AN ABSTRACT OF THE THESIS OF

Chaiyanut Aueamnuay for the degree of Master of Science in Electrical and Computer Engineering presented on August 11, 2021.

Title: g_m/I_D Design Techniques for Low-power High-performance CMOS Analog Integrated Circuits

Abstract approved:

_____

David J. Allstot

The g_m/I_D-based design of analog integrated circuits introduced in 1996 employs an empirical transistor sizing methodology using *SPICE*-generated lookup tables that enables good agreement between simulations and specifications. This research introduces a *SPICE* lookup table that extends the g_m/I_D design approach to process, voltage, and temperature-insensitive Miller pole-splitting frequency compensation of the classical CMOS two-stage operational transconductance amplifier (OTA).

In the design of ultra-low-power amplifiers for optimum performance, the iconic plots of g_m/I_D vs. the excess voltage, V_{OV}, suggest that some devices should be operated deep in weak inversion where g_m/I_D is near maximum. Performance parameters such as gain, bandwidth, thermal noise, and power dissipation benefit from this choice. In applications where small-signal settling time is critical, the unity-gain phase margin (PM) is a parameter of paramount importance; PM vs. V_{OV} design considerations are also presented in this thesis. A key result is that as the design choice of V_{OV} moves the region of operation from strong to moderate to weak inversion, the PM is reduced substantially and the settling time is increased dramatically. In addition to new design and synthesis insights, area-efficient minimum parasitic capacitance device layout techniques are illustrated that improve performance.

g_m/I_D Design Techniques for Low-power High-performance CMOS Analog Integrated Circuits

by

Chaiyanut Aueamnuay

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented August 11, 2021
Commencement June 2022

Master of Science thesis of Chaiyanut Aueamnuay presented on August 11, 2021

APPROVED:

Major Professor, representing Electrical and Computer Engineering

Head of the School of the Electrical Engineering and Computer Science

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of the Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Chaiyanut Aueamnuay, Author

ACKNOWLEDGEMENTS

I would like to express my special thanks to my advisor Professor David Allstot who gave me an opportunity to do a wonderful MS research project. It was very challenging for me since I started my academics at OSU as an INTO student. I decided to change my program from MEng to MS after one year of my master's program. During my research work, Professor Allstot always patiently guided me in terms of knowledge and passion to become a good IC designer. He always encouraged and supported me every time when I was struggling on the problem. Professor never denied any idea that I proposed to apply on the research. On the contrary, he is willing to spend time to discuss the idea with me. Because of this I respect him as my role model, and I wish to be able to follow his footsteps in the future.

I would not have been able to start and complete my master's degree without my parents. They are one of the main keys behind my success. When I decided to pursue a master's degree, I told my parents of this idea. They agreed with my decision and have not hesitated to financially support me. I know this has been a big task for them to take to respond to me, especially when I decided to change to do a research degree which meant I needed to extend another year to complete it. They have also provided constant encouragement. They always cheer me up and stay by my side when I face any difficulties in my life. I would like to dedicate this section of my thesis to my parents and my family.

Next, I would like to thank Professors Un-Ku Moon, Arun Natarajan and Gabor Temes who have taught fundamental analog IC design knowledge to me. Their teaching and lectures have helped me to emphasize simplifying a complex problem which is a useful tool to gain insight understanding and to be able to analyze the circuit operation intuitively. I would also want to thank my committee members Professors Matt Johnson, Gabor Temes and Megumi Kawasaki.

Lastly, I would like to thank Bhagirath, Ajmal, Julie, Bo and Jialin who are the fantastic research teammates. All of them were willing to help me clarify any doubts and support me for the thesis project. Also, the people in the office are very nice,

friendly and easy to talk with, especially Professor Teme’s team (Jenny, Emanuel, Alexander and Manju). Thanks especially for their friendship.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

Chapter 1: Introduction

The basic g_m/I_D-based design methodology for analog integrated circuits introduced by Silveira, et al. in 1996 [1] is advanced in a 2017 book by Jespers and Murmann [2]. It is emerging as an essential design technique as MOSFETs scale to 7nm and beyond and the corresponding device models become ever more complex. For example, the shrinking of the channel length introduces many short-channel effects such as mobility degradation, drain-induced barrier lowering, etc. As a result, the expected small-signal transconductance of the MOSFET derived from the large-signal square-law model deviates significantly from simulation results (20-60%) [2]. As a consequence, the classical analysis and design techniques based on the simple square-law approximations are no longer adequate.

The g_m/I_D design flow is an empirical transistor sizing methodology that employs *Simulation Program with Integrated Circuits Emphasis* (*SPICE*)-generated lookup tables and plots [3]. It allows the designer to explore the space of performance specifications such as small-signal voltage gain, transition frequency (f_T), small-signal transconductance (g_m), power consumption, etc., in all of the possible saturation or non-saturation (i.e., triode) operating regions (weak (i.e., subthreshold), moderate, and strong inversion). This enables designers to quickly determine and compare first-pass design choices and trade-offs when selecting transistor sizes. Moreover, the g_m/I_D-based methodology provides good agreement with *SPICE* simulations.

The growing interest in g_m/I_D design techniques is demonstrated by many recently published papers. For example, the systematic design and optimization of a CMOS single-stage operational transconductance amplifier (OTA) using the g_m/I_D methodology was described by Sabry, et al. in 2018 [4], and related design automation synthesis techniques were presented by Kumar, et al. in 2019 [5]. Of course, as CMOS technology continues to scale to deep sub-micron minimum feature sizes (e.g., 7nm), the total power supply voltage also scales down dramatically. Owing to minimum DC bias voltage headroom and output voltage swing requirements, therefore, the use of single-stage cascode OTA topologies is ever more restricted. These limitations can be overcome using multiple gain-stage topologies such as the classical CMOS two-stage OTA presented by Black, et al. [6]. Although the design of these low-voltage high-gain

topologies is arguably more difficult because frequency compensation is required to guarantee closed-loop stability, the g_m/I_D methodology offers an easier and more systematic exploration of the available design space which speeds synthesis.

Following this introductory chapter, Chapter 2 reviews both the g_m/I_D prior art and the CMOS two-stage OTA frequency compensation techniques. The g_m/I_D design flow for a classical two-stage opamp with process, voltage and temperature (PVT)-insensitive tracking Miller pole-splitting frequency compensation is detailed in Chapter 3 [6]. A key contribution of this work is the development of new *SPICE*-generated lookup tables for sizing MOSFETs operated in the triode region as needed with this compensation technique. Chapter 4 explores the advantages and limitations of subthreshold OTA designs and proposes device layout techniques that minimize critical parasitic capacitances. Finally, Chapter 5 concludes this thesis.

Chapter 2: g_m/I_D Methodology and CMOS Two-stage OTA Frequency Compensation

2.1 g_m/I_D Design Methodology Prior Art

The g_m/I_D methodology disclosed by Silveira, et al. [1] enables the design of CMOS analog integrated circuits such as operational amplifiers wherein the constituent MOSFETs can be operated in all possible regions from weak to strong inversion. The method starts by considering the relationship between g_m/I_D and the normalized drain current of the MOSFET, $I_D/(W/L)$, which can be expressed mathematically as

$$\frac{g_m}{I_D} = \frac{\partial \left\{ ln\left[\frac{I_D}{\left(\frac{W}{L}\right)} \right] \right\}}{\partial V_G} \tag{2.1}$$

This normalization effectively means that g_m/I_D is independent of the size (i.e., aspect ratio) of the transistor and therefore specifies a unique characteristic of the NMOS and PMOS devices for each scaled technology node. In addition to being size independent, the g_m/I_D ratio is also a measure of the efficiency of converting DC drain current to small-signal transconductance. As illustrated in Fig. 2.1, the highest values of g_m/I_D are obtained with the transistors operating in weak inversion (i.e., the subthreshold region) where I_D is an exponential function of V_{GS}. Conversely, g_m/I_D decreases dramatically as the operating region moves from weak to moderate or strong inversion where I_D is an approximately quadratic function of V_{GS}. Hence, biasing transistors in the subthreshold region reduces power consumption because greater g_m is the obtained for the same DC bias current. Of course, this benefit trades off against larger transistor sizes with greater parasitic capacitances. This tradeoff and layout techniques to mitigate it are detailed in Chapter 4 which explores the practical limitations of subthreshold two-stage OTA designs.

The g_m/I_D versus $I_D/(W/L)$ graphs are useful design aids for determining the initial sizing of the transistors for the various operating regions. A detailed design example for the simple single-stage current-steering OTA of Fig. 2.2 is presented in [1] for typical unity-gain frequency, load capacitor (C_L), slew rate (SR), DC power dissipation, etc., specifications. The g_m/I_D ratio of the input pair is chosen near the boundary between moderate and weak inversion which is near optimum for the low-frequency voltage gain, unity-gain frequency, and unity-gain phase margin

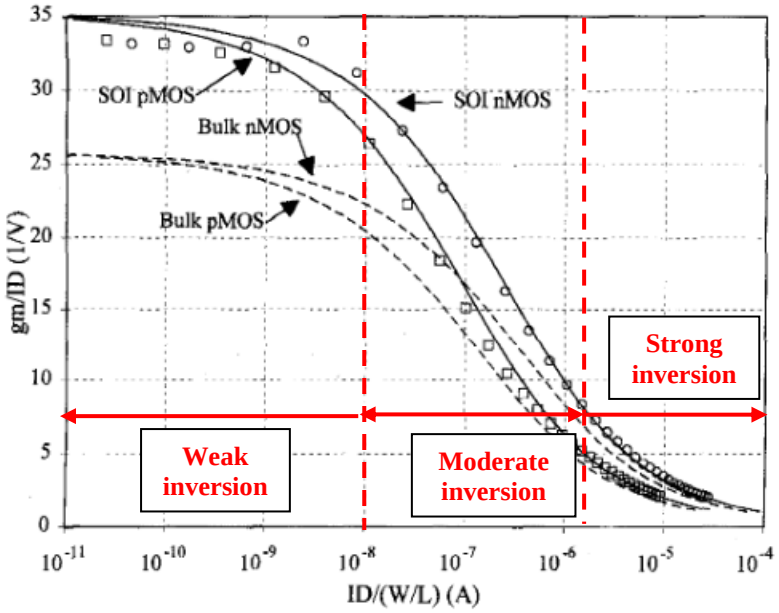


Fig. 2.1: Calculated and measured g_m/I_D vs. $I_D/(W/L)$ curves for both bulk and fully-depleted silicon-on-insulator (SOI) [7] NMOS and PMOS transistors (Fig. 1 from [1])

specifications. The g_m/I_D ratios for the current mirror and cascode transistors are selected for strong inversion operation, primarily to save chip area. The transistor channel lengths (L) ranged from 3µm to 12µm for the 3µm SOI CMOS process. The Early voltage factor which relates to the intrinsic output resistance in saturation is about 7V/µm. The overall g_m/I_D design flow for this example is depicted in Fig. 2.3.

