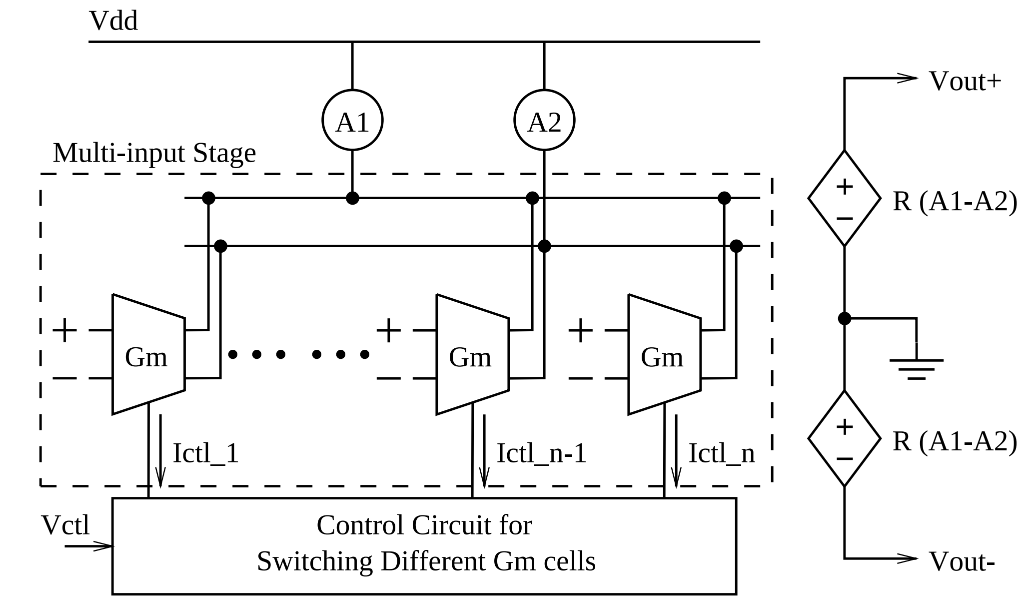


Figure 5.1 Schematic of an operational amplifier with a non-ideal multi-input-stage.

Section 3.1.3. The THD under this situation can be reduced if the third G_{m} cell is switched on earlier. The proper transition from the first to the third cell is depicted in Fig. 5.2.

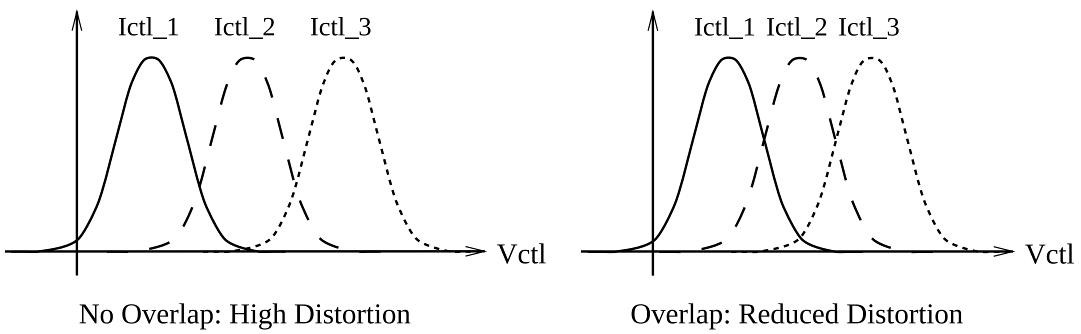


Figure 5.2 The proper current signals for switching three transconductance cells.

Moreover, the distortion is proportional to the third power of $\Delta A_v/|A_v|$. Therefore, to reduce the distortion, $\Delta A_v/|A_v|$ should be reduced between consecutive stages. If the variable gain range is fixed, then reducing $\Delta A_v/|A_v|$ requires the addition of more stages. These preliminary simulations verify that the average THD reduces if the number of the G_{m} cells is increased from four to eight, and reduces again if increased from eight to ten. In order to restrict the die area, only 10 G_{rmm} cells are used.

Switching off all the Unused $\mathbf{G_m}$ Cells On the other hand, the control circuit should ensure all the unused G_{m} cells are switched off. If this is not the case, there will be a highly-distorted current generated from the partially ON cells that will degrade performance. To avid this, an offset current is added to force the bias current off. This was explained in Section 2.9.

The simulated control currents are shown in Fig. 5.3. Each bell-shaped current signal is used to control one G_{m} cell. The simulation was performed when the VGA inputs was tied to the common-mode potential and the control signal (V_{ctl}) was swept from 0-V to 3-V (1.5-V ± 1.5-V) This simulation shows that the control currents indeed drop to zero outside of their active range.

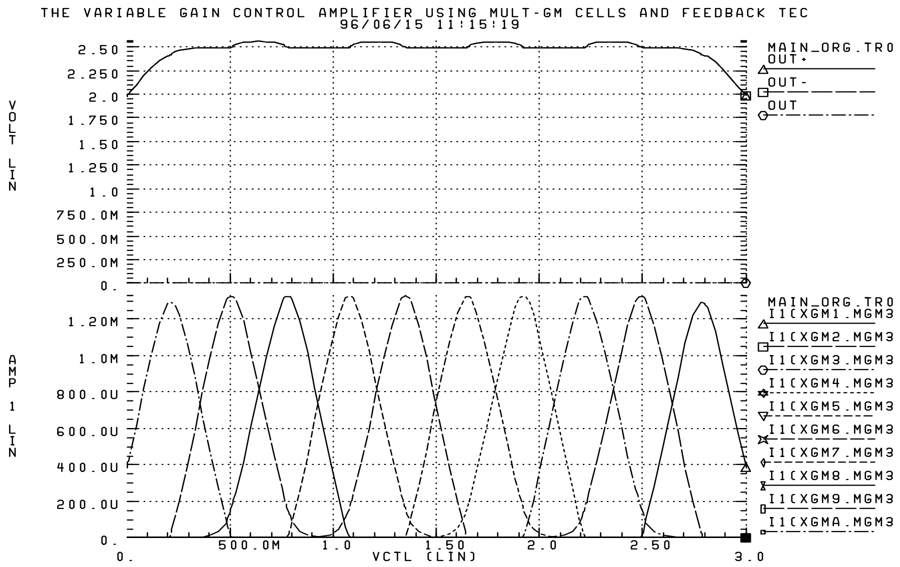


Figure 5.3 The simulated current signals for controlling the transconductance cells in the multi-input-stage.

Other Sub-Circuits After the design of the multi-input-stage and the control circuit were chosen, different output stages were considered. The selected output stage must also have enough current-handling capability and its distortion cannot significantly degrade the overall performance.

In order to prevent the input signals driving the feedback resistors directly, two simple opamps with unity feedback are used as input voltage buffers. Based on simulations, the THD is expected to be in the range from −70-dB to −65-dB differential. Their bandwidth should be larger than that of the complete VGA.

Finally, all the sub-circuits are connected together and extensive simulations are performed for verifying the overall circuit performance. Specific simulations are performed to determine THD, two-tone intermodulation, and closed-loop stability. After an initial layout is completed, the aspect ratio of some composite transistors need adjustments in order to better utilize the die area. After modifications and parasitic extractions, the simulations are repeated and the circuit is reoptimized.

5.1.2 Simulation Results

After the test chip was submitted for fabrication, one serious mistake was found. When the layout was drawn, the sheet resistance of the polysilicon layer was assumed to be 4-$\Omega/\square$. Actually, it should be 2-$\Omega/\square$ which had been verified by measurement. Therefore, the resistance of all resistors are only half of the expected values.

Because the size of each feedback resistor is reduced by half, the performance of all output stages are affected seriously. In order to solve this problem, the master bias current of the VGA is increased from 100-μA to 150-μA. The following simulation results are based on this new bias current.

On the test chip, there are two identical circuits with different values of compensation capacitors. From simulation, the size of the compensation capacitors in the input voltage buffers and in the differential output stage are optimized to be 2.0-pF and 2.5-pF, respectively.[1] The test circuit with these capacitors is called TC_1. Its performance will be studied in detail. The other will be called TC_2, the size of the compensation capacitors in the input voltage buffers and in the differential output stage are chosen to be 3.0-pF and 4.0-pF, respectively.

Also, in order to compare the simulation results with the measured results, the conditions used for simulation and for testing are the same. Unless otherwise stated, the circuit is simulated at a temperature of 25°C and the Berkeley level 4 model is used. Also, the loading capacitance at each output terminal of the VGA is 1.0-pF.

Stability Frequency response and step response of the VGA can determined the stability of the circuit. The variable gain range of the VGA is from 1 to 2; its phase margin is lowest when the closed-loop gain is unity. The following results are obtained when the voltage gain of the VGA is unity.

For TC_1, its step response is shown in Fig. 5.4 when a 4-V_{pp} differential input signal at 10-MHz is applied. ("PAD_XVON" and "PAD_XVOP" are the single-ended responses and "OUT" is equal to (PAD_XVON − PAD_XVOP).) The frequency response is shown in Fig. 5.5.

Fig. 5.4 shows that the overshooting in the single-ended response is very large, but there is little overshoot in the differential output. This indicates that the common-mode feedback circuit is marginally stable with a phase margin of 30°.[2] The phase margin for the fully-differential circuit is approximately 70°.

[1] Referring to Section 3.2, the size of the compensation capacitor affects the THD directly. Since a stable circuit with the lowest THD is expected, the size of the compensation capacitor can not be too large or the circuit is over-compensated, and THD will increase.

[2] For simplicity, most higher-order systems are approximated to be a second-order system for analysis [21, Section 6.5].

An example for finding the phase margin of a second-order system according to the overshooting in the step response and the resonant peaking in the frequency response is illustrated in [22, Fig. 5.33]. On the other hand, referring to [21, (6.5-13)], the relationship between the resonant

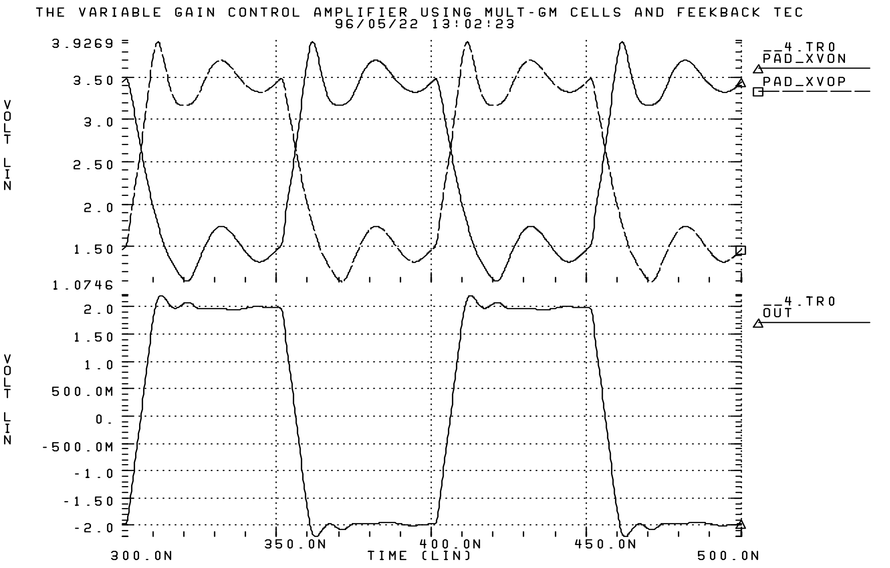


Figure 5.4 The simulated large step response of TC_1 at 10-MHz.

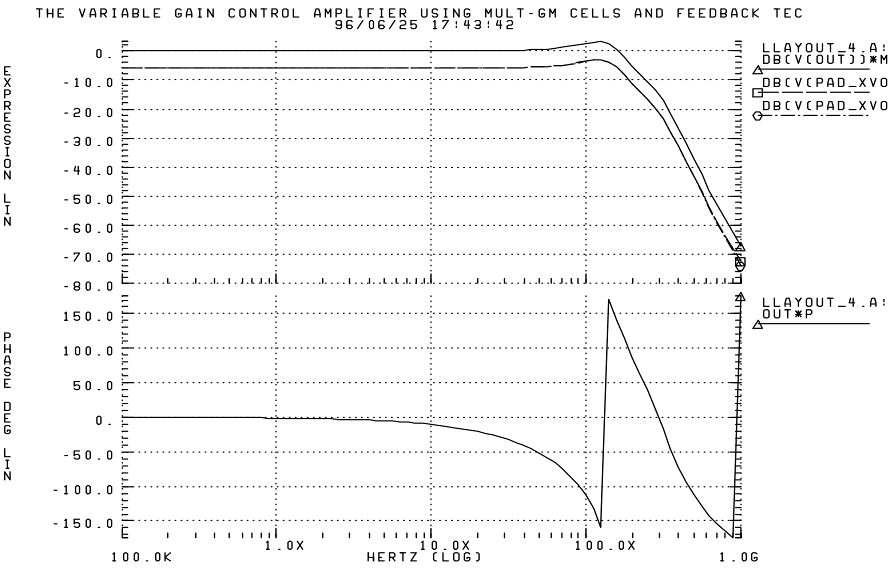


Figure 5.5 The simulated frequency response of TC_1.

These simulations are repeated for TC_2, which has larger compensation capacitors than TC_1. Its step response and its frequency response are shown in Fig. 5.6 and Fig. 5.7, respectively. This shows less overshoot, but also a much longer rise-time which indicates slew-rate limitation.

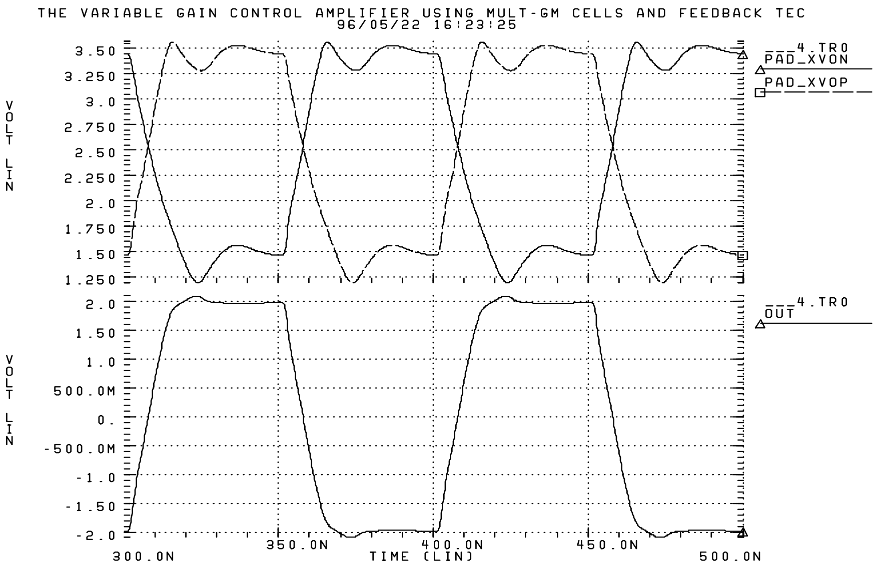


Figure 5.6 The simulated large step response of TC_2 at 10-MHz.

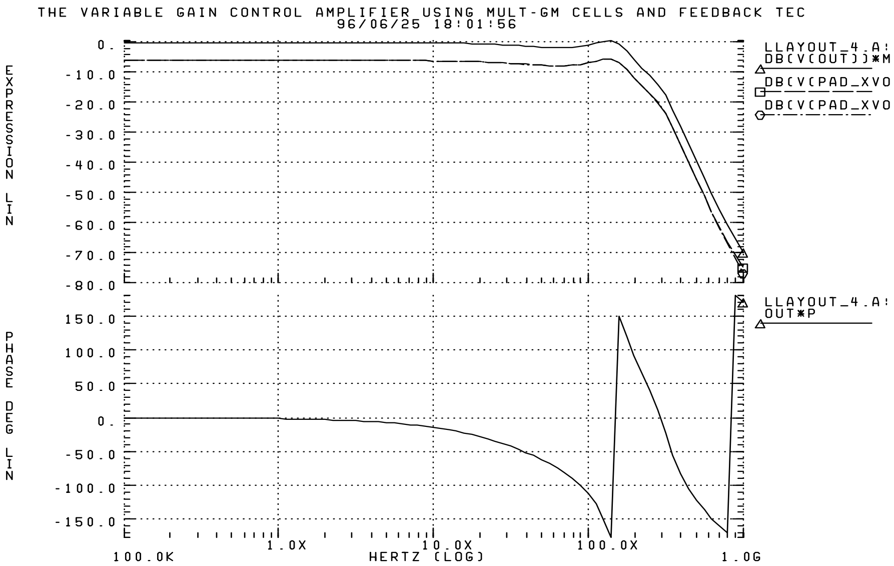


Figure 5.7 The simulated frequency response of TC_2.

peaking in the frequency response (M_f) and the phase margin (ϕ_m) can be written as:

$$M_f \geq \frac{1}{2\sin(\phi_m/2)}.$$

Both step responses in Fig. 5.4 and Fig. 5.6 show that the overshooting in the differential output is reduced because of cancellation. This means that the underdamping is due to the common-mode oscillation and/or the CMFB circuit is undercompensated.

In order to clarify the source of oscillation, the inputs of the CMFB circuit are reconnected to $V_{\text{out}+}$ and $V_{\text{out}-}$. By prediction, if a square wave signal is applied to both inputs of the VGA ($V_{\text{in}+}$ and $V_{\text{in}-}$), both outputs ($V_{\text{out}+}$ and $V_{\text{out}-}$) are zero because the input signals do not have any differential component. However, it is not the case. Once the step is applied, glitches are found in both outputs. This means that the response of the CMFB circuit is not fast enough to track the common-mode error. The simulated common-mode step response is shown in Fig 5.8 when a 10-MHz 2-V_{pp} input signal is applied.

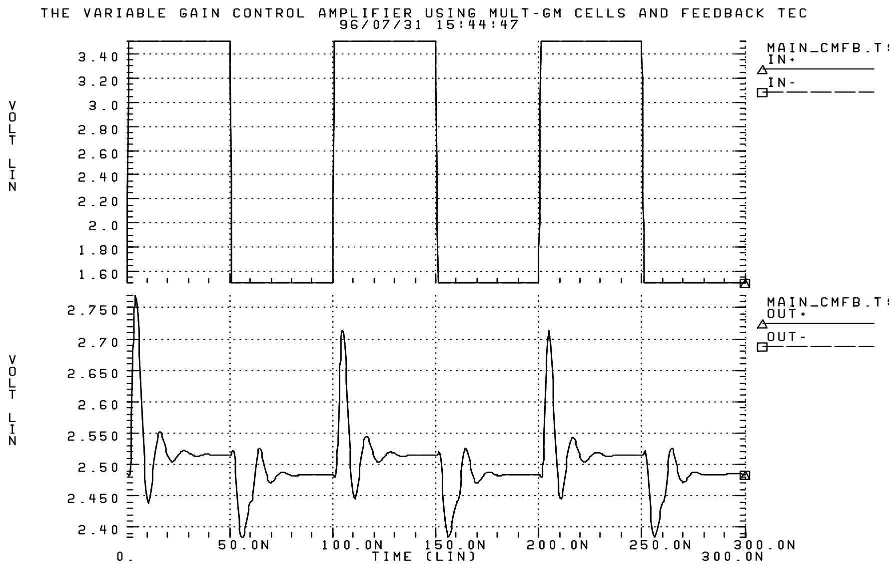


Figure 5.8 The simulated common-mode step response of the modified TC_1.

Open-Loop Frequency Response There are multiple pairs of input terminals in the multi-input-stage opamp. The transconductance of the multi-input-stage is a vector representation (2.1). Instead of simulating the open-loop frequency response of the opamp, the transimpedance from the telescopic-cascode stage to the differential output stage is investigated. The result is shown in Fig. 5.9. Each single-ended output terminal is loaded with a 1-kΩ resistor and a 1-pF capacitor connected in parallel. The size of the compensation capacitors is 2.5-pF. The impedance at DC is more than 6-MΩ. At 10-MHz, the equivalent impedance is 5-kΩ. The second dominate pole and the transmission zero are located at around 250-MHz. Since the multi-input stage has parasitic poles well beyond 250-MHz, this simulation accurately reflects the dominant dynamic behavior of the VGA.

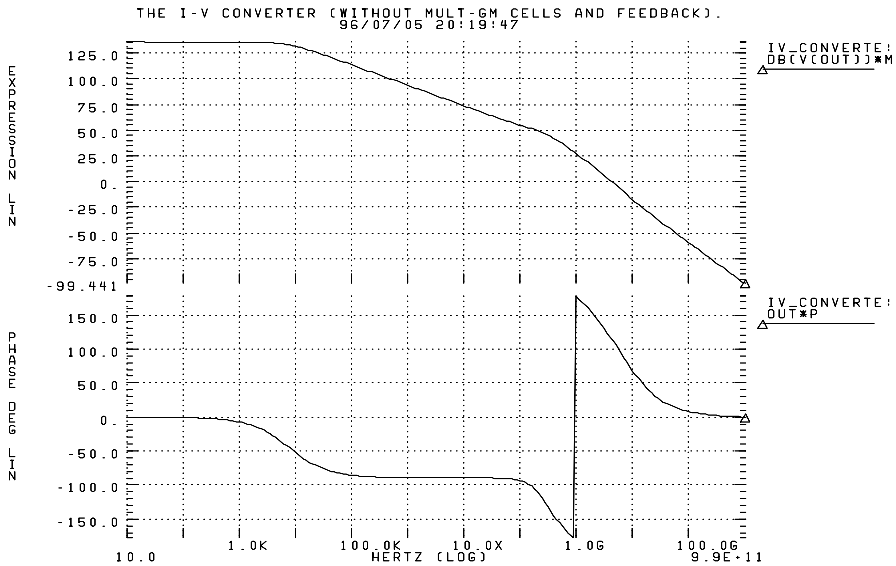
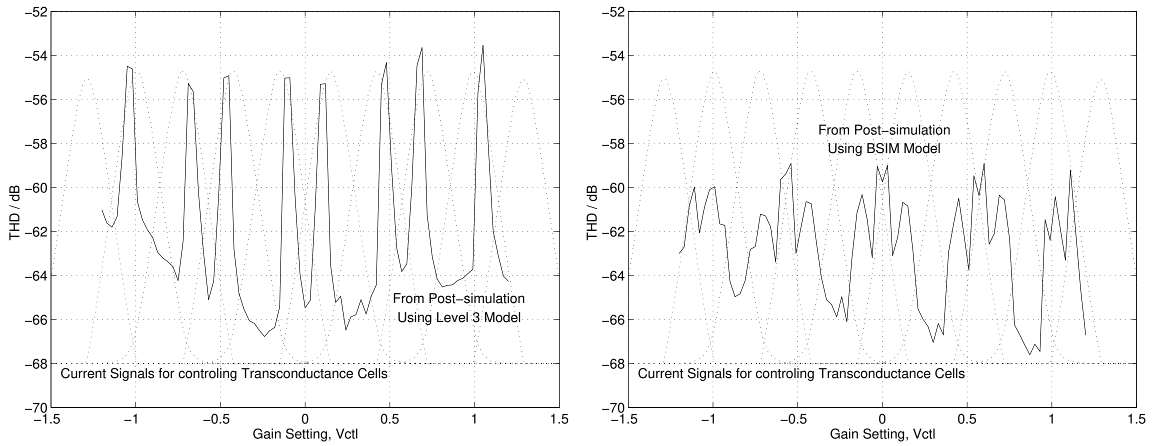


Figure 5.9 The frequency response of transimpedance from the telescopic-cascode stage to the differential output stage. The size of compensation capacitor is 2.5-pF.