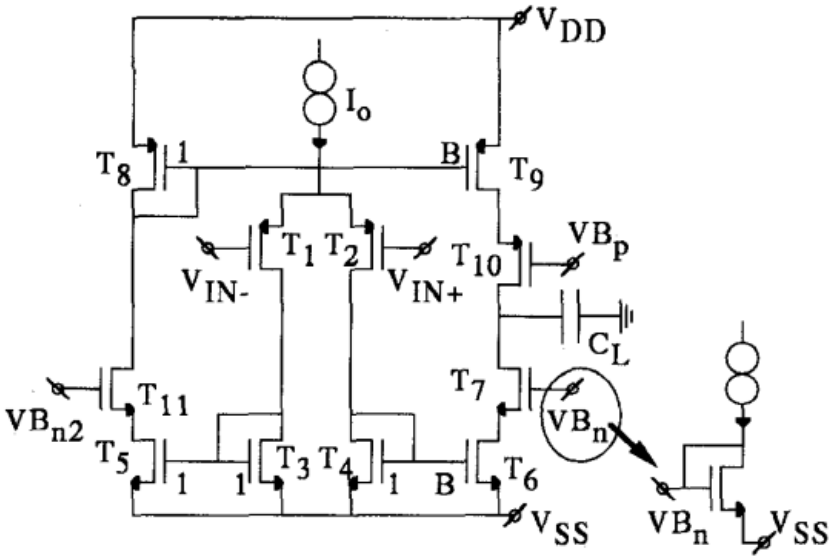


Fig. 2.2: The single-stage OTA designed using the g_m/I_D methodology (Fig. 2 in [1])

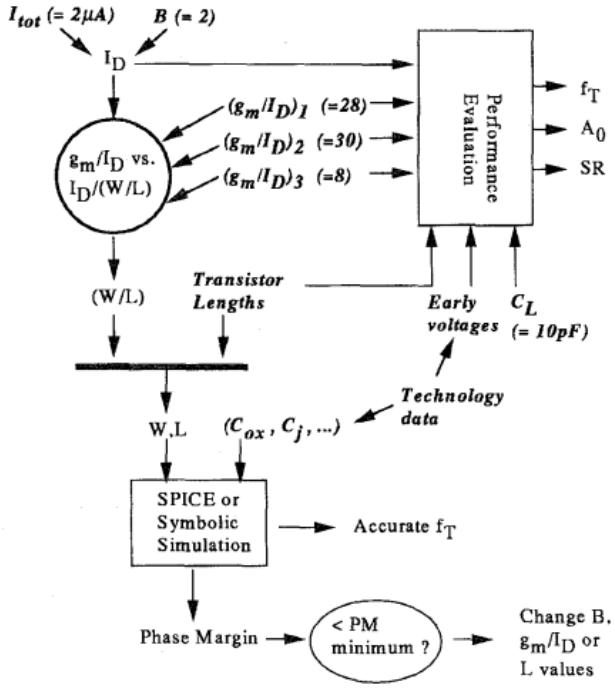


Fig. 2.3: g_m/I_D design flow for the single-stage OTA (Fig. 3 from [1])

2.2 Two-stage OTA Frequency Compensation

The two-stage OTA requires the addition of compensation circuitry in order to guarantee closed-loop stability in the worst-case unity-gain closed-loop negative feedback configuration. The simplest technique is Miller frequency compensation wherein a compensation capacitor, C_C, is connected between the relatively high-impedance output nodes of the first and second gain stages as shown in Fig. 2.4(a) [8]. It effectively creates dominant and non-dominant pole frequencies via the pole-splitting mechanism.

The transfer function of the system is derived using the small-signal model of Fig. 2.4(b) as

$$\frac{Vout(s)}{Vin(s)} = \frac{g_{m1}g_{m6}R_1R_2(1-\frac{SC_C}{g_{m6}})}{S^2R_1R_2(C_cC_1+C_1C_L+C_1C_c)+S\left(g_{m6}R_1R_2C_c+ R_2(C_C+C_L) + R_1(C_1+C_C)\right)+1} \qquad (2.2)$$

It comprises two left-half plane (LHP) poles and one right-half plane (RHP) zero as depicted in Fig. 2.5:

$$P_1' = -\frac{1}{g_{m6}R_2R_1C_C}\ ,\ P_2' = -\frac{g_{m6}}{C_L} \text{ and } Z_1 = \frac{g_{m6}}{C_c} \qquad (2.3)$$

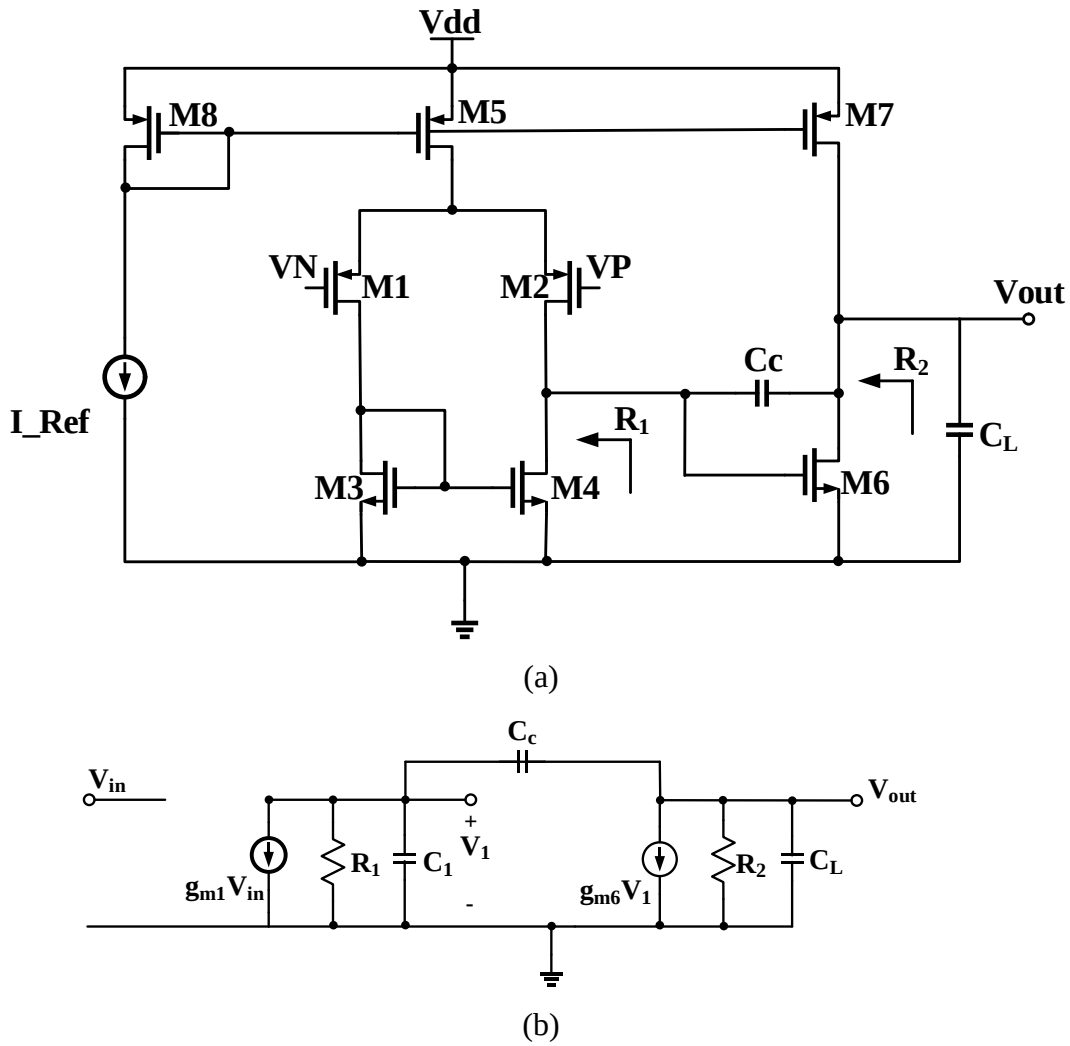


Fig. 2.4: (a) The classic two-stage CMOS OTA with pole-splitting Miller frequency compensation via C_C (after Fig. 9 in [8]) and (b) a simple small-signal model

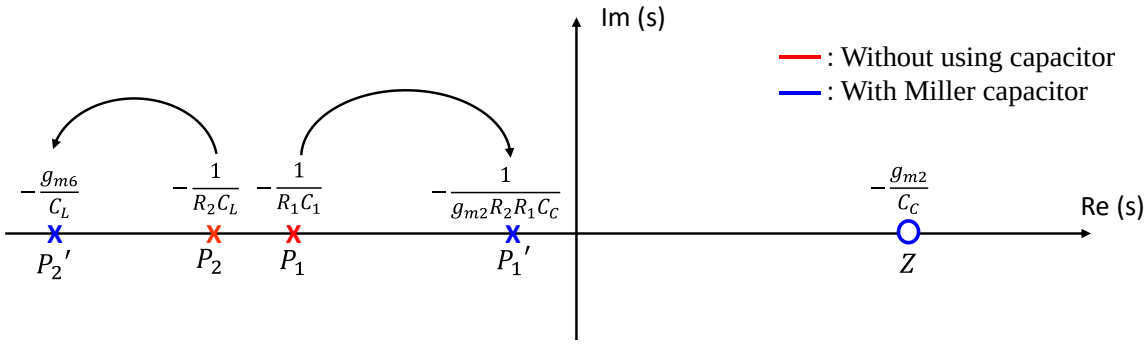


Fig. 2.5: Pole and zero locations of the two-stage OTA before and after frequency compensation using only a Miller capacitor C_C

where P_1' and P_2' are the dominant and non-dominant pole locations, respectively, after adding the Miller capacitor, C_C (Fig. 2.5). If the system only has two poles split widely in frequency, its total phase shift at the unity-gain frequency is usually much less than 180 degrees; i.e., it is stable. However, the simple Miller compensation not only splits the two poles but also introduces a RHP zero which degrades closed-loop stability by decreasing both the unity-gain phase and gain margins. Four different design techniques have been proposed to mitigate the effects of the RHP zero while maintaining the advantages of pole splitting compensation.

2.2.1 Cascode Feedback and Voltage Buffer Feedback Compensation Techniques

As mentioned above, the Miller compensation capacitor, Cc, creates a direct path between the high-impedance output nodes of the first and second gain stages. The resulting RHP zero caused by the direct coupling through C_C around the second stage is explained as follows: At low frequencies, C_C approximates an open circuit so the common-source second stage is inverting with a phase shift of -180°. At high frequencies, however, Cc approaches a short circuit which effectively bypasses the second stage and gives it a phase shift of 0°. Thus, with a negative feedback circuit connected around the OTA, the overall negative feedback at low frequencies becomes positive feedback at high frequencies due to the 180° total phase shift—the OTA becomes unstable.

One technique that blocks feedforward through Cc but still keeps the feedback current flow, $C_c \frac{dV_{out}}{dt}$, from the second stage output back to the first stage output is depicted in Fig. 2.6(a) [9]. M9 and M11 are biased at equal DC currents, grounded-gate cascode device M10 connected between them provides a current steering path, and V_a functions as a virtual ground node [9][10]. As a result, the feedback compensation current flows from the output of the second stage through C_C into the source of M10, and then because M9 and M11 operate at constant DC currents, out of the drain of M10 into the output node of the first gain stage. The transfer function of the equivalent small-signal model of Fig. 2.6(b) is found as:

$$\frac{Vout(s)}{Vin(s)} = \frac{g_{m1}g_{m6}R_1R_2}{S^2R_1R_2C_1(C_c+C_L)+S(R_1C_1+R_2C_c+R_2C_L+g_{m6}R_2R_1C_c)+1} \quad (2.4)$$

$$P_1 = -\frac{1}{g_{m6}R_2R_1C_C} \text{ and } P_2 = -\frac{g_{m6}C_C}{C_1(Cc+C_L)} \quad (2.5)$$

As expected, the RHP zero is eliminated because feedforward through C_C is blocked. The dominant pole frequency is unchanged but the non-dominant pole frequency is increased because of the relatively small C_1 term in the denominator of P_2. Thus, for the same degree of unity-gain stability, an OTA compensated using the cascode feedback technique can drive a larger load in accordance with (2.5).