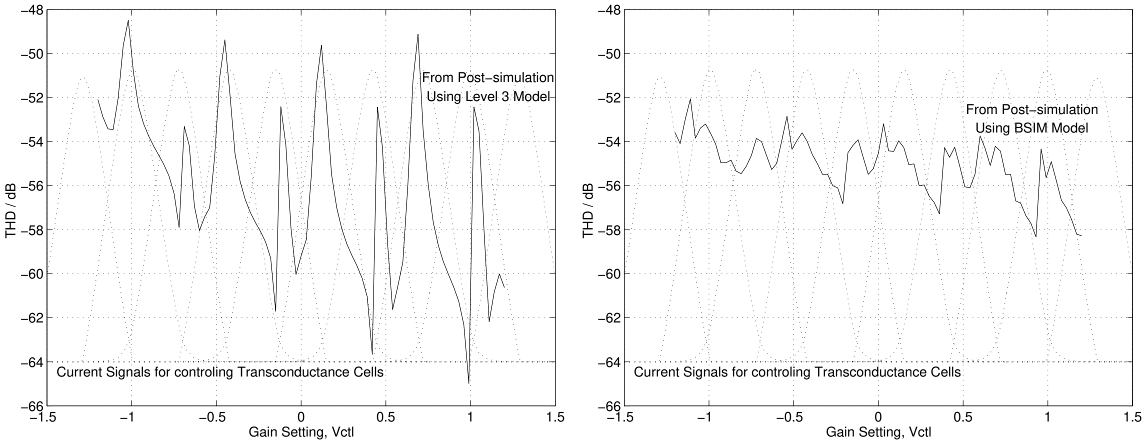
Total Harmonic Distortion The non-linearity of the VGA is investigated by measuring the THD of the circuit. This is the primary limitation on the dynamic-range. After the signal is amplified, its fidelity should be maintained.

Fig. 5.10 and Fig. 5.11 plot the THD of TC_1 at different voltage gain when an input signal at 5-MHz and at 10-MHz are applied. The results are simulated at room temperature. Each single-ended output voltage is 2-V_{pp} and the differential output voltage is 4-V_{pp}. The gain is lowest ($A_v = 1$) when V_{ctl} is equal to -1.15-V; and the gain is highest ($A_v = 2$) when V_{ctl} is equal to 1.15-V.



(a) (b)

Figure 5.10 The simulated THD of TC_1 (a) using level-3 model and (b) using BSIM model. The circuit is tested at room temperature using a 5-MHz signal.

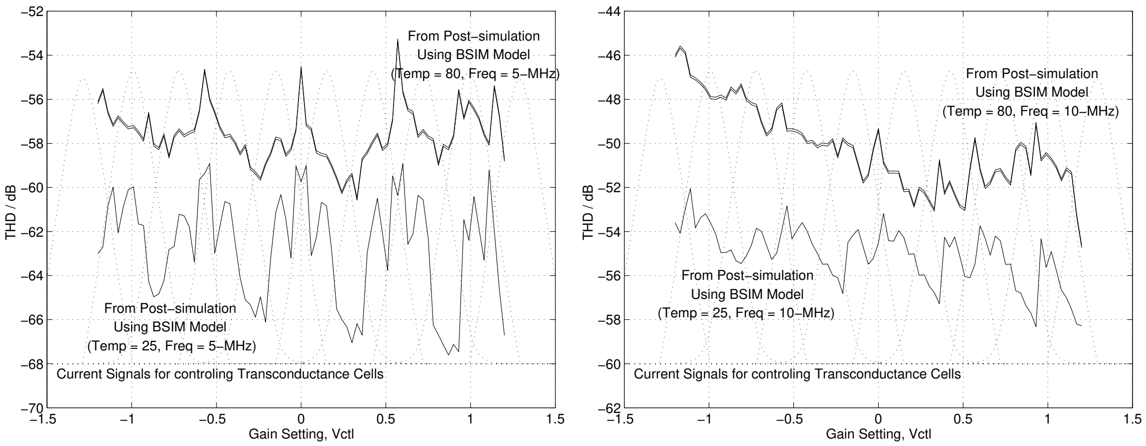


(a) (b)

Figure 5.11 The simulated THD of TC_1 (a) using level 3 Model and (b) using BSIM model. The circuit is tested at room temperature by a 10-MHz signal.

The simulation results are shown that the average THD at 5-MHz and at 10-MHz are found to be -62.24-dB and -55.67-dB, respectively. The numerical data demonstrates that the even harmonics are attenuated. When the two voltage buffers are isolated from the VGA and simulated differentially, the THD at 5-MHz and at 10-MHz are found to be -71-dB and -64-dB, respectively. (These results are obtained when each output terminal is loaded with a 1-kΩ resistor and the differential output voltage is 4-V_{pp}.)

However, when the circuit is operated in a ceramic DIP package in ambient air at room temperature, the estimated die temperature is 80°C. The simulations are repeated and the results are shown in Fig. 5.12. Fig. 5.12 shows that the THD at 80°C increases by 5-dB in average. This means that almost one bit of linearity is



(a) (b)

Figure 5.12 The simulated THD of TC_1 (a) using a 5-MHz signal and (b) using a 10-MHz signal. The BSIM model is used and the circuit is tested at room temperature and at 80°C.

lost. As temperature increases, the mobility of both carriers go down and so does the transconductance of each transistor. Therefore, the gain bandwidth product of the circuit is smaller. Generally, the distortion reduced by feedback is proportional to the open-loop voltage gain. Therefore, the distortion will become larger when the gain bandwidth product drops.

Two-tone Intermodulation In crowded communication systems, signal interference of one device with another is a common problem. For example, third-order intermodulation of two signals often is a problem in narrow-band systems. When two signals (f_1 and f_2) are present in a system, they will mix with the second harmonics and create a third-order intermodulation products which are located close to the original signals. Higher order intermodulation will also occur.

In simulation, two equal amplitude tones at 9.5-MHz and 10.5-MHz are applied to TC_1 such that the peak amplitude of the differential output signal is 2-V_p. The control signal V_{ctl} is set to -1.00-V, 0.00-V and 1.00-V where the voltage gain is close to 1 (low), $\sqrt{2}$ (median) and 2 (high). The third-order intermodulation products (IM_3) are below -59.82-dBc for $V_{ctl} = -1.00$-V ($A_v \approx 1$), below -59.43-dBc for $V_{ctl} = 0.00$-V ($A_v \approx \sqrt{2}$) and below -59.40-dBc for $V_{ctl} = 1.00$-V ($A_v \approx 2$). The spectrum of the output signals for the above three cases are shown in Fig. 5.13.

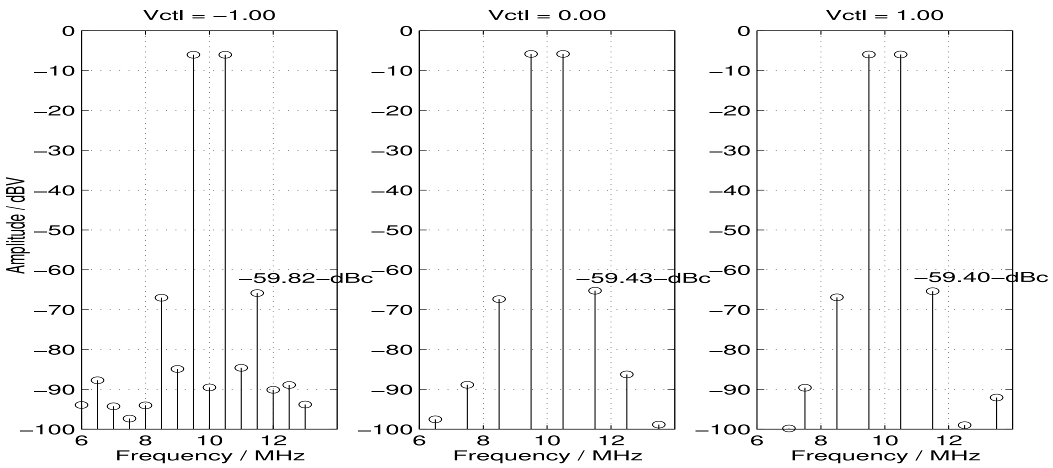
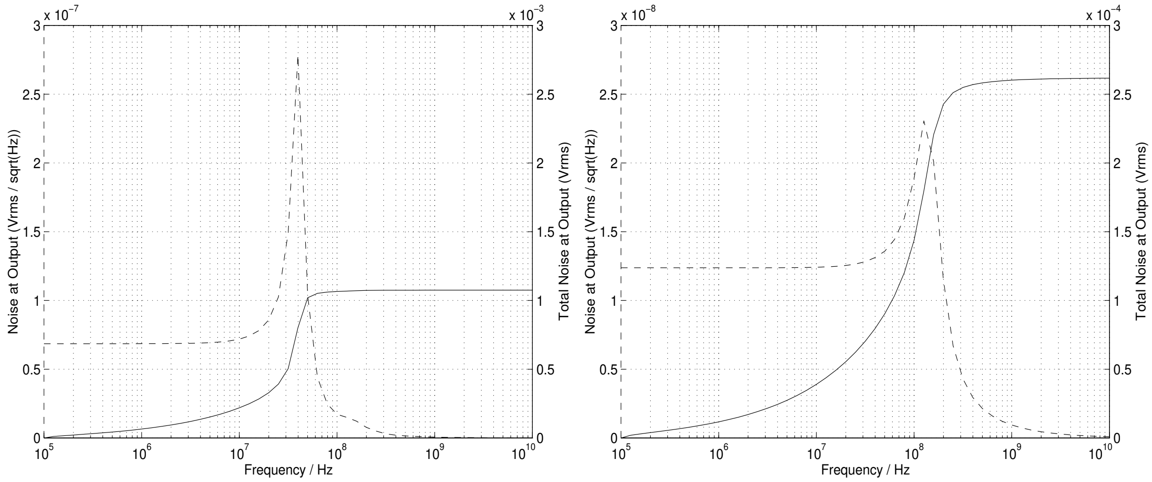


Figure 5.13 The simulated output spectrums for the intermodulation of TC_1.

Noise Noise is always unwanted in a communication system. It is usually unpredictable. It comes from the devices and also from signal coupling. The following simulation only considers the noise from devices.

Firstly, the control signal V_{ctl} is set to -1.00-V ($A_v \approx 1$). For TC_1, the simulated output noise spectral density and the total output noise are shown in Fig. 5.14. The total noise integrated over 1-GHz bandwidth is found to be 1.075-mV_{rms} at each single-ended output and 261.65-μV_{rms} at the differential output. Since the VGA is a

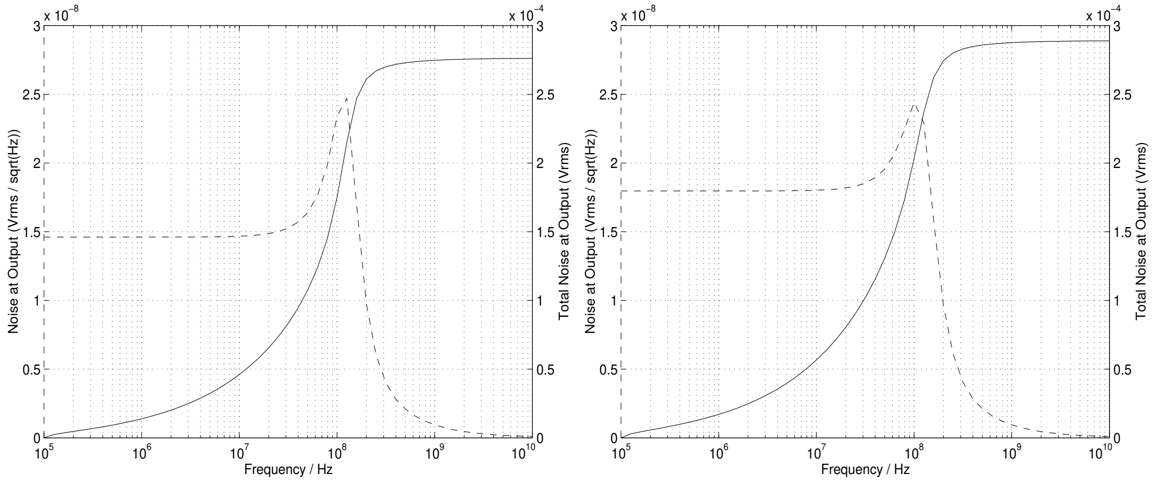


(a) (b)

Figure 5.14 The simulated output noise spectral density (dashed line) and the total output noise (solid line) when $V_{\mathrm{ctl}} = -1.00$-V, (a) single-ended and (b) differential.

fully differential circuit, the common-mode noise (e.g., noise from the CMFB circuit and from the control circuit) will not appear in the differential output.

At the differential output only, when V_{ctl} is set to 0.00-V ($A_v \approx \sqrt{2}$) and 1.00-V ($A_v \approx 2$), the simulated output noise spectral density and the total output noise are shown in Fig. 5.15. The total noise integrated over 1-GHz bandwidth are 276.19-$\mu\mathrm{V}_{\mathrm{rms}}$ and 288.97-$\mu\mathrm{V}_{\mathrm{rms}}$, respectively. From the results, the noise voltage increases with the closed-loop gain; the noise from the early stages will be amplified. However,



(a) (b)

Figure 5.15 The simulated output noise spectral density (dashed line) and the total output noise (solid line) at the differential output when (a) $V_{\mathrm{ctl}} = 0.00$-V and (b) $V_{\mathrm{ctl}} = 1.00$-V.

even if the closed-loop gain is increased from 1 to 2, the total noise only increases by 10.44%. With respect to the nominal signal level of 4-V_{pp} (1.414-$\mathrm{V}_{\mathrm{rms}}$), the signal-to-noise ratio (SNR) of the VGA is above 73.79-dB, and therefore, is not a limitation to achieve a 60-dB dynamic range.

5.2 Circuit Testing and Measurements

5.2.1 Test Equipments

The following are the major equipments used in testing.

Tektronix 11403A Digitizing Oscilloscope Bandwidth up to 3-GHz and a 10-bit resolution. Used for monitoring signals.

Hewlett Packard 8563E Spectrum Analyzer Bandwidth of 30-Hz to 26.5-GHz and its internal distortion products less than −72-dBc for most range settings. Used to display the spectrum of a signal and measure the magnitude of harmonics.

Hewlett Packard 8752C Network Analyzer Frequency range from 300-kHz to 6-GHz. Used to measure the frequency response of a circuit.

Tektronix HFS9003 Stimulus System Can generate a square wave with a frequency from 50-kHz to 630-MHz and an amplitude up to 3-V_{pp}. Used to test the step response of a circuit.

Hewlett Packard 8657B Synthesized Signal Generator Can generate a tone from 0.1-MHz to 2.06-GHz. After the harmonics are attenuated by a LPF, spectrum measurements indicate that the THD of the output signal is less than −65-dB when generating a 4-V_{pp} differential sinusoid at 10-MHz across a 50-Ω load.

The test chip is mounted on a prototyping PCB with various resistors and capacitors for biasing. In order to have wide bandwidth and clean input signals, each high frequency signal pin is connected to the ground plane by means of a 50-Ω resistor. For convenience, the supplies are set to +2.50-V and −2.50-V with respect to the ground plane. In real applications, simple level shifting could be performed. The bypass capacitors are placed within 0.5”, or less, to the power supply pins for maintaining a wideband low-impedance characteristic and insuring the fidelity of high-speed transient signals. The conversion between a single-ended signal and two out-of-phase signals is performed by a power combiner/splitter (Mini-Circuits ZFSCJ-2-2 and ZFSCJ-2-4).

The VGA does not have analog output voltage buffers for driving the external capacitive load. For solving this problem, the prototyping PCB is fixed on a probe station and the output signals are reconstructed by two high speed and high input impedance active probes. The probe is manufactured by the GGB Industries, Inc.

The output signal amplitude is 1/20 of the detected signal amplitude. The stated input capacitance is 0.1-pF.

5.2.2 Measured Results

Step Response The setup for testing the step response of the VGA is shown in Fig. 5.16. The voltage gain of the VGA is set to 1. For TC_1, its large step response

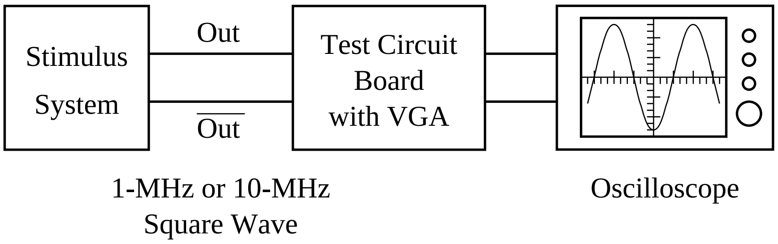


Figure 5.16 Setup for the step response testing.

to a 10-MHz 4-V_{pp} input signal is shown in Fig. 5.17. The single-ended outputs and the differential output are plotted on the upper graticule and the lower graticule, respectively. The output waveforms agree with the corresponding simulation results. If the input frequency is slowed down to 1-MHz, the measured results for the large (4-V_{pp}) and the small (1-V_{pp}) step response are shown in Fig. 5.18. The output

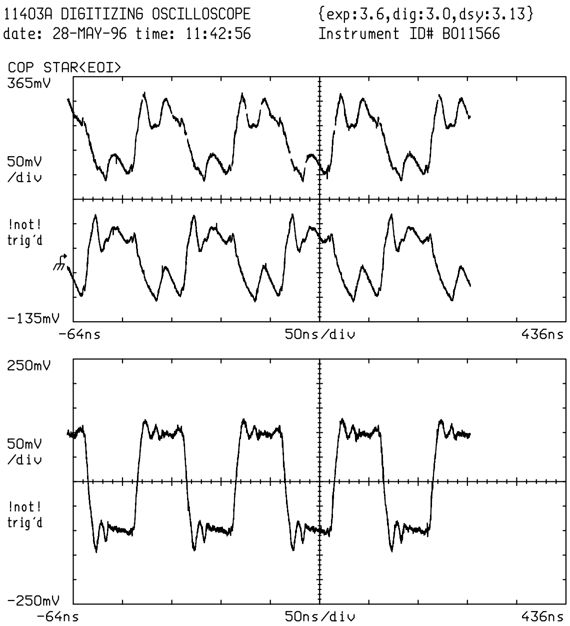


Figure 5.17 Measured large step response of TC_1 at 10-MHz.

waveforms show that the settling of the single-ended response is not fast enough. It is related to the stability or the compensation of the circuit.

In the test chip, TC_2 has bigger compensation capacitors. If the input frequency is

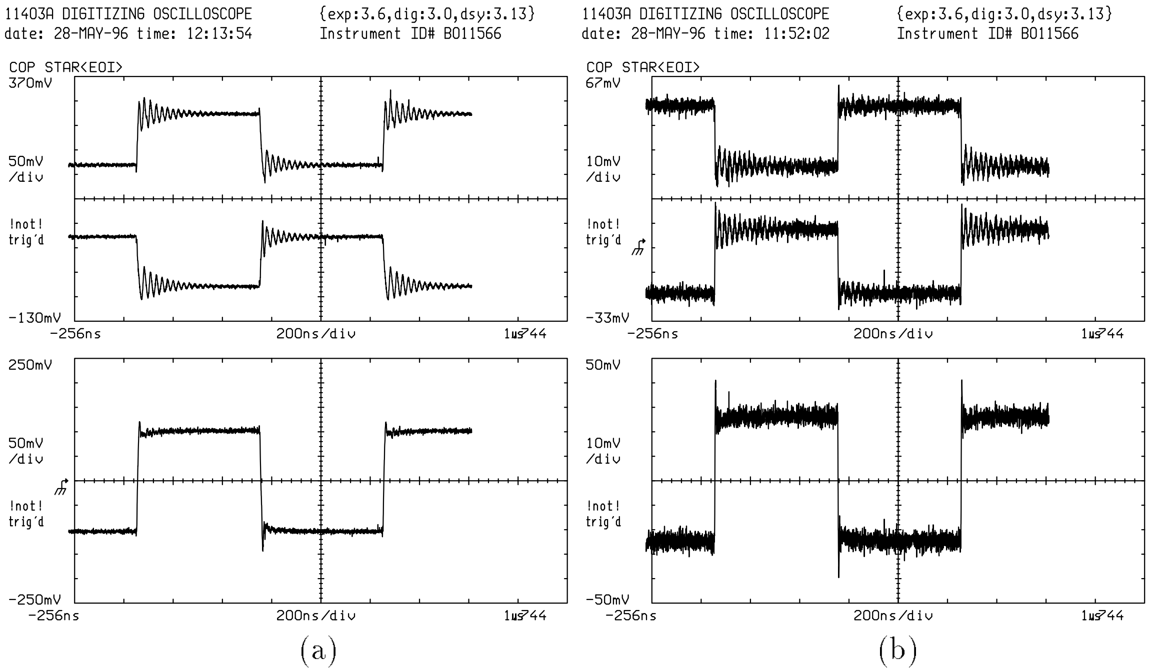


Figure 5.18 Measured step response of TC_1 at 1-MHz, (a) for large step response and (b) for small step response.

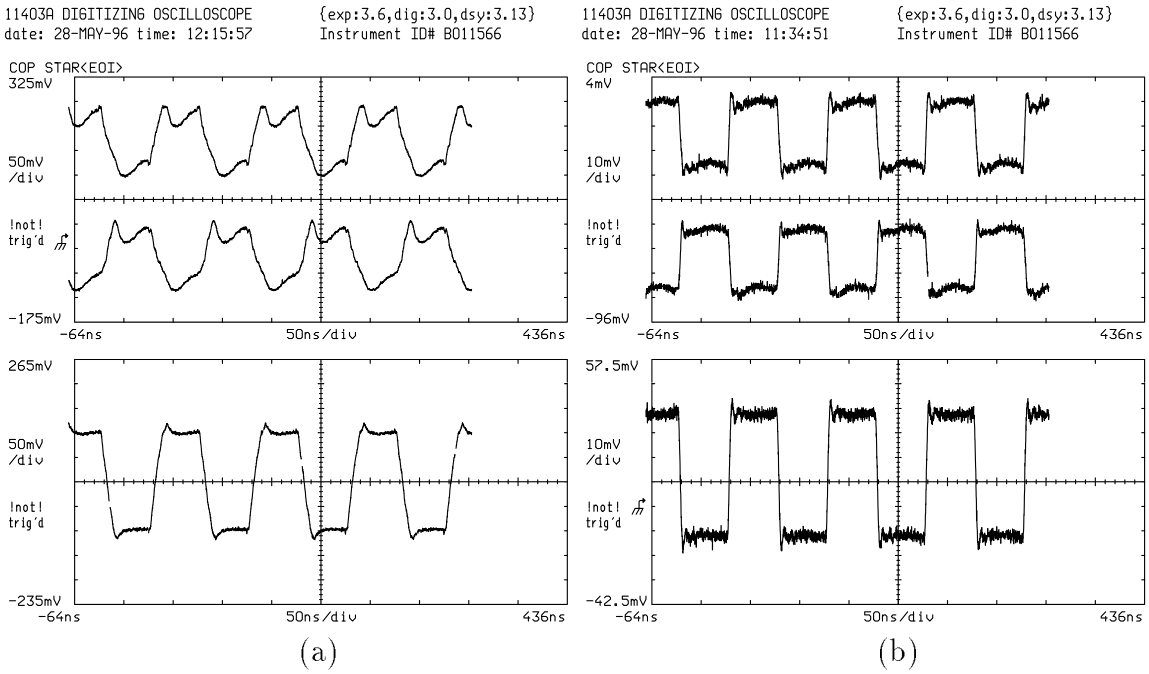


Figure 5.19 Measured step response of TC_2 at 10-MHz, (a) for large step response and (b) for small step response.