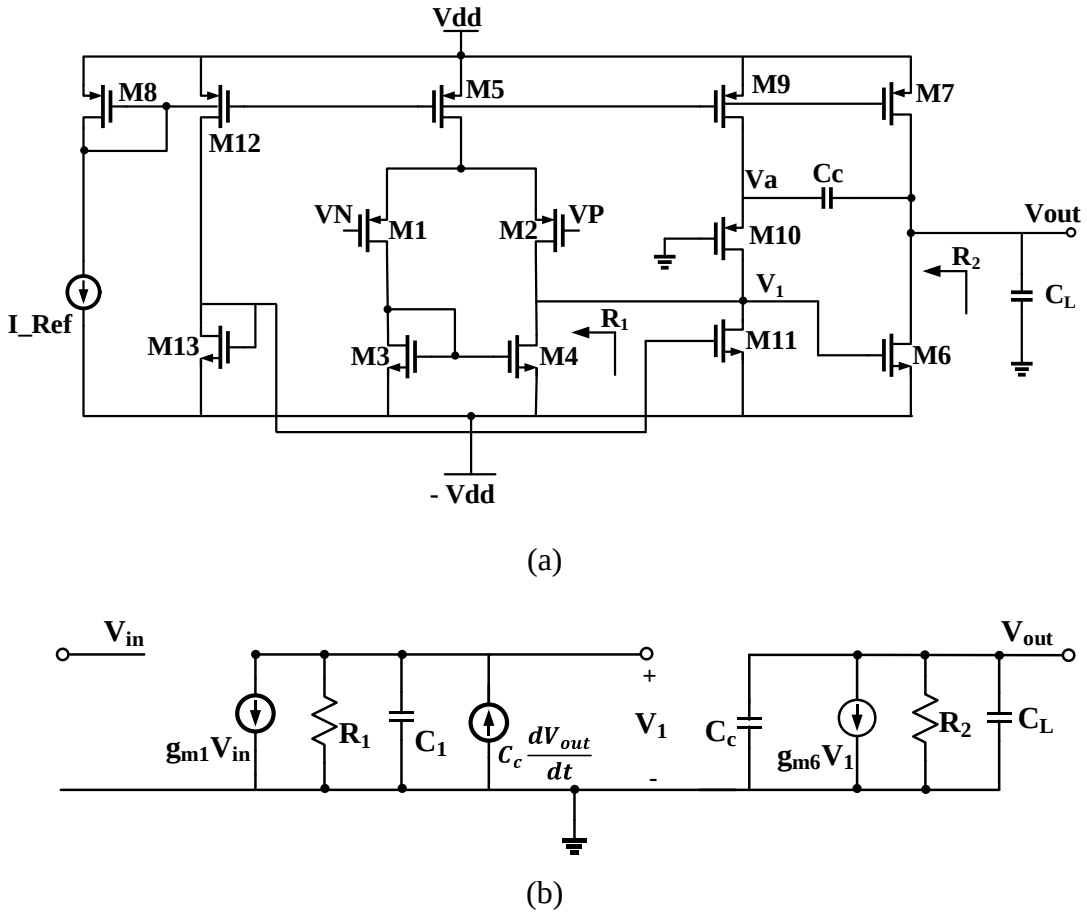


Fig. 2.6: (a) Cascode feedback frequency compensation technique (after Fig. 3(b) in [9]) and (b) equivalent small-signal circuit model

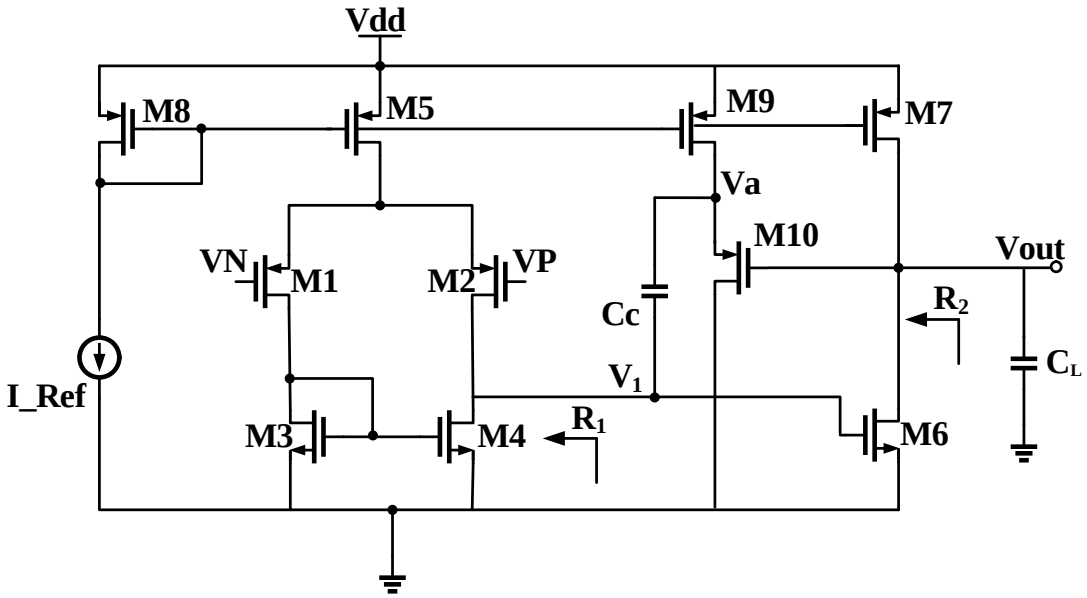
A second technique that blocks feedforward through the compensation network uses a uni-directional unity-gain buffer as shown in Fig. 2.7(a) [11]. Here transistor M10 connected to V_{OUT} acts as a source follower; thus, the small-signal voltage V_a at the output of the buffer which drives C_C, is approximately equal to the feedback voltage, V_{OUT}. By redrawing the small-signal circuit as in Fig. 2.7(b), the transfer function is:

$$\frac{Vout(s)}{Vin(s)} = \frac{g_{m1}g_{m6}R_1R_2}{S^2R_1R_2C_1(C_c+C_L)+S(R_1(C_c+C_L)+R_2C_L+g_{m6}R_2R_1C_c)+1} \quad (2.6)$$

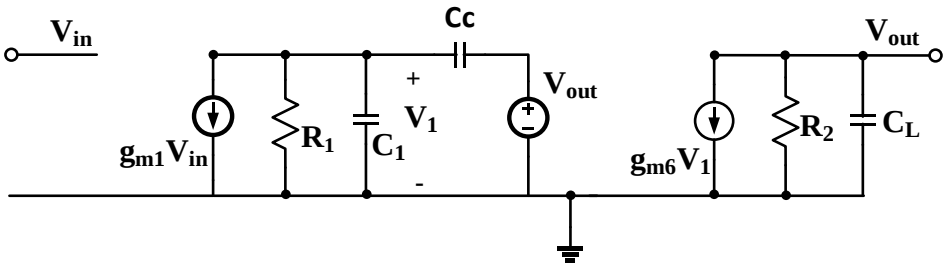
where

$$P_1 = -\frac{1}{g_{m6}R_2R_1C_C} \text{ and } P_2 = -\frac{g_{m6}C_C}{C_L(C_C+C_1)} \quad (2.7)$$

Again, the RHP zero is eliminated. The main drawbacks of this technique are increased power dissipation and decreased output voltage swing owing to the source-follower stage. As compared to the simple Miller compensation of Fig. 2.4 in which the highest output voltage swing is (V_{dd} - $V_{OV\text{-}M7}$), the highest output voltage swing using the voltage buffer is (V_{dd} - $V_{OV\text{-}M9}$ - $V_{SG\text{-}M10}$) which is less by $V_{SG\text{-}M10}$.



(a)



(b)

Fig. 2.7: (a) Indirect feedforward blocking using a unity-gain voltage buffer (after Fig. 6 in [11]), and (b) the corresponding small-signal equivalent circuit

2.2.2 Two-stage OTA with RC Frequency Compensation

Rather than adding power-hungry circuitry to eliminate the RHP zero, an alternative series RC frequency compensation technique (Fig. 2.8(a)) can be used to control its position so that it can be cancelled [6][10][12]. An important advantage of this technique is no additional power consumption. The idea is illustrated via the transfer functions derived from the small-signal model of Fig. 2.8(b):

$$\frac{Vout(s)}{Vin(s)} = \frac{g_{m1}g_{m6}R_1R_2\,[1-s\left(\frac{C_C}{g_{m6}} - R_cC_C\right)]}{aS^3 + bS^2 + cS + 1} \tag{2.8}$$

where

$$a = R_1R_2R_cC_1C_LC_C \tag{2.9}$$

$$b = R_1R_2(C_1C_L + C_CC_1 + C_CC_L) + R_CC_C(R_1C_1 + R_2C_L) \tag{2.10}$$

$$c = R_2(C_L + C_C) + R_1(C_1 + C_C) + g_{m2}R_1R_2C_C + R_CC_C \tag{2.11}$$

Using the dominant-pole approximation,

$$P_1 \simeq -\frac{1}{g_{m6}R_2R_1C_C}\ ,\ P_2 \simeq -\frac{g_{m6}C_C}{C_1C_2 + C_CC_1 + C_CC_2} \simeq -\frac{g_{m6}}{C_L} \text{ and } P_3 \simeq -\frac{1}{R_CC_1} \tag{2.12}$$

$$Z_1 = \frac{1}{C_C\,(\frac{1}{g_{m6}} - R_C)} \tag{2.13}$$

The zero frequency, Z_1, can be controlled by adjusting R_C. Two possible locations for Z_1 are of particular interest. First, with $R_C = 1/g_{m6}$, it is placed at infinity and effectively eliminated. The system now has three poles (2.12) with the highest-frequency pole, P_3, due to R_C; since $P_3 >> P_2$ by design, the unity-gain phase margin is determined primarily by P_2. As a second (more attractive) option, the zero can be moved from the RHP into the LHP and placed on top of P_2 which results in a pole-zero cancellation. The required resistance is

$$R_C = \frac{C_C + C_L}{g_{m6}C_C} \tag{2.14}$$

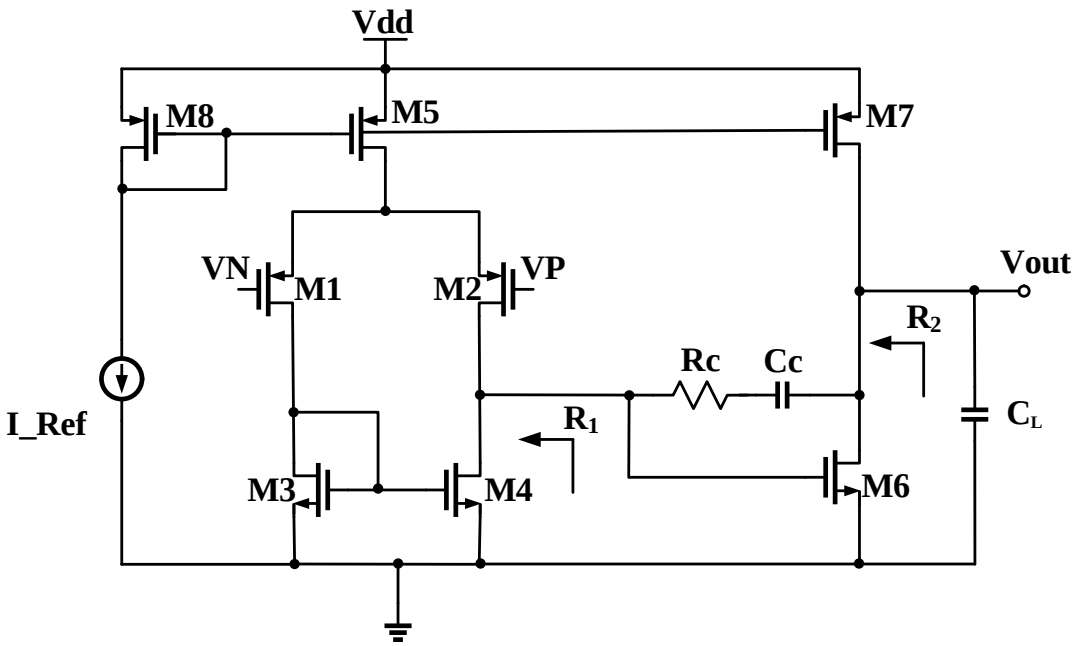
Comparing closed-loop stability in the two cases, the two-stage OTA is more stable when P_2 is cancelled and the non-dominant pole P_3 is placed above the unity-gain frequency [6]. More specifically, when the RHP zero is placed at infinity, the stability is determined by:

Case I: $P_2 > Av\,P_1$ or $Cc > \frac{g_{m1}}{g_{m2}} C_L$

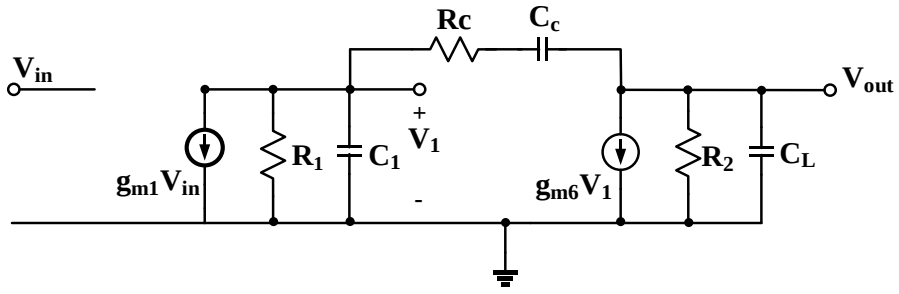
Conversely, if the RHP zero is moved into the LHP to cancel the second pole, the stability follows from:

Case II: $P_3 > Av\,P_1$ or $Cc > \sqrt{\frac{g_{m1}}{g_{m2}} C_1 C_L}$

If the target unity-gain bandwidth is the same in both cases, smaller Miller capacitance means smaller g_{m1} for Case II than Case I. Thus, Case II has several advantages including lower power dissipation and smaller chip area.



(a)



(b)

Fig. 2.8: (a) Two-stage OTA with RC compensation, and (b) the corresponding small-signal circuit equivalent circuit model [6][10][12]

2.2.3 Opamp with Nulling Resistance RC Compensation

In an extension of the previous passive RC compensation technique, the compensation resistance can be implemented using a MOSFET as disclosed by Senderowicz, et al. [12]. More specifically, in the NMOS operational amplifier of Fig. 2.9, depletion-mode MOSFET M31 operates in the triode region and realizes a compensation resistance of value:

$$R_{triode} = \frac{1}{\mu_x C_{ox} \frac{W}{L}(V_{DD} - V_{DS-M18} - V_{th})} \tag{2.15}$$

Note however that since R_{triode} depends on V_{DD}, any variation in V_{DD} changes the value of R_C. It is also obvious that because of the direct connection to V_{DD}, this technique is susceptible to power supply noise coupling into the sensitive signal path. Moreover, variations in V_{DD} also create a frequency doublet with concomitant slow settling-time due to the inexact pole-zero cancellation [13].

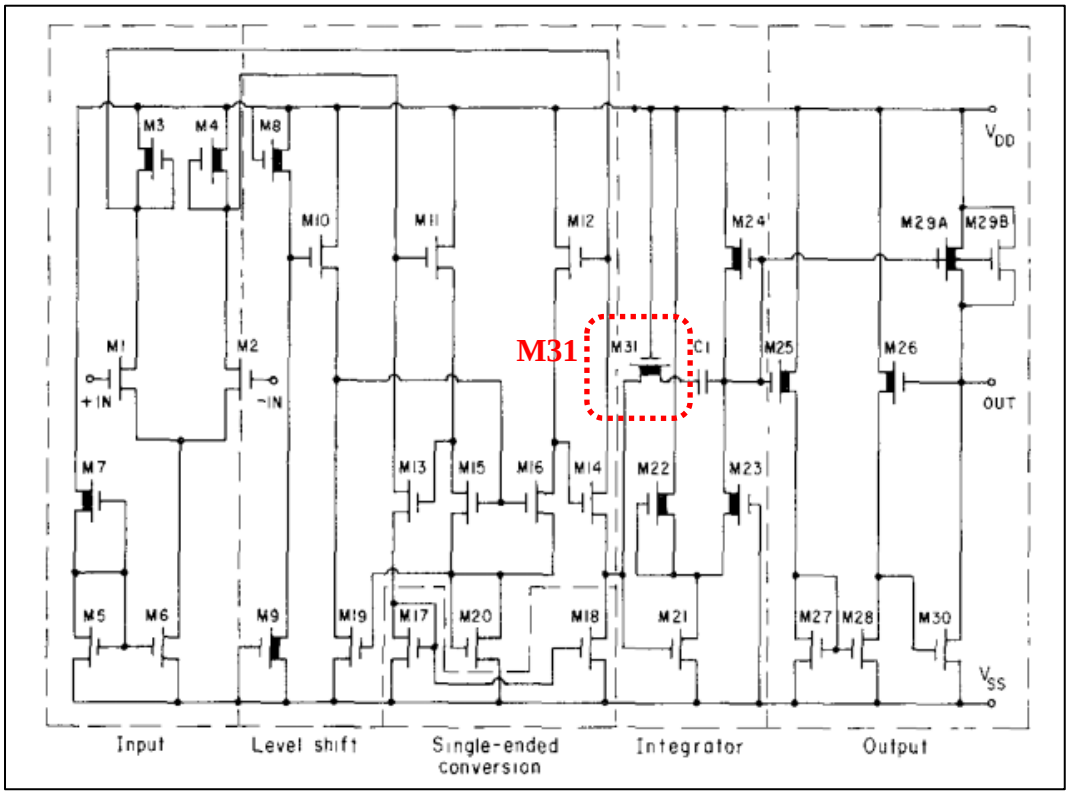


Fig. 2.9: NMOS depletion-load three-stage operational amplifier with active RC compensation via M31 (Fig. 10 in [12])

2.2.4 Two-stage OTA with Nulling Resistor PVT-insensitive Compensation

A robust technique of frequency compensation using an active nulling resistor that also provides PVT insensitivity is introduced by Black, et al. [6].

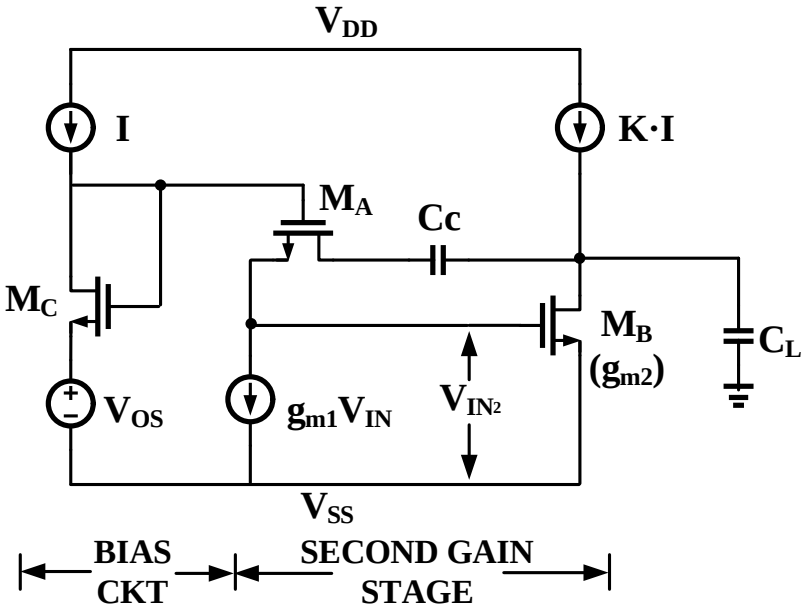


Fig. 2.10: Pole-zero PVT-tracking RC compensation and simplified DC bias circuitry (after Fig. 7 in [6])

With reference to the two-stage CMOS OTA of Fig. 2.8, the compensation resistance R_C is implemented using NMOS M_A in Fig. 2.10 biased in the triode region. Rather than connecting the gate terminal to V_{DD} as in Fig. 2.9, however, M_A is DC biased by diode-connected NMOS M_C in series with V_{OS}. The bias circuitry is designed using replication so that $V_{OS} = V_{GS\_MB}$ and $V_{GS\_MC} = V_{GS\_MA}$. Using simple square-law derivations neglecting channel-length modulation, I_D and g_m are expressed as

$$I_D = \frac{\mu_x C_{ox}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_{th})^2 \tag{2.16}$$

$$g_m = \sqrt{2 \cdot I_D \cdot \mu_x C_{ox} \cdot \frac{W}{L}} \tag{2.17}$$

The necessary condition for pole-zero cancellation is

$$\left(\frac{W}{L}\right)_A \cong \left[\left(\frac{W}{L}\right)_B \cdot \left(\frac{W}{L}\right)_C \cdot K\right]^{\frac{1}{2}} \cdot \frac{C_c}{C_c + C_L} \tag{2.18}$$

Where

$$K \cong \frac{I_{MB}}{I_{MC}} \tag{2.19}$$

Since the pole-zero cancellation depends only on (W/L) ratios and a capacitance ratio, this compensation method guarantees tracking of the ideally cancelled pole and zero frequencies versus PVT variations. This technique has been used in the production of billions of embedded CMOS two-stage opamps for more than four decades [6][14].

Chapter 3: g_m/I_D-based Design of a CMOS Two-stage OTA with PVT-insensitive Nulling Resistor Frequency Compensation

3.1 MOSFET Triode-Region g_m/I_D Techniques [15]

The use of a MOSFET operated in the triode region to realize PVT-insensitive RC frequency compensation is detailed in Chapter 2. With reference to Fig. 3.1, it is achieved when the triode region small-signal conductance of M_{12} (g_{12}) takes the same form as the saturation region small-signal transconductance of M_6 (g_{m6}) [6]. PVT-insensitivity is evident because the aspect ratio of M_{12} depends only on capacitance and (W/L) ratios:

$$\left(\frac{W}{L}\right)_{12} \cong \left[\left(\frac{W}{L}\right)_{6} \cdot \left(\frac{W}{L}\right)_{11} \cdot K\right]^{\frac{1}{2}} \cdot \frac{C_c}{C_c + C_L} \tag{3.1}$$

This chapter extends the g_m/I_D methodology to the design of MOSFETs operated in the triode region for PVT-insensitive RC frequency compensation. For example, the two-stage OTA of Fig. 3.1 shows the NMOS triode transistor M_{12} and its associated DC bias circuitry used to realize R.