10-MHz, the measured results for the large and the small step response are shown in Fig. 5.19. The output waveforms in the large step response agree with the simulation results.

As in simulation, the overshooting in the differential output is reduced because of common-mode rejection. The phase margin for this fully-differential circuit is approximately 70°. However, the ringing in the single-ended outputs illustrates that there exist some problems in the CMFB circuit.

The slew rates obtained from simulation and from measurement are compared in Table 5.1. Their values are very close. If the size of the compensation capacitors are increased, the slew rate will be reduced.

Table 5.1 The slew rates for the differential output obtained from simulation and from measurement.

Compensation C_c	From Simulation		From Measurement	
	Slew Rate	FPBW	Slew Rate	FPBW
2.0-pF, 2.5-pF	445.0-V/μs	35.41-MHz	$\approx$ 475-V/μs	$\approx$ 37.7-MHz
3.0-pF, 4.0-pF	294.2-V/μs	23.41-MHz	$\approx$ 310-V/μs	$\approx$ 24.7-MHz

Frequency Response The setup for testing the frequency response of the VGA is shown in Fig. 5.20. The voltage gain of the VGA is set to 1. For TC_1, its frequency

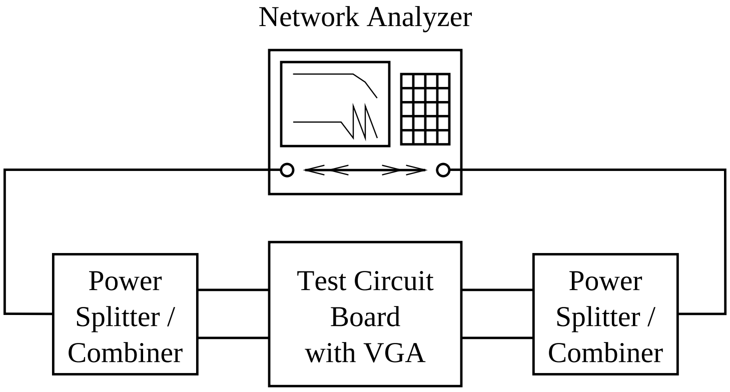
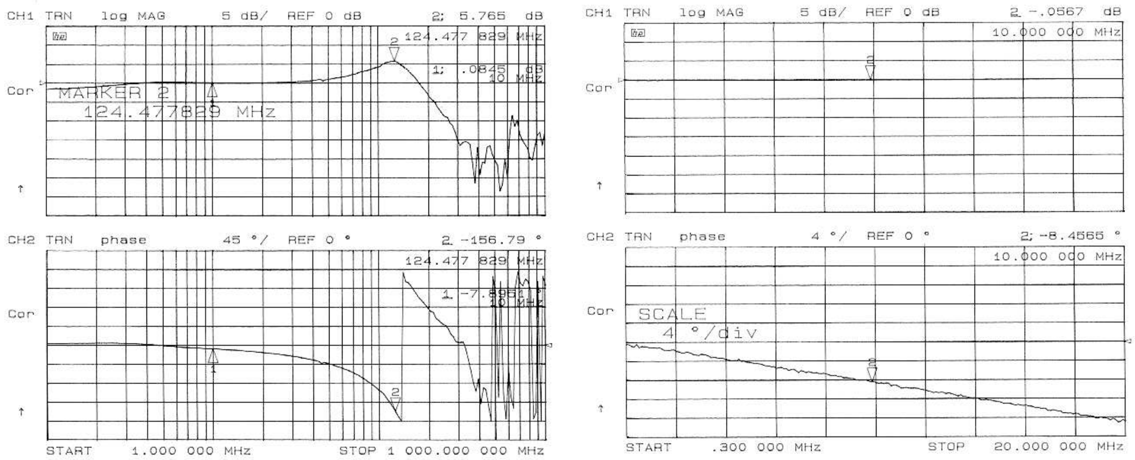


Figure 5.20 Setup for the frequency response testing.

response is shown in Fig. 5.21. Fig. 5.21a shows that the cut-off frequency is close to 200-MHz. From Fig. 5.21b, by approximation, the transfer function of the VGA has a linear phase response characteristic for the frequency from DC to 20-MHz. This means that the output signals have a constant time delay if the signal frequencies are below 20-MHz. This is a very important property for pulsed-data systems. If the VGA does not have this property, the sharp edges of the waveform will be distorted.



(a) (b)

Figure 5.21 The measured frequency response of TC_1, (a) from 1-MHz to 1-GHz and (b) from 300-kHz to 20-MHz.

Total Harmonic Distortion The setup for measuring the THD of the VGA is shown in Fig. 5.22. Table 5.2 and Table 5.3 list the THD measured from TC_1 when

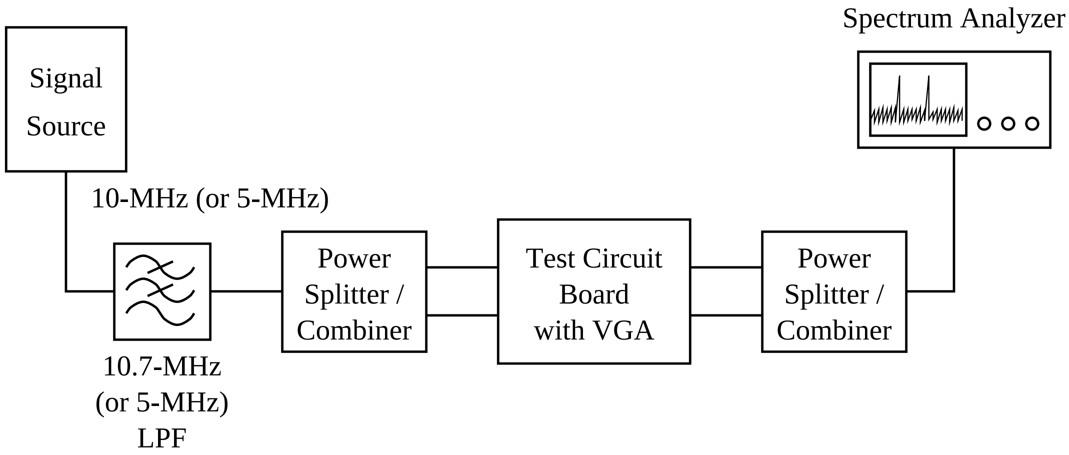


Figure 5.22 Setup for the THD measurement.

an input signal at 5-MHz and at 10-MHz are applied and the differential output is 4-V_{pp}. The voltage gain is approximately equal to 1 when V_{ctl} is equal to -1.25-V; and the voltage gain is approximately equal to 2 when V_{ctl} is equal to 1.00-V. The plots for comparing the simulation and the measured results are shown in Fig. 5.23 and Fig. 5.24.

From the results, if the voltage gain is set between 1.15 and 1.85 and the signal frequency is below 10-MHz, the VGA has 10-bit linearity (60-dB). The worst case is found when the signal frequency is 10-MHz and the voltage gain is unity. The linearity is only 8.5-bits. In general, the measured results are close to the simulation results. The differences may arise from the device modeling because the device models cannot exactly reproduce the transfer characteristics of each transistor.

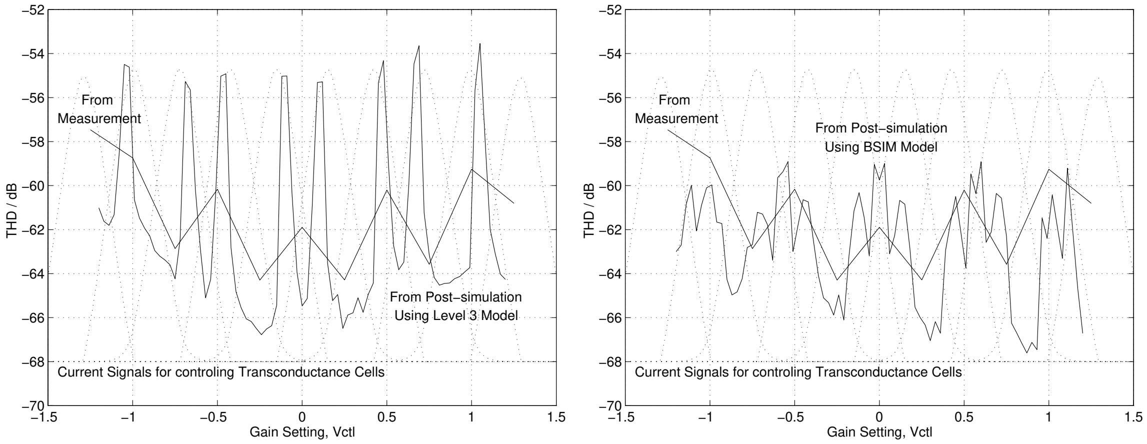
Table 5.2 The measured THD of TC_1 when an input signal at 5-MHz is applied.

V_{ctl} with respect to 1.5-V above V_{ss}	THD of V_{in} (NF = −70-dB; RL = 20-dB; ATT = 30-dB)	THD of V_{out+} and V_{out-} (NF = −70-dB, RL = −5-dB; ATT = 10-dB)	THD of V_{out} (NF = −78-dB, RL = −5-dB; ATT = 10-dB)
−1.25-V	4.00-V_{pp}, −61.46-dB (HD = −62.88, −67.01)	−38.77-dB (HD = −38.84, −59.21, −60.39)	−57.46-dB (HD = −63.38, −58.70)
−1.00-V	3.67-V_{pp}, −62.46-dB (HD = −63.71, −68.46)	−53.71-dB (HD = −55.38, −58.66, –)	−58.74-dB (HD = −68.63, −59.21)
−0.75-V	3.50-V_{pp}, −62.89-dB (HD = −64.05, −69.17)	−62.03-dB (HD = −64.55, −65.59, –)	−62.86-dB (HD = −67.80, −64.54)
−0.50-V	3.33-V_{pp}, −63.06-dB (HD = −64.13, −69.67)	−60.88-dB (HD = −68.50, −61.71, –)	−60.16-dB (HD = −67.17, −61.13)
−0.25-V	3.00-V_{pp}, −64.03-dB (HD = −65.04, −70.87)	−64.18-dB (HD = −68.45, −66.21, –)	−64.30-dB (HD = −68.12, −66.62)
0.00-V	2.83-V_{pp}, −64.69-dB (HD = −65.58, −72.00)	−61.93-dB (HD = −68.13, −63.12, –)	−61.90-dB (HD = −67.84, −63.17)
0.25-V	2.50-V_{pp}, −65.54-dB (HD = −66.38, −73.08)	−61.76-dB (HD = −63.17, −67.34, –)	−64.28-dB (HD = −67.38, −67.21)
0.50-V	2.33-V_{pp}, −66.28-dB (HD = −67.01, −74.38)	−59.82-dB (HD = −64.34, −61.71, –)	−60.21-dB (HD = −65.87, −61.58)
0.75-V	2.17-V_{pp}, −66.92-dB (HD = −67.63, −75.13)	−56.28-dB (HD = −56.71, −66.54, –)	−63.57-dB (HD = −66.63, −66.54)
1.00-V	2.00-V_{pp}, −67.46-dB (HD = −68.09 −76.17)	−58.66-dB (HD = −62.58, −60.92, –)	−59.26-dB (HD = −65.42, −60.46)
1.25-V	1.83-V_{pp}, −69.26-dB (HD = −70.04, −77.09)	−43.11-dB (HD = −43.17, −63.54, −65.67)	−60.80-dB (HD = −63.96, −63.67)

Remark :
1. NF — Noise Floor; RL — Reference Level and ATT — Attenuation.
2. The single-ended output signals (V_{out+} and V_{out-}) are 2-V_{pp}. In measurement, the circuit interfaces to external voltage buffers. The output signal amplitudes would be 1/20 of the original values.
3. (HD = HD_2, HD_3, $\cdots$). The unit for each HD is "dBc".

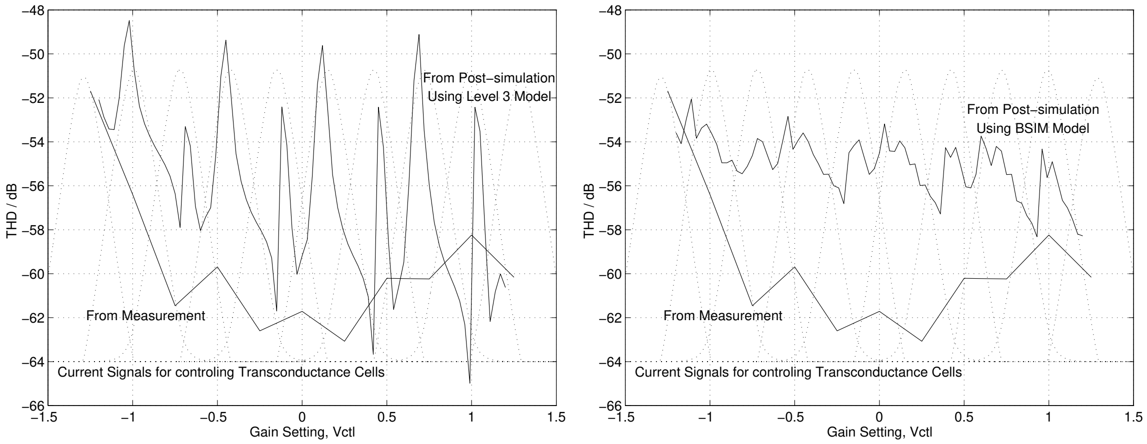
Table 5.3 The measured THD of TC_1 when an input signal at 10-MHz is applied.

V_{ctl} with respect to 1.5-V above V_{ss}	THD of V_{in} (NF = −70-dB; RL = 20-dB; ATT = 30-dB)	THD of V_{out+} and V_{out-} (NF = −70-dB, RL = −5-dB; ATT = 10-dB)	THD of V_{out} (NF = −78-dB, RL = −5-dB; ATT = 10-dB)
−1.25-V	4.00-V_{pp}, −63.48-dB (HD = −65.94, −67.11)	−31.97-dB (HD = −32.55, −50.92, −41.63 −64.63, −57.92)	−51.68-dB (HD = −60.67, −52.39 −72.56, −69.79)
−1.00-V	3.67-V_{pp}, −64.07-dB (HD = −65.89, −68.73)	−43.39-dB (HD = −44.25, −55.38, −52.75)	−56.40-dB (HD = −68.61, −56.67)
−0.75-V	3.50-V_{pp}, −64.68-dB (HD = −66.33, −69.67)	−49.94-dB (HD = −51.38, −61.29, −56.75)	−61.45-dB (HD = −68.17, −62.50)
−0.50-V	3.33-V_{pp}, −64.77-dB (HD = −66.39, −69.83)	−46.70-dB (HD = −47.54, −58.54, −56.25)	−59.69-dB (HD = −66.89, −60.61)
−0.25-V	3.00-V_{pp}, −65.75-dB (HD = −67.05, −71.61)	−48.30-dB (HD = −48.79, −63.46, −59.50)	−62.60-dB (HD = −67.22, −64.44)
0.00-V	2.83-V_{pp}, −66.27-dB (HD = −67.44, −72.53)	−48.51-dB (HD = −49.13, −62.04, −59.04)	−61.71-dB (HD = −67.17, −63.17)
0.25-V	2.50-V_{pp}, −67.25-dB (HD = −68.39, −73.61)	−47.70-dB (HD = −48.17, −64.04, −58.75)	−63.07-dB (HD = −67.17, −65.22)
0.50-V	2.33-V_{pp}, −67.72-dB (HD = −68.56, −75.27)	−49.20-dB (HD = −50.05, −60.17, −59.25)	−60.21-dB (HD = −66.44, −61.39)
0.75-V	2.17-V_{pp}, −68.66-dB (HD = −69.50, −76.22)	−47.32-dB (HD = −47.92, −60.75, −58.09)	−60.24-dB (HD = −66.78, −61.33)
1.00-V	2.00-V_{pp}, −69.71-dB (HD = −70.61, −77.00)	−46.90-dB (HD = −47.71, −57.88, −57.29)	−58.24-dB (HD = −64.78, −59.33)
1.25-V	1.83-V_{pp}, −69.15-dB (HD = −69.72, −78.28)	−39.16-dB (HD = −39.25, −67.61, −56.59)	−60.16-dB (HD = −60.55, −70.83)



(a) (b)

Figure 5.23 Comparison between the simulated and the measured THD of TC_1 (a) if the level 3 model is used and (b) if the BSIM model is used. The circuit is tested at room temperature by a 5-MHz signal.



(a) (b)

Figure 5.24 Comparison between the simulated and the measured THD of TC_1 (a) if the level 3 model is used and (b) if the BSIM model is used. The circuit is tested at room temperature by a 10-MHz signal.

The THD for the differential signal at 5-MHz and at 10-MHz are similar. However, the THD for the single-ended signals have a greater improvement if the signal frequency is slowed down. When the signal frequency is 10-MHz, each single-ended signal has around 8-bit linearity. If the frequency is slowed down to 5-MHz, the linearity will be improved to around 10-bit because of the reduction of the second harmonic.

Furthermore, when the measured results are compared with the simulation results, the measured THD has less variation (only 3-dB). Since the variation of THD is always due to the multi-input-stage (refer to Section 3.1.3), the simulation results may over-estimate its non-linearity.

For the purpose of comparison, Table 5.4 and Table 5.5 list the THD measured from TC_2 when an input signals at 5-MHz and at 10-MHz are applied. The THD for the differential signal is close to the results obtained from TC_1. It increases by 3-dB in average. This means that half a bit of linearity is lost, which is tolerable.

Table 5.4 The measured THD of TC_2 when an input signal at 5-MHz is applied.

V_{ctl} with respect to 1.5-V above V_{ss}	THD of V_{in} (NF = −70-dB; RL = 20-dB; ATT = 30-dB)	THD of V_{out+} and V_{out-} (NF = −70-dB, RL = −5-dB; ATT = 10-dB)	THD of V_{out} (NF = −78-dB, RL = −5-dB; ATT = 10-dB)
−1.00-V	3.67-V_{pp}, −62.46-dB (HD = −63.71, −68.46)	−48.04-dB (HD = −48.63, −57.96, −63.88)	−57.47-dB (HD = −69.34, −57.76)
−0.50-V	3.33-V_{pp}, −63.06-dB (HD = −64.13, −69.67)	−51.37-dB (HD = −52.07, −61.33, −64.54)	−60.21-dB (HD = −67.42, −61.13)
0.00-V	2.83-V_{pp}, −64.69-dB (HD = −65.58, −72.00)	−53.76-dB (HD = −54.87, −61.12, −67.62)	−60.21-dB (HD = −67.62, −61.08)
0.50-V	2.33-V_{pp}, −66.28-dB (HD = −67.01, −74.38)	−53.51-dB (HD = −54.96, −59.58, −67.79)	−58.85-dB (HD = −68.33, −59.37)
1.00-V	2.00-V_{pp}, −67.46-dB (HD = −68.09, −76.17)	−50.19-dB (HD = −51.13, −57.75, −67.34)	−58.29-dB (HD = −70.04, −58.59)

However, increasing the size of compensation capacitors have a great influence on the THD for the single-ended signals. When the frequencies are 5-MHz and 10-MHz, the linearity is only 8.5-bit and 5-bit, respectively. It is because the operating frequency is close to the FPBW ($\approx$ 24.7-MHz), and the magnitude of the positive sand the negative slew rate are not equal (see Fig. 5.6 and Fig. 5.19).

At high frequency, if the magnitude of the positive and the negative slew rate are different, the waveform at each half cycle will not be symmetrical. The resulting waveform appears to be a fundamental plus a second harmonic. This means that the second harmonic distortion will arise.

Table 5.5 The measured THD of TC_2 when an input signal at 10-MHz is applied.

V_{ctl} with respect to 1.5-V above V_{ss}	THD of V_{in} (NF = −70-dB; RL = 20-dB; ATT = 30-dB)	THD of $V_{\mathrm{out+}}$ and $V_{\mathrm{out-}}$ (NF = −70-dB, RL = −5-dB; ATT = 10-dB)	THD of V_{out} (NF = −78-dB, RL = −5-dB; ATT = 10-dB)
−1.00-V	3.67-$\mathrm{V_{pp}}$, −64.07-dB (HD = −65.89, −68.73)	−29.13-dB (HD = −29.17, −51.34, −54.17)	−52.37-dB (HD = −66.54, −52.54)
−0.50-V	3.33-$\mathrm{V_{pp}}$, −64.77-dB (HD = −66.39, −69.83)	−30.51-dB (HD = −30.54, −54.67, −54.76)	−55.56-dB (HD = −66.58, −55.92)
0.00-V	2.83-$\mathrm{V_{pp}}$, −66.27-dB (HD = −67.44, −72.53)	−31.44-dB (HD = −31.46, −57.58, −57.58)	−58.40-dB (HD = −65.71, −59.29)
0.50-V	2.33-$\mathrm{V_{pp}}$, −67.72-dB (HD = −68.56, −75.27)	−31.90-dB (HD = −31.92, −56.71, −59.54)	−58.31-dB (HD = −66.21, −59.08)
1.00-V	2.00-$\mathrm{V_{pp}}$, −69.71-dB (HD = −70.61, −77.00)	−31.61-dB (HD = −31.63, −55.88, −61.96)	−56.94-dB (HD = −68.50, −57.25)

Two-Tone Intermodulation The setup for measuring the two-tone intermodulation in the VGA is shown in Fig. 5.25. Different lowpass filters are employed to attenuate the unwanted harmonics. The measurement is performed when the control