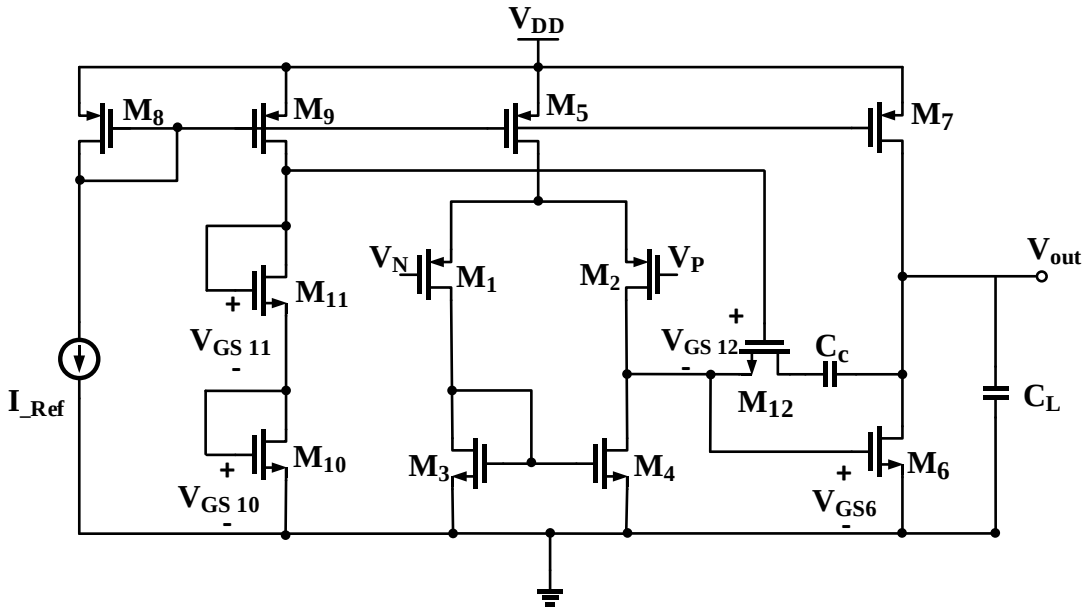


Fig. 3.1: CMOS two-stage OTA with PVT-insensitive RC frequency compensation wherein M_{12} operated in the triode region acts as the compensation resistance (after Fig. 8(a) in [6])

The small-signal conductance of triode transistor M_{12} which acts as the series compensation conductance (i.e., $1/R_{12}$) is derived as

$$g_{12} = \frac{\partial I_D}{\partial V_{DS}}\Big|_{V_{DS}=0} = \frac{\partial}{\partial V_{DS}} \left\{ \mu_x C_{ox} \left(\frac{W}{L}\right)_{12} \left[(V_{GS} - V_{TH})V_{DS} - \frac{{V_{DS}}^2}{2}\right]\right\} \quad (3.2)$$

$$= \mu_x C_{ox} \left(\frac{W}{L}\right)_{12} V_{OV12} \quad (3.3)$$

and

$$R_{12} = {g_{12}}^{-1} = \frac{1}{\mu_x C_{ox} \left(\frac{W}{L}\right)_{12} V_{OV12}} \quad (3.4)$$

where $V_{OV} = V_{GS} - V_{th}$ is the excess or overdrive voltage. From (3.3), g_{12} depends on design variables $(W/L)_{12}$ and V_{OV12}, where $V_{GS12} = V_{GS10} + V_{GS11} - V_{GS6}$ (Fig. 3.1). For PVT-insensitivity, $V_{GS6} = V_{GS10}$ and $V_{GS12} = V_{GS11}$; thus, $V_{OV12} = V_{OV11}$. Eqn. (3.3) is normalized to conductance per unit width as

$$\frac{G12}{W_{12}} = \frac{\mu_x C_{ox}}{L} Vov_{12} \quad (3.5)$$

With $V_{OV12} = V_{OV11}$, (3.5) can also be written as:

$$\frac{G12}{W_{12}} = \frac{\mu_x C_{ox}}{L} Vov_{11} \quad (3.6)$$

g_m/I_D decreases from weak to strong inversion (Fig. 3.2). In this design, all devices are operated in moderate inversion with $V_{OV} = 0.1$ V where $g_m/I_D = 10$ S/A for the NMOS transistors. As a lower-power alternative design, however, the input pair, M_1-M_2 (Fig. 3.1) may be operated in weak inversion with the optimum value of V_{OV} (Chapter 4).

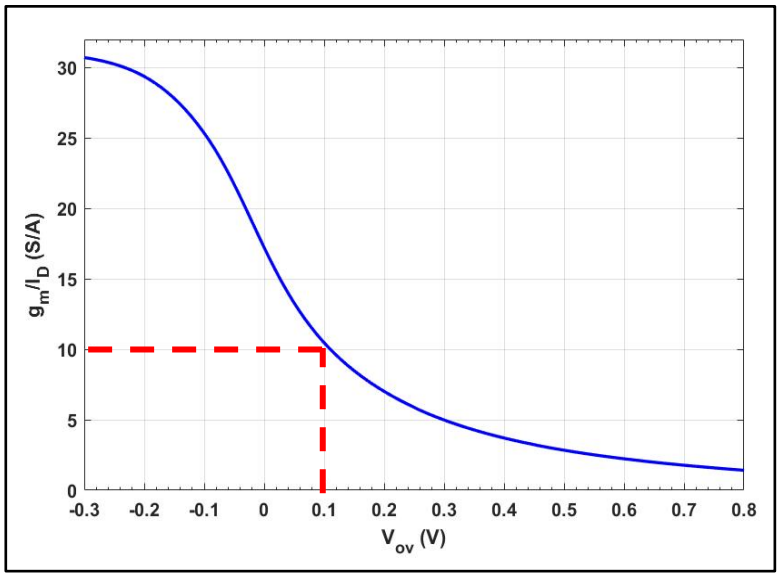


Fig. 3.2: NMOS g_m/I_D vs. V_{OV} (180nm CMOS process)

Thus, from (3.5), (3.6) and Fig. 3.2, there is a one-to-one correspondence between G_{12}/W_{12} and g_m/I_D. The *SPICE* plot of G_{triode}/W vs. g_m/I_D for a practical range of L values (Fig. 3.4) can be generated using identical NMOS devices biased as in Fig. 3.3.

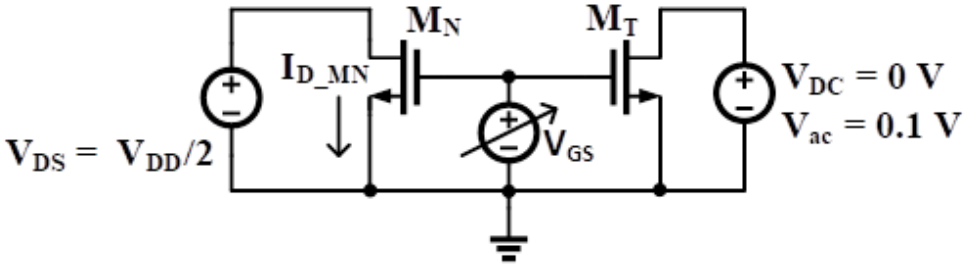


Fig. 3.3. Bias circuits used to determine (via *SPICE*) the normalized NMOS small-signal triode region conductance G_{triode}/W vs. g_m/I_D (Fig. 5 from [15])

MN and MT are driven by gate bias voltage, V_{GS}, which is swept from 0 to V_{DD} = 1.8V for the 180nm CMOS process. The drain bias of MN is set to $V_{DD}/2$ for operation in saturation as in the standard g_m/I_D lookup tables [1][2]. On the other hand, the drain bias of MT (like M_{12} in Fig. 3.1) is set to 0 V in the triode region. The main purpose of the *SPICE* simulations is to extract G_{triode} vs. V_{OV} for MT and g_m/I_D vs. V_{OV} for MN. Because MN and MT have equal V_{OV} values, the g_m/I_D values of MN and MT are also the same as derived above. Finally, G_{triode}/W vs. g_m/I_D for a range of L values can be plotted as in Fig. 3.4.

3.2 Systematic Design of a Two-stage OTA using the Enhanced g_m/I_D Methodology

In accordance with standard practice, several calculations and assumptions are made before beginning the g_m/I_D design process. Typical target specifications are listed in Table I for the prototype OTA in Fig. 3.1, and several performance parameters are then calculated based on the specifications.

A. Low-frequency Small-signal Voltage Gain

The overall low-frequency gain of the two-stage OTA is the product of the voltage gains of the first and second stage*s*:

$$A_{total} = A_1 A_2 \tag{3.7}$$

where

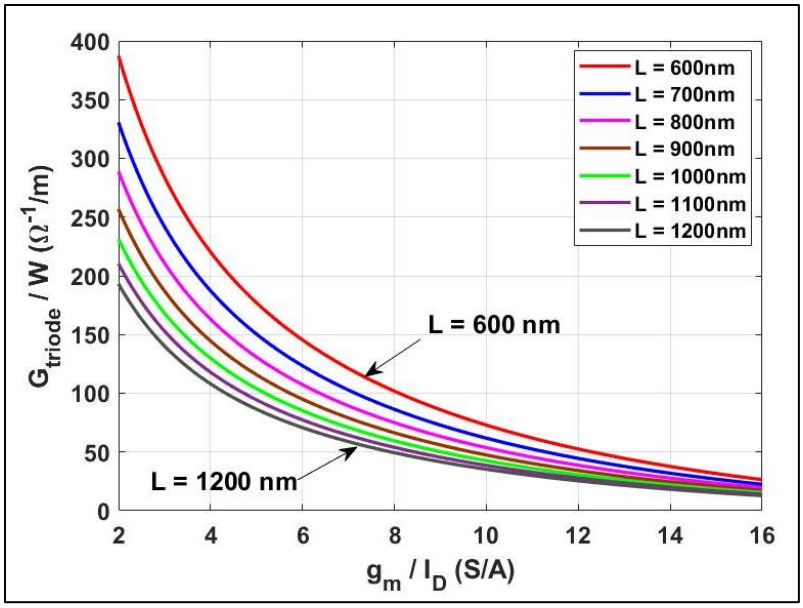


Fig. 3.4: NMOS G_{triode}/W vs. g_m/I_D over a typical range of L values (Fig. 6 from [15])

Design parameter	Specification	Target with margin
Voltage Gain (dB)	> 60	70
UGBW (MHz)	> 50	70
Phase Margin (◦)	> 60	70
Input-referred Noise @ 10MHz (nV/√Hz)	<15	10
C_L (pF)	5	5

Table 3.1. Target Design Specifications

$$A_1 = g_{m1}(r_{o1} || r_{o4}) = \left[\frac{(g_m/I_D)_{m1}}{(g_m/I_D)_{m4}} \cdot g_{m4}r_{o4}\right] = A_{IP} \,||\, \left[\frac{(g_m/I_D)_{m1}}{(g_m/I_D)_{m4}} A_{IN}\right] \quad (3.8)$$

$$A_2 = g_{m6}(r_{o6} || r_{o7}) = \left[\frac{(g_m/I_D)_{m6}}{(g_m/I_D)_{m7}} \cdot g_{m7}r_{o7}\right] = A_{IN} \,||\, \left[\frac{(g_m/I_D)_{m6}}{(g_m/I_D)_{m7}} \cdot A_{IP}\right] \quad (3.9)$$

and A_{IP} and A_{IN} are the *intrinsic* voltage gains of the PMOS and NMOS devices.

B. Unity-gain Bandwidth (UGBW)

$$\omega_u = A_{total}\,\omega_{p1}$$

$$= \frac{g_{m1}}{C_c} = 2\pi \text{ UGBW} \quad (3.10)$$

where C_C is determined by the input-referred thermal noise specification (Table I).