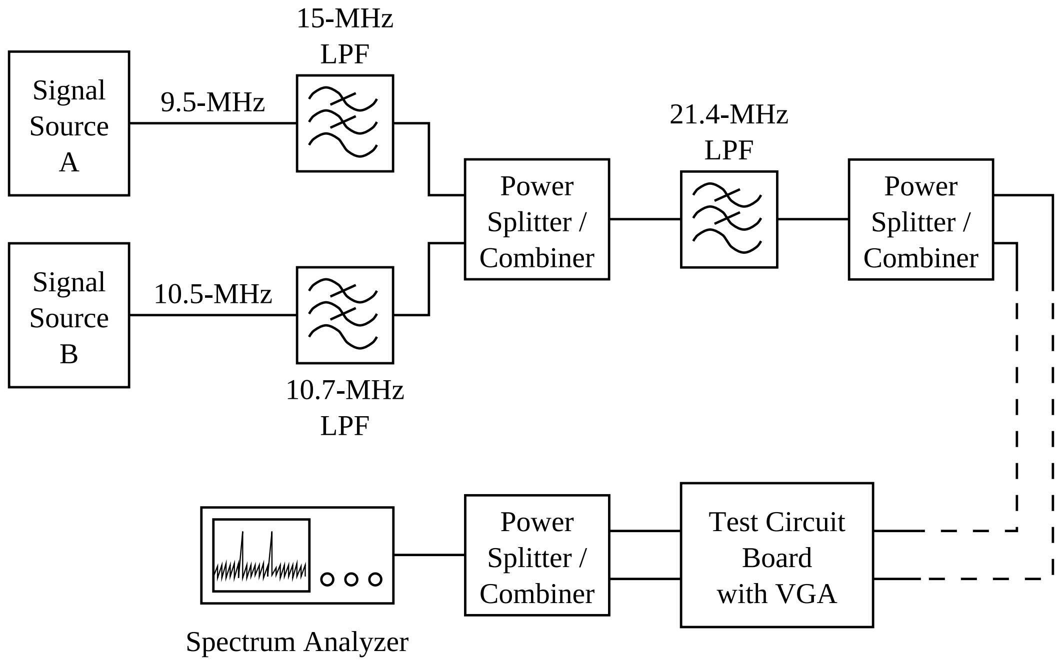
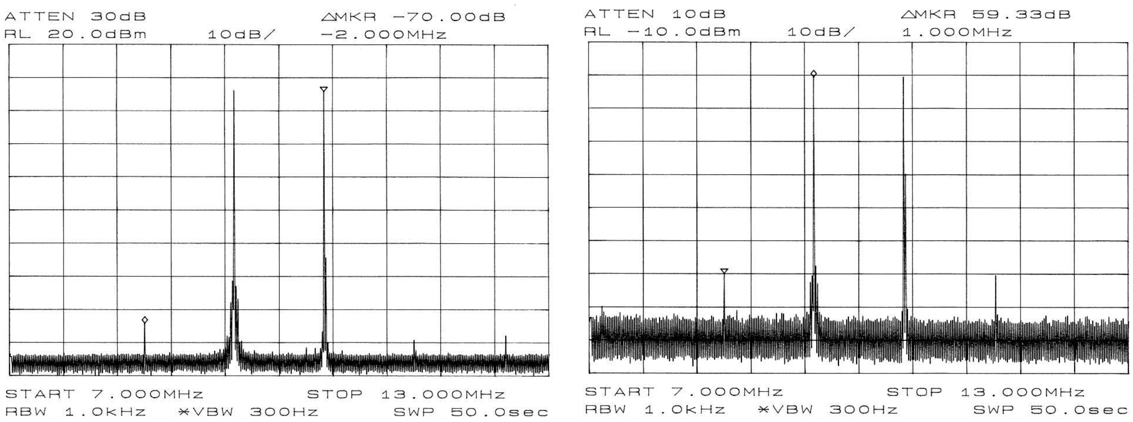


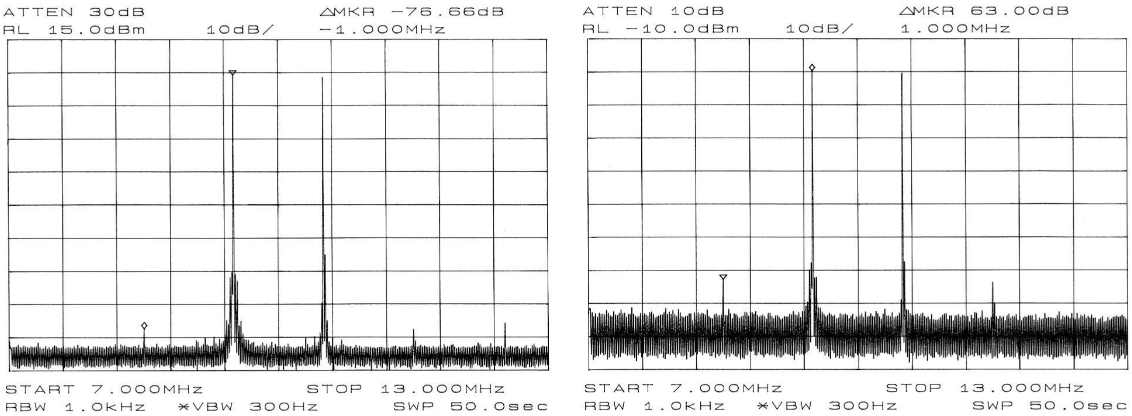
Figure 5.25 Setup for two-tone intermodulation measurement.

signal V_{ctl} is set to −1.00-V ($A_v \approx 1$), 0.00-V ($A_v \approx \sqrt{2}$) and 1.00-V ($A_v \approx 2$). Two equal amplitude tones at 9.5-MHz and 10.5-MHz are applied to TC_1 such that the amplitude of the differential output signal is 4-$\mathrm{V_{pp}}$. The spectrums for the input and the output signal are shown in Fig. 5.26, Fig. 5.27 and Fig. 5.28. The third-order



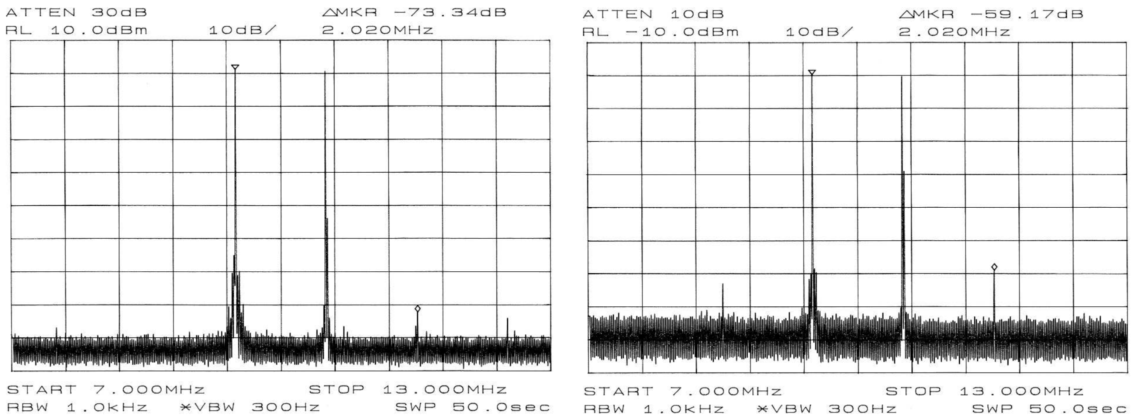
(a) (b)

Figure 5.26 Intermodulation distortion of TC_1 when $V_{ctl} = -1.00$-V, (a) input spectrum and (b) output spectrum.



(a) (b)

Figure 5.27 Intermodulation distortion of TC_1 when $V_{ctl} = 0.00$-V, (a) input spectrum and (b) output spectrum.



(a) (b)

Figure 5.28 Intermodulation distortion of TC_1 when $V_{ctl} = 1.00$-V, (a) input spectrum and (b) output spectrum.

intermodulation products (IM_3) are below −59.33-dBc for $V_{ctl} = -1.00$-V, below −63.00-dBc for $V_{ctl} = -0.00$-V and below −59.17-dBc for $V_{ctl} = 1.00$-V. These values are close to the corresponding third harmonic distortion (HD_3) obtained in Table 5.3. Generally, HD_3 should be 2.5-dB greater than IM_3.[3]

5.3 Conclusions

In this thesis, a novel idea in designing a CMOS VGA is presented. The VGA is constructed from a fully-differential multi-input-stage opamp with resistive feedback. The smoothly variation of the voltage gain is achieved by interpolation between multiple taps of a feedback resistor. Various aspects related to the design, analysis and optimization of this VGA have been presented in previous chapters. In this chapter, the simulation and the measured results are used to verify the performance of the VGA. In closing, a summary of the circuit performance is made and a few directions for improvement to the work presented are suggested.

5.3.1 Summary

The performance of the VGA (TC_1) obtained from measurements is summarized in Table 5.6. Measurements show a close correspondence to the simulation results. The VGA shows IMD of nearly −60-dB over all gain settings, which demonstrate that the VGA obtains a linearity of 10-bit for video bandwidths. Simulated noise performance shows an SNR greater than 73.79-dB. All the results show that the

Table 5.6 Summary for the performance of the VGA from measurement.

Parameter	Measured Results
Closed-loop Gain	1–2 or 0–6-dB
Output Voltage Swing	4-V_{pp} differential
Phase Margin	$\approx 70^o$
Slew Rate (FPBW)	$\approx$ 475-V/μs ($\approx$ 37.7-MHz)
THD at 5-MHz	< -60-dB when $A_v \in [1.15, 1.85]$
THD at 10-MHz	< -60-dB when $A_v \in [1.15, 1.85]$
IM_3, 4-V_{pp} at 9.5-MHz and 10.5-MHz	< -59.17-dBc
Supply Voltage	$\pm$2.5-V
DC Power Dissipation	80-mW
Transistor Count	167
Active Die Area	0.6×1-mm^2

[3] When the total amplitude of two-tone waveform equals the amplitude of a single-tone waveform, $IM_3 = (3/4)HD_3$.

VGA can achieve 60-dB dynamic-range in the middle part of the variable gain range [1.15,1.85]. Combined with further design improvements, it can be used as a variable gain buffer in 10-bit resolution video systems and digital broadband communication signal-conditioning front-end circuitry.

5.3.2 Further Improvements

According to the observed weaknesses in the present design and the aspects that were not stressed in the prototype, the improvements are itemized as follows.

Device Models The THD from simulation has a greater variation when it is compared with the measured results. This means that the non-linearity due to the multi-input-stage may be over-estimated at the beginning.

THD is a measure of the circuit non-linearity. For the multi-input-stage, it is very important to have a good model of the transistor's transfer characteristics when it is switching between the different regions. If an accurate model is available, the prediction of THD at the outputs and the optimization of the circuit performance will be performed more properly.

CMFB Circuit The results show that the CMFB circuit has some problems. The existing CMFB circuit is a general design which assumes the input signals are taken at the output terminals; and the selections of the transistors' aspect ratio are non-optimal. Moreover, the input signals to the CMFB circuit are obtained from the mid-point of two feedback resistors. Their swing may be too small or the sensitivity of this circuit is not fast enough to track the immediate common-mode error. If the CMFB circuit is redesigned according to the above problems, the stability of the VGA would be improved.

On the other hand, as regards the simulation result in the upper part of Fig. 5.3, the common-mode output potential is moving up and down by 50-mV when $V_{\rm ctl}$ is changing. It would be treated as a DC noise from the control signal $V_{\rm ctl}$. This is because the total bias current of the multi-input-stage is not a constant which does not match the design of the telescopic-cascode stage and the CMFB circuit. In solving this problem, the sum of all bias currents in the multi-input-stage is duplicated and fed as the reference current for the CMFB circuit.

Power and Die Size In order to maintain a large operating bandwidth and a low noise characteristic, the bias current and the aspect ratio of transistors in all input stages and in the telescopic-cascode stage are chosen to be fairly large.

If the bias currents and the aspect ratios are reduced, the VGA will consume less power and require a smaller die area. At the same time, if longer feedback resistors are used, less output current will be drawn from each output stage. All output stages will have a smaller current modulation index and contribute less distortion. As a result, the improved performance in all output stages can lessen the performance degradation due to the adjustment of the bias currents and the aspect ratios.