C. Unity-gain Phase Margin

The overall frequency response of the two-stage OTA of Fig. 3.1 comprises three LHP poles and one zero. Ideally, the non-dominant pole, P_2, is cancelled by Z_1 after which the system has only two poles. Thus, P_1 is still the dominant pole but P_3 is now the non-dominant pole which determines the unity-gain phase margin (PM):

$$\text{Phase margin (PM)} = 180 - tan^{-1}(\frac{\omega_u}{\omega_{p1}}) - tan^{-1}(\frac{\omega_u}{\omega_{p3}}) \quad (3.11)$$

Since $\omega_{p1} << \omega_u$, $tan^{-1}(\frac{\omega_u}{\omega_{p1}}) \simeq 90^\circ$. Thus, (3.11) can be approximated as

$$\text{Phase margin (PM)} = 90^\circ - tan^{-1}(\frac{\omega_u}{\omega_{p3}}) \quad (3.12)$$

For a specified value of PM, the location of ω_{p3} is

$$\omega_{p3} = \frac{\omega_u}{tan(90^\circ - PM)} \quad (3.13)$$

Substituting $\omega_{p3} = \frac{g_{m6}}{C_1(1+\frac{C_L}{C_c})}$ in (3.13) gives

$$\frac{g_{m6}}{C_1} = \frac{\omega_u}{tan(90^\circ - PM)} (1 + \frac{C_L}{C_c}) \quad (3.14)$$

where $C_1 = C_{dd-M2} + C_{dd-M4} + C_{gg-M6}$ (C_{ss-M12} is neglected) is the total parasitic capacitance at the output node of the first stage. Because these devices are all designed to operate in saturation in moderate inversion, C_{gg-M6} is the dominant component of C_1. Therefore, $C_1 \simeq C_{gg-M6}$ substituted in (3.14) gives:

$$f_{T\_m6} = \frac{g_{m6}}{2\pi \cdot C_{gg-M6}} = \frac{\omega_u}{2\pi \cdot tan(90^\circ - PM)} (1 + \frac{C_L}{C_c}) \quad (3.15)$$

From (3.15), high f_T is required for a large PM. Referring to Fig. 3.5 which plots f_T vs. g_m/I_D for the NMOS devices, f_T increases with shorter L values as is well known. Thus, once g_m/I_D is determined, the final channel length of M_6 (L_6) is found from the PM specification [16].

D. Input-referred Thermal Noise

The equivalent input-referred thermal noise power spectral density of the two-stage OTA of Fig. 3.1 is

$$S_n(\text{f}) = \frac{8KT\gamma}{g_{m1}} \left[1 + \frac{g_{m4}}{g_{m1}}\right] \quad (3.16)$$

where $\gamma = 2/3$ for long-channel transistors and is higher for submicron MOSFETs.

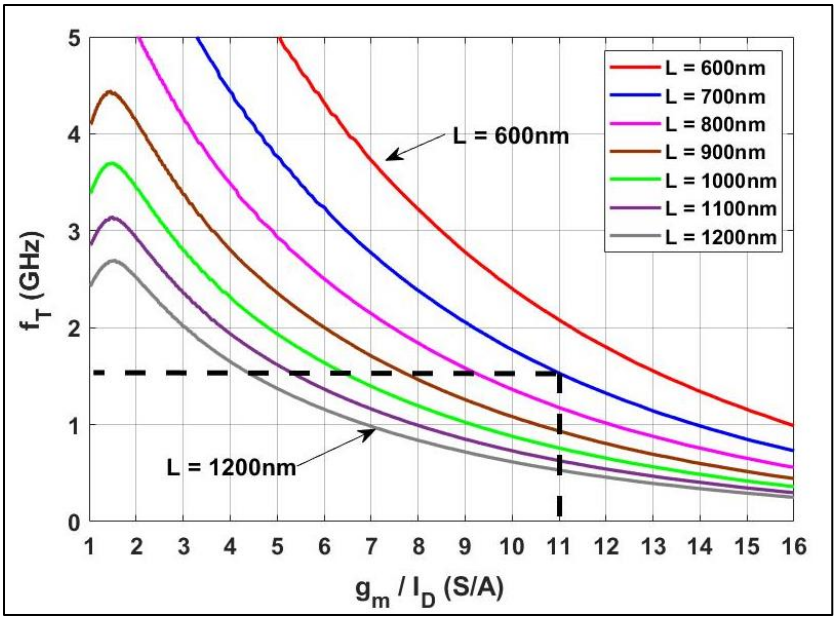


Fig. 3.5: NMOS f_T vs. g_m/I_D for a range of L values. f_T = 1.5 GHz at g_m/I_D = 11 S/A and L = 700 nm (Fig. 7 in [15])

Using (3.10) with $I_{D1} = I_{D2} = I_{D4}$ gives:

$$S_n(\mathrm{f}) = \frac{8KT\gamma}{\omega_u C_C}\left[1 + \frac{\left(\frac{g_m}{I_D}\right)_4}{\left(\frac{g_m}{I_D}\right)_1}\right] \tag{3.17}$$

3.2.1 Step-by-step g_m/I_D Design Flow for a Two-stage OTA [15]

Step 1: Specify V_{OV} for all devices

Generally, V_{OV} is determined by either the output voltage swing specification or the region of operation as follows: weak (V_{OV} < 0 V), moderate (0 < V_{OV} < 0.2 V) or strong inversion (V_{OV} > 0.2 V) [1][2]. For simplicity in this design, all transistors operate with V_{OV} = 0.1 V. Hence, the corresponding g_m/I_D values are 11 S/A, and 10 S/A for the NMOS (Fig. 3.6) and PMOS (Fig. 3.7) devices, respectively.

Step 2: Determine the Compensation Capacitance, C_C

C_C is determined from the input-referred noise (3.17) and ω_u (3.10) specifications. The desired thermal noise voltage spectral density is 10 nV/√Hz with a UGBW = 70 MHz. These target values include practical margins in order to satisfy the specifications over all possible PVT process corners. Finally, C_C = 1 pF for this design.

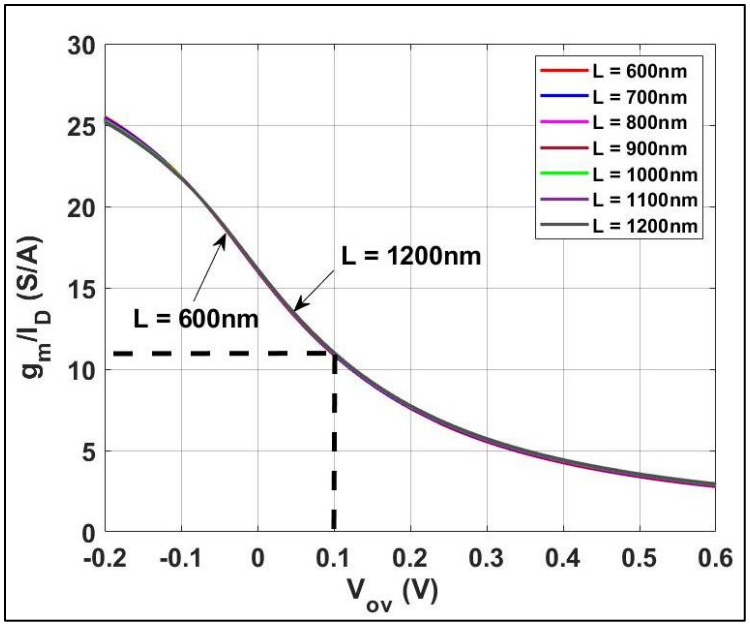


Fig. 3.6: NMOS g_m/I_D vs. V_{OV}. For all devices V_{OV} = 0.1 V and g_m/I_D = 11 S/A (Fig. 3 in [15])

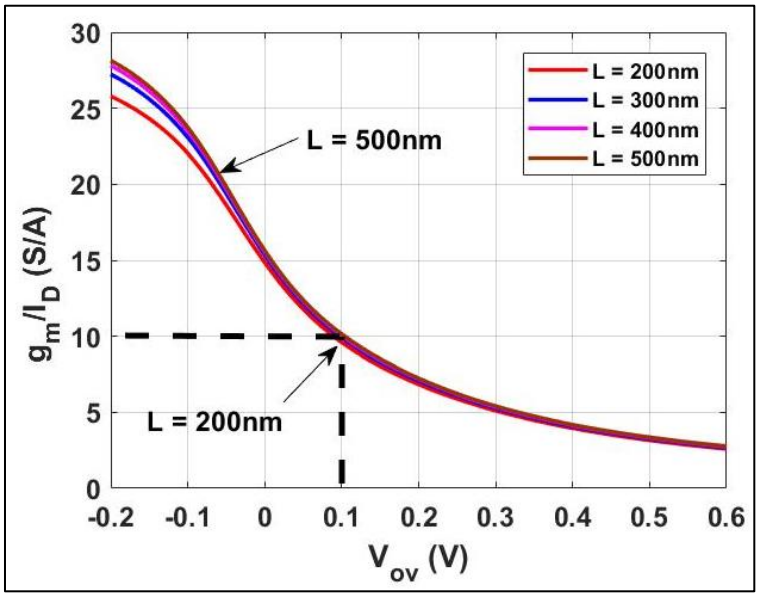


Fig. 3.7: PMOS g_m/I_D vs. V_{OV}. For all devices V_{OV} = 0.1 V and g_m/I_D = 10 S/A (Fig. 4 in [15])

Step 3: Find the PMOS and NMOS L Values

From Table 3.1, key targets are UGBW = 70 MHz and PM = 70°. PM = 70° also enables near-minimum small-signal settling time as depicted in Fig. 3.8 [17]. From (3.15), PM = 70° requires $f_T \simeq 1.15$ GHz. Thus, Fig. 3.5 shows that at g_m/I_D = 11 S/A,

the NMOS length (L) should be L ≤ 800 nm. As described earlier, (3.15) assumes that $C_{gg\text{-}M6}$ is the dominant component of C_1. To ensure a sufficient design margin, therefore, the NMOS devices are designed with L = 700 nm for $f_T \simeq 1.5$ GHz. The corresponding intrinsic gain is A_{IN} = 45 dB as shown in Fig. 3.9.

The PMOS length (L) is now chosen to meet the overall voltage gain specification assuming the small-signal drain resistance r_{ds} α L. The target total voltage gain is A_V = 70 dB. Initially, A_V = 70 dB (3162 V/V) is (arbitrarily) divided equally between the first (A_1) and second (A_2) stages with each at 35 dB (56 V/V). But, the length of the NMOS (L = 700 nm) which corresponds to A_{IN} = 45 dB (178 V/V) has already been determined. Thus, A_{IP} = 39 dB (89 V/V) can be calculated using (3.7) or $A_{IP} \simeq 37$ dB (71 V/V) using (3.9). Since A_{IN} > 35 dB gives some margin, the PMOS L = 300 nm can be chosen which gives A_{IP} = 38 dB (79 V/V) with g_m/I_D = 10 S/A as shown in Fig. 3.10. Substituting A_{IN} = 45 dB (178 V/V) with g_m/I_D = 11 S/A and A_{IP} = 38 dB (79 V/V) with g_m/I_D = 10 S/A in (3.8) and (3.9) gives $A_{total} = A_1A_2 \simeq [(53\ \text{V/V})(58\ \text{V/V})] \simeq 3074\ \text{V/V} \simeq 70$ dB.

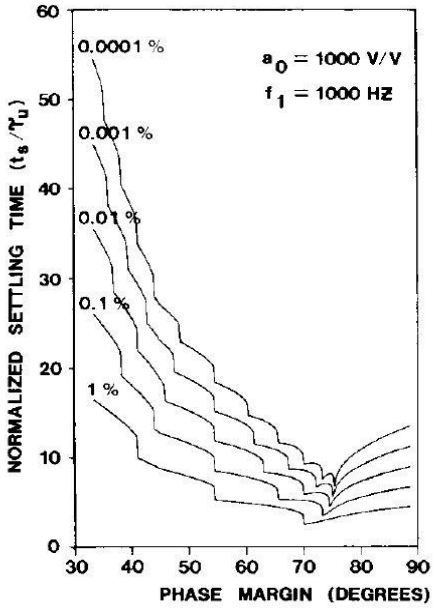


Fig. 3.8: Normalized small-signal settling time vs. unity-gain phase margin for a two-stage OTA (Fig. 3 from [17])

Step 4: Determine the Small-signal Transconductances

The small-signal transconductance is calculated using (3.10), g_{m1-2} = 0.44 mS

with UGBW = 70 MHz. The transconductance of the second stage g_{m6} is determined by the required cancellation of P_2 and Z_1. By (somewhat arbitrarily) choosing ω_{p2} = 30 MHz, g_{m6} = 0.94 mS is calculated from (2.12).

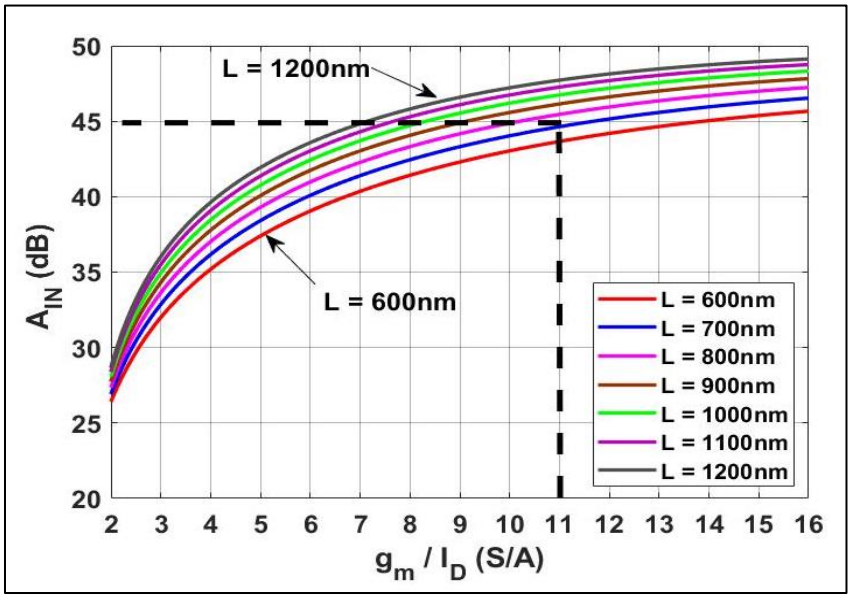


Fig. 3.9: Intrinsic NMOS small-signal voltage gain A_{IN} vs. g_m/I_D. A_{IN} = 45 dB for L = 700 nm and g_m/I_D = 11 S/A (Fig. 8 in [15])

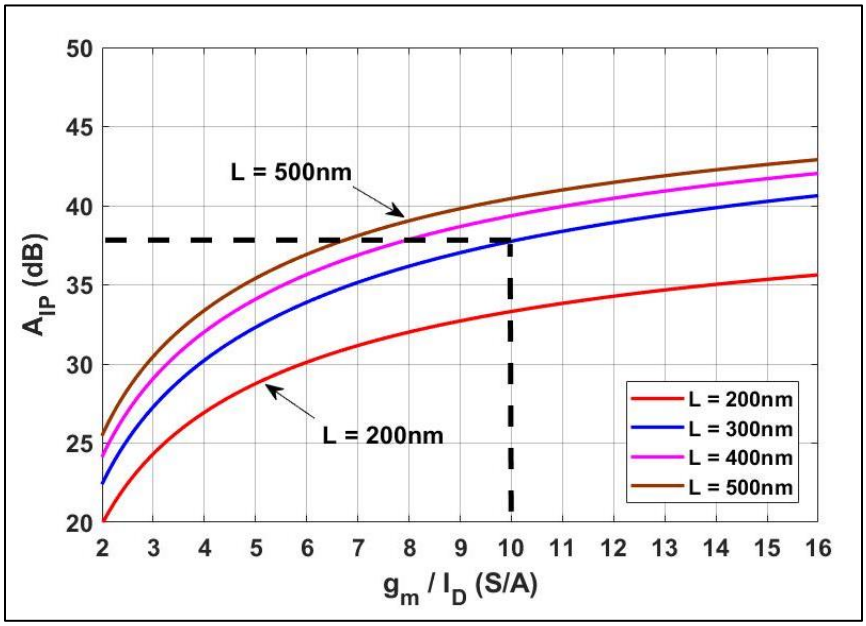


Fig. 3.10: Intrinsic PMOS small-signal voltage gain A_{IP} vs. g_m/I_D. A_{IP} = 38 dB for L = 300 nm and g_m/I_D = 10 S/A (Fig. 9 in [15])

Step 5: Sizing Transistors M_1-M_{12}

The final PMOS W/L ratios are found using the corresponding current density plot in Fig. 3.11. It is assumed for this design (Fig. 3.1) that $I_{Ref} = I_8 = I_9 = 20$ µA. Thus, from Fig. 3.11, $I_D/W = 4$ A/m with $g_m/I_D = 10$ S/A, and $W_8 = W_9 = 5$ µm. $g_{m1,2} = 0.44$ mS requires $I_1 = I_2 = 44$ µA with $W_1 = W_2 = 11$µm. $I_5 = 2I_1 = 88$ µA so $W_5 = 22$ µm. W_7 is determined below.

The final NMOS W/L ratios are found from Fig. 3.12 where $g_m/I_D = 11$ S/A and $I_D/W = 4$ A/m. NMOS devices M_{10} and M_{11} are biased at $I_{10} = I_{11} = 20$ µA so $W_{10} = W_{11} = 5$ µm. NMOS simple current mirror devices M_3 and M_4 are biased at $I_3 = I_4 = 44$ µA so $W_3 = W_4 = 11$ µm. With $g_{m6} = 0.94$ mS, $I_6 = 85$ µA with $W_6 = 21.25$ µm. For PMOS M_7 with $I_6 = I_7 = 85$ µA, $W_7 = 21.25$ µm. M_{12} is sized using Fig. 3.13 and (2.14) which corresponds to $R_{12} = 6383$ Ω:

$$W_{12} = \left.\frac{G_{12}}{\frac{G_{Triode}}{W_{12}}}\right|_{\left(\frac{g_m}{I_D} = 11\right)} \tag{3.18}$$

Therefore, $W_{12} = 3$ µm. The final design values ($I_m/W = 4$ A/m for M_1 - M_{11}) are given in Table 3.2.

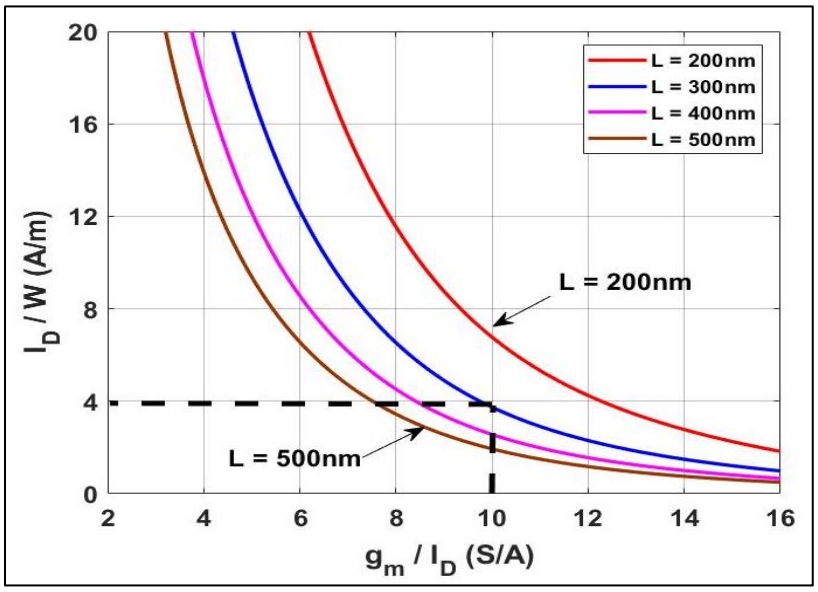


Fig. 3.11: Normalized PMOS I_D/W vs. g_m/I_D for a range of L values. $I_D/W = 4$ A/m at $g_m/I_D = 10$ S/A and L = 300 nm (Fig. 10 in [15])

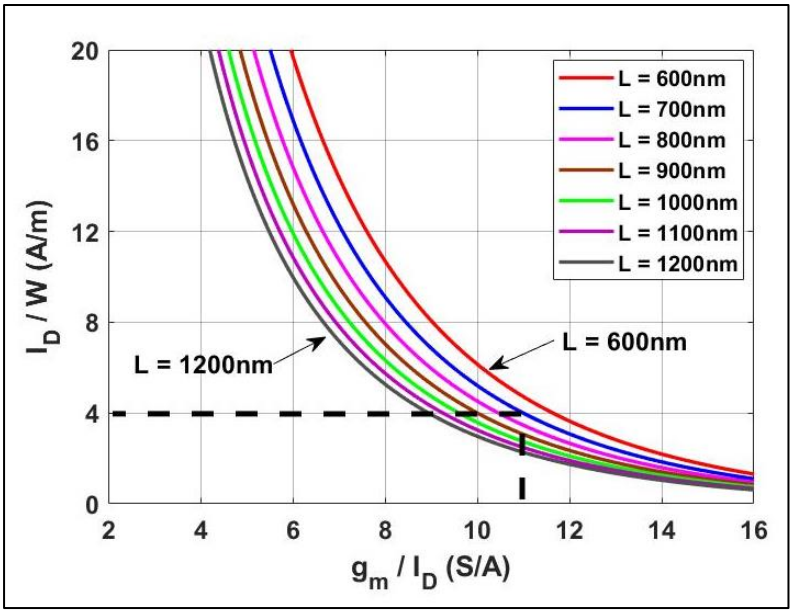


Fig. 3.12: Normalized NMOS I_D/W vs. g_m/I_d for a range of L values. I_D/W = 4 A/m at g_m/I_D = 11 S/A and L = 700 nm (Fig. 11 in [15])

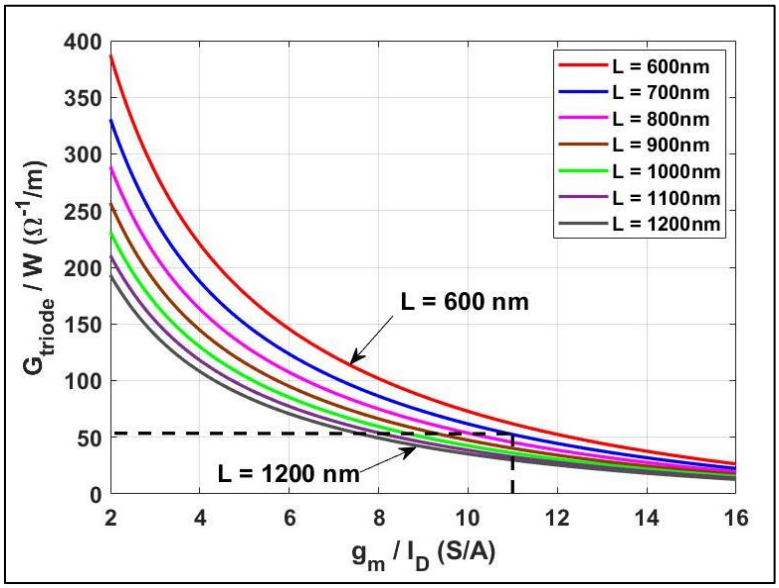


Fig. 3.13: Normalized NMOS G_{triode}/W vs. g_m/I_D for a range of L values. G_{triode}/W = 52 1/Ωm at g_m/I_D = 11 S/A with L = 700 nm (Fig. 12 in [15])

	g_m/I_D (S/A)	I_D (µA)	G_{triode}/W (Ω^{-1}/m)	W/L (µm/µm)
$M_{1,2}$ (PMOS)	10	20	-	11/0.3
$M_{3,4}$ (NMOS)	11	44	-	11/0.7
M_5 (PMOS)	10	88	-	22/0.3
M_6 (NMOS)	11	20	-	21.25/0.7
M_7 (PMOS)	10	44	-	21.25/0.3
$M_{8,9}$ (PMOS)	10	85	-	5/0.3
$M_{10,11}$ (NMOS)	11	85	-	5/0.7
M_{12} (NMOS)	11	-	52	3/0.7

Table 3.2. Nominal W and L Values for the 180 nm CMOS OTA of Fig. 3.1 [15]

3.3 *SPICE* Validation of the g_m/I_D-based Design of a CMOS Two-stage OTA with PVT-insensitive Nulling Resistor Frequency Compensation

The device sizes (Table 3.2) for the g_m/I_D-based design of the 180 nm CMOS two-stage OTA of Fig. 3.1 were determined assuming a nominal power supply voltage (V_{DD} = 1.8 V), room temperature (T = 27°C), and typical process parameters; i.e., the TT2 "process corner" in Table 3.3.