Layout The layout of the capacitors is not perfect because the polysilicon layer is used. A large grounding capacitance is found at each output terminal. In future test chips, the capacitors should be formed by only metal layers, and then the grounding capacitance will be minimized, even if a large capacitor area is required.

The digital noise is harmful to all analog circuits. If this VGA is integrated with any digital circuits (such as ADC) on the same substrate, it must have guard rings for collecting the injected minority carriers. Also, V_{dd} and V_{ss} of the guard rings should not be connected to any analog circuit.

Variable Voltage Gain In the design stage, the variable gain range is assigned to be from 0.95 to 2.10 such that it can achieve 10-bit linearity for the gain from 1 to 2. In circuit testing and measurement, its performance demonstrates that 10-bit linearity is achieved only when the gain is set between 1.15 and 1.85. This means that the safety margins are not large enough.

If the VGA should meet the expected specification (10-bit linearity for the gain from 1 to 2), the safety margin of the variable gain range must be enlarged. At the same time, the first and the last G_{m} cell are not in use because their contribution to distortion cannot be canceled by the other G_{m} cells.

Appendix A

Harmonic Distortion and Predistortion

This appendix is the summary of [11, Chapter 4]. It is talking about the relationship between the harmonic distortion and the harmonic predistortion. The equations for converting their coefficients are given at the end.

For a SISO nonlinear system, if its transfer function is a bijective mapping,[1] it can be written as

$$\Delta y = f(\Delta x) \qquad \text{and} \qquad \Delta x = g(\Delta y) = f^{-1}(\Delta y), \tag{A.1}$$

where Δy and Δx are output and input, respectively.

When the transfer functions are rewritten in power series,

$$\Delta y = a_1\Delta x + a_2\Delta x^2 + a_3\Delta x^3 + a_4\Delta x^4 + a_5\Delta x^5 + \cdots, \tag{A.2}$$
$$\Delta x = b_1\Delta y + b_2\Delta y^2 + b_3\Delta y^3 + b_4\Delta y^4 + b_5\Delta y^5 + \cdots. \tag{A.3}$$

For a linear system, $\Delta y = a_1\Delta x$ or $\Delta x = b_1\Delta y = \Delta y/a_1$. The others terms are unwanted. (a_2, a_3, a_4, a_5, ...) and (b_2, b_3, b_4, b_5, ...) are classified as the *distortion* and the *predistortion* coefficients, respectively.

For simplicity, the harmonic distortion is defined in terms of the unwanted output harmonics produced when a circuit is driven by a pure sinusoidal input signal. The harmonic predistortion measures the proportion of all harmonics must be added to predistort the input signal $x = x_0 + \rho_x \cos\omega t$ to produce a pure sinusoidal output signal with a DC offset.

In some cases, the transfer function of a system can be expressed as (A.3) instead of (A.2). This is a general problem of reverting a power series mathematically.

[1] A bijective mapped transfer function is restricted to its mathematical operation. It does not means that the input and the output of the system are reversible.

We substitute (A.3) into (A.2),

$$\begin{aligned}
\Delta y &= a_1(b_1\Delta y + b_2\Delta y^2 + b_3\Delta y^3 + b_4\Delta y^4 + b_5\Delta y^5 + \cdots) \\
&\quad + a_2(b_1\Delta y + b_2\Delta y^2 + b_3\Delta y^3 + b_4\Delta y^4 + b_5\Delta y^5 + \cdots)^2 \\
&\quad + a_3(b_1\Delta y + b_2\Delta y^2 + b_3\Delta y^3 + b_4\Delta y^4 + b_5\Delta y^5 + \cdots)^3 \\
&\quad + a_4(b_1\Delta y + b_2\Delta y^2 + b_3\Delta y^3 + b_4\Delta y^4 + b_5\Delta y^5 + \cdots)^4 \\
&\quad + a_5(b_1\Delta y + b_2\Delta y^2 + b_3\Delta y^3 + b_4\Delta y^4 + b_5\Delta y^5 + \cdots)^5 \\
&\quad + \cdots, \\
&= a_1(b_1\Delta y + b_2\Delta y^2 + b_3\Delta y^3 + b_4\Delta y^4 + b_5\Delta y^5 + \cdots) \\
&\quad + a_2[b_1^2\Delta y^2 + 2b_1b_2\Delta y^3 + (2b_1b_3 + b_2^2)\Delta y^4 + (2b_1b_4 + 2b_2b_3)\Delta y^5 + \cdots] \\
&\quad + a_3[b_1^3\Delta y^3 + 3b_1^2b_2\Delta y^4 + (3b_1^2b_3 + 3b_1b_2^2)\Delta y^5 + \cdots] \\
&\quad + a_4[b_1^4\Delta y^4 + 4b_1^3b_2\Delta y^5 + \cdots] \\
&\quad + a_5[b_1^5\Delta y^5 + \cdots] + \cdots, \\
&= \Delta y(a_1b_1) \\
&\quad + \Delta y^2(a_1b_2 + a_2b_1^2) \\
&\quad + \Delta y^3(a_1b_3 + 2a_2b_1b_2 + a_3b_1^3) \\
&\quad + \Delta y^4(a_1b_4 + 2a_2b_1b_3 + a_2b_2^2 + 3a_3b_1^2b_2 + a_4b_1^4) \\
&\quad + \Delta y^5(a_1b_5 + 2a_2b_1b_4 + 2a_2b_2b_3 + 3a_3b_1^2b_3 + 3a_3b_1b_3^2 + 4a_4b_1^3b_2 + a_5b_1^5) \\
&\quad + \cdots. \qquad \text{(A.4)}
\end{aligned}$$

If we compare the coefficients on L.H.S. and R.H.S.,

$$\begin{aligned}
1 &= a_1b_1, \\
0 &= a_1b_2 + a_2b_1^2, \\
0 &= a_1b_3 + 2a_2b_1b_2 + a_3b_1^3, \qquad \text{(A.5)} \\
0 &= a_1b_4 + 2a_2b_1b_3 + a_2b_2^2 + 3a_3b_1^2b_2 + a_4b_1^4, \\
0 &= a_1b_5 + 2a_2b_1b_4 + 2a_2b_2b_3 + 3a_3b_1^2b_3 + 3a_3b_1b_3^2 + 4a_4b_1^3b_2 + a_5b_1^5.
\end{aligned}$$

After solving the system equations of (A.5),

$$\begin{aligned}
a_1 &= \frac{1}{b_1}, \\
a_2 &= -\frac{b_2}{b_1^3}, \\
a_3 &= \frac{2b_2^2 - b_1b_3}{b_1^5}, \qquad \text{(A.6)} \\
a_4 &= \frac{-b_1^2b_4 + 5b_1b_2b_3 - 5b_2^3}{b_1^7}, \\
a_5 &= \frac{-b_1^3b_5 + 6b_1^2b_2b_4 + 3b_1^2b_3^2 - 21b_1b_2^2b_3 + 14b_2^4}{b_1^9}.
\end{aligned}$$

We can verify (A.6) by

$$1 + y = \sqrt{1+x},$$

where

$$\begin{aligned}
\sqrt{1+x} &= 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \frac{7}{256}x^5 + \cdots, \\
(1+y)^2 &= 1 + 2y + y^2.
\end{aligned}$$

If the system is a fully differential design, all the coefficients of the even power term (b_2, b_4, ...) would be attenuated or they are equal to zero. (A.6) are modified to be

$$\begin{aligned}
a_1 &= \frac{1}{b_1}, \\
a_3 &= -\frac{b_3}{b_1^4}, \\
a_5 &= \frac{-b_1 b_5 + 3b_3^2}{b_1^7}, \\
a_{2i} &= 0 \quad \text{where } i \text{ is a positive integer.}
\end{aligned} \tag{A.7}$$

The above equations demonstrate that the conversion between the coefficients in (A.2) and (A.3) is possible under the bijective mapping condition.

Bibliography

[1] S.-I. Liu and Y.-S. Hwang, "CMOS four-quadrant multiplier using bias feedback techniques," *IEEE J. Solid-State Circuits*, vol. 29, pp. 750–752, June 1994.

[2] N. Saxena and J. J. Clark, "A four-quadrant CMOS analog multiplier for analog neural network," *IEEE J. Solid-State Circuits*, vol. 29, pp. 746–749, June 1994.

[3] H.-J. Song and C.-K. Kim, "An MOS four-quandrant analog multiplier using simple two-input squaring circuits with source followers," *IEEE J. Solid-State Circuits*, vol. 25, pp. 841–848, June 1990.

[4] Z. Wang, "A CMOS four-quadrant analog multiplier with single-ended voltage output and improved temperature performance," *IEEE J. Solid-State Circuits*, vol. 26, pp. 1293–1301, Sept. 1991.

[5] M. Wakayama, "Variable gain amplifier by interpolating between dual-input stages of an opamp, *private communication*," Jan. 1995.

[6] A. B. Grebene, *Bipolar and MOS analog integrated circuit design*. John Wiley & Sons, Inc., first ed., 1984.

[7] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. John Wiley & Sons, Inc., third ed., 1993.

[8] J. F. Duque-Carrillo, J. M. Valverde, and R. Pérez-Aloe, "Constant-Gm rail-to-rail common-mode range input stage with minimum CMRR degradation," *IEEE J. Solid-State Circuits*, vol. 28, pp. 661–666, June 1993.

[9] R. Hogervorst, J. P. Tero, R. G. H. Eschauzier, and J. H. Huijsing, "A compact power-efficient 3-V CMOS rail-to-rail input/output operational amplifier for VLSI cell libraries," *IEEE J. Solid-State Circuits*, vol. 29, pp. 1505–1513, Dec. 1994.

[10] K. Bult and A. Buchwald, "A 250-mW 10-bit 50-MS/s CMOS AD-converter in 1-mm^2, *unpublished*," Sept. 1995.

[11] S. D. Willingham and K. Martin, *Integrated Video-Frequency Continuous-Time Filters: High-Performance Realizations in BiCMOS*. Kluwer Academic Publishers, first ed., 1995.

[12] A. Buchwald and K. W. Martin, *Integrated Fiber-Optic Recivers*. Kluwer Academic Publishers, first ed., 1995.

[13] S. Ross, *A First Course in Probability.* Macmillan Publishing Company, third ed., 1988.

[14] S. Rosenstark, *Feedback Amplifier Principles.* Macmillan Publishing Company, first ed., 1986.

[15] M. Ismail and T. Fiez, *Analog VLSI: Signal and Information Processing.* McGraw-Hill, Inc., international ed., 1994.

[16] J. E. Franca and Y. Tsividis, *Design of Analog-Digital VLSI Circuits for Telecommunications and Signal Processing.* Prentice Hall, second ed., 1994.

[17] S. R. Vemuru, "Layout comparsion of MOSFETs with large W/L ratios," *Electronic. Lett.*, vol. 28, pp. 2327–2329, Dec. 1992.

[18] N. H. E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design: A Systems Perspective.* Addison-Wesley Publishing Company, second ed., 1993.

[19] R. S. Muller and T. I. Karmins, *Device Electronics for Integrated Circuits.* John Wiley & Sons, Inc., second ed., 1986.

[20] S. Satyanarayana and K. Suyama, "Resistive interpolation biasing: A technique for compensating linear variation in an array of MOS current sources," *IEEE J. Solid-State Circuits*, vol. 30, pp. 595–598, May 1995.

[21] C. E. Rohrs, J. L. Melsa, and D. G. Schultz, *Linear Control Systems.* McGraw-Hill, Inc., international ed., 1993.

[22] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems.* Addison-Wesley Publishing Company, second ed., 1991.