Process Corner	Operating Temperature (°C)	V_{DD} Supply Voltage (V)
SS1	85	1.7
SS2	27	1.8
SS3	-40	1.9
TT1	85	1.7
TT2	27	1.8
TT3	-40	1.9
FF1	85	1.7
FF2	27	1.8
FF3	-40	1.9

Table 3.3. 180 nm CMOS Process Corners [15]

SPICE simulations were used to validate the design. The open-loop voltage gain and inverting phase shift responses of Fig. 3.14 predict a nominal low-frequency gain of 68 dB, unity-gain bandwidth of 58.5 MHz, and unity-gain phase margin of 68°. Fig. 3.15 plots the input-referred voltage noise spectral density, which is 13.3 nV/[Hz]$^{0.5}$ at 10 MHz. The closed-loop unity-gain frequency response is plotted in Fig. 3.16. All of the target design specifications are met at the nominal TT2 process corner as validated via *SPICE* (Table 3.4).

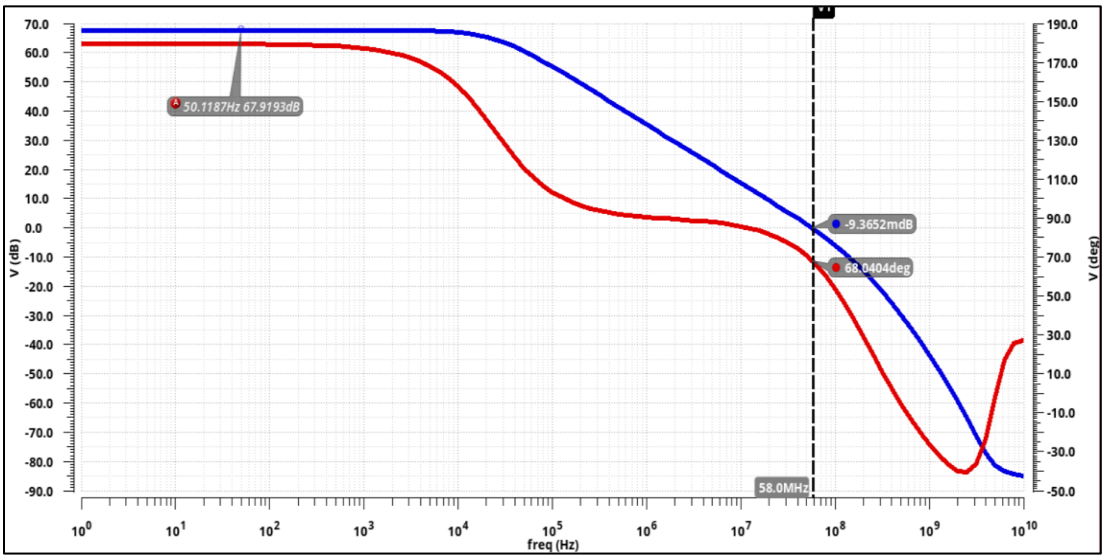


Fig. 3.14: Open-loop frequency response at the TT2 process corner

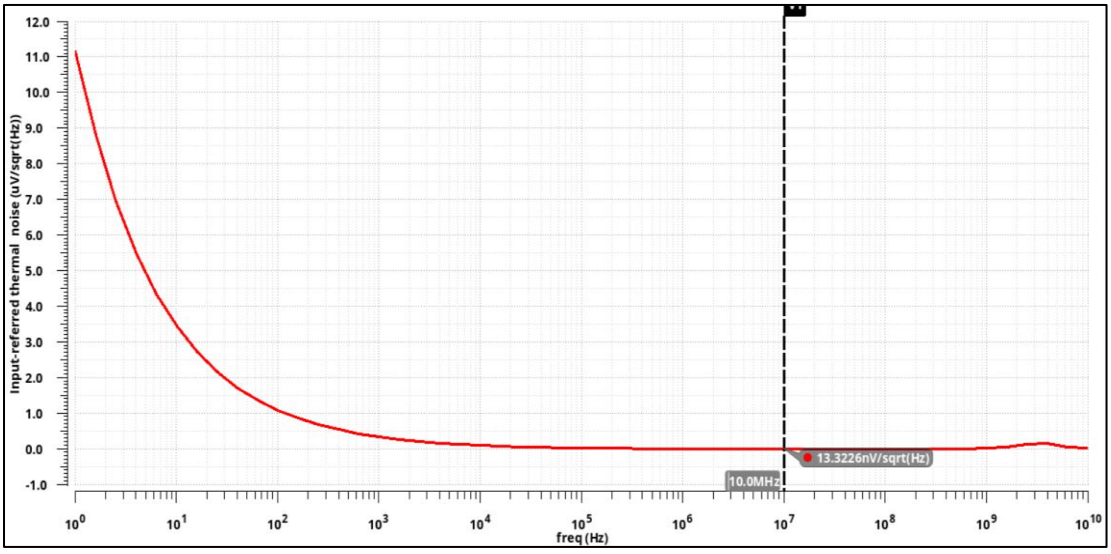


Fig. 3.15: Input-referred noise voltage spectral density vs. frequency

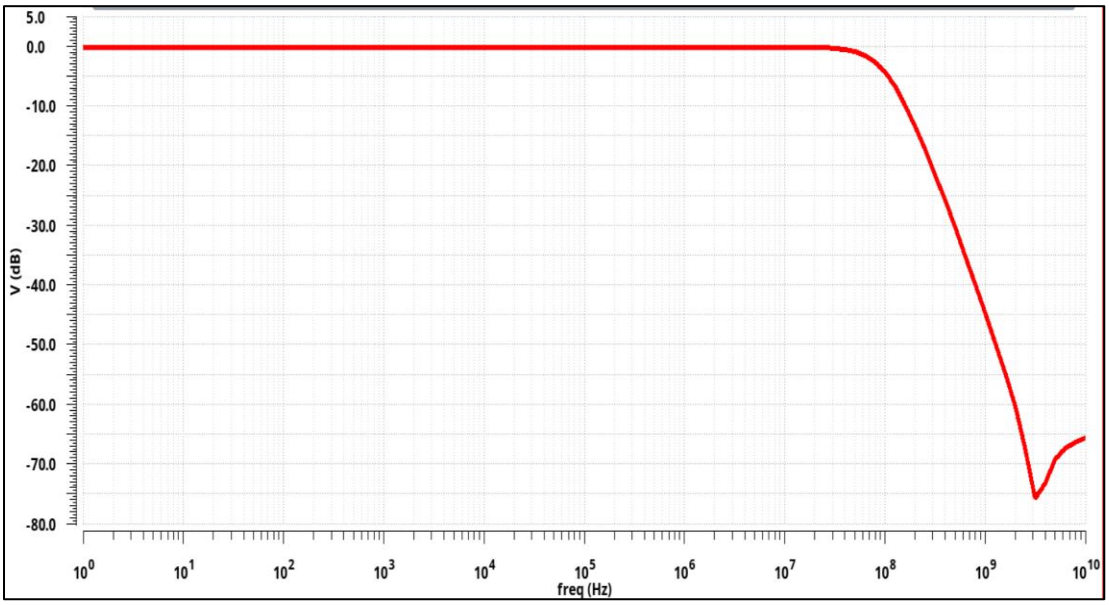


Fig. 3.16: Closed-loop unity-gain frequency response at the TT2 process corner

	Specifications	***SPICE* Results**
DC Gain (dB)	> 60	68
UGBW (MHz)	> 50	58.5
PM (°)	> 60	68
Input-referred noise voltage spectral density @ 10MHz (nV/√Hz)	< 15	13.3
C_L (pF)	5	5
C_C (pF)	-	1

Table 3.4. *SPICE* Simulation Results vs. Target Specifications at TT2 [15]

For high-volume production in industry applications, the final design should also meet the target specifications of Table 3.1 over wide ranges of CMOS process parameter values (i.e., slow (SS1, SS2, SS3), typical (TT1, TT2, TT3) and fast (FF1, FF2, FF3)), temperatures, - 40°C < T < 85°C, and power supply voltages, 1.7 V < V_{DD} < 1.8 V. Thus, extensive *SPICE* simulations are used to validate the final design over the nine "process corners" listed in Table 3.3. The final *SPICE* simulations are

summarized in Table. 3.5. MOSFET device performance is, of course, characterized by a fundamental gain/bandwidth tradeoff. Thus, it is interesting to note that the margin for UGBW is minimum for the slowest process corner SS1, whereas the margin for the DC gain is minimum for the fastest process corner FF1. The final g_m/I_D-based design exceeds the target specifications over all nine of the 180 nm CMOS process corners.

Process Corner	**DC gain (dB)**	**UGBW (MHz)**	**PM (°)**
Spec.	> 60	> 50	> 60
SS1	67.6	53.3	68
SS2	69	56.7	68.7
SS3	71.8	66.4	69
TT1	66.4	56	67
TT2	68	58.5	68
TT3	70.3	67.5	69
FF1	65.6	59	66
FF2	67	61	67
FF3	69	69	68.2

Table 3.5. Two-stage OTA Performance vs. 180 nm CMOS Process Corners [15]

Chapter 4: g_m/I_D-based Design of an Ultra-low-power CMOS Two-stage OTA with PVT-insensitive Nulling Resistor Frequency Compensation

4.1 Overview of MOSFET Operation in Weak Inversion

The two-stage OTA of Fig. 3.1 was a critical analog IC building block in the first CMOS PCM Filter chip introduced in 1980 [6][14]; it contributed to the paradigm shift from analog to digital telephony worldwide. More than four decades later, and after the production of tens of billions of units, it continues to play a key role in the majority of modern commercial mixed-signal/RF IC products. For a typical design of the OTA of Fig. 3.1, M_1-M_{11} are operated in the saturation region in moderate or strong inversion whereas M_{12} is operated in the triode region to realize the resistance required for frequency compensation. Based on "conventional wisdom" and for simplicity, V_{OV} = 0.1 V was chosen for all twelve MOSFETs in the design of Chapter 3.

The aim of this chapter is to extend the g_m/I_D-based synthesis techniques to the optimum design of a low-voltage ultra-low-power OTA as required for biomedical applications, for example. Rather than making an educated guess for V_{OV} as before, *SPICE* simulations are first used to plot a figure-of-merit defined as the product of g_m/I_D (power efficiency) times f_T (speed) vs. V_{OV} as shown in Fig. 4.1 [2]. Clearly, the choice for optimum speed/power performance is V_{OV} = 0.15 V in the moderate inversion region. In addition, tradeoffs are also considered for the operation of the M_1-M_2 input pair in weak inversion for ultra-low-power applications.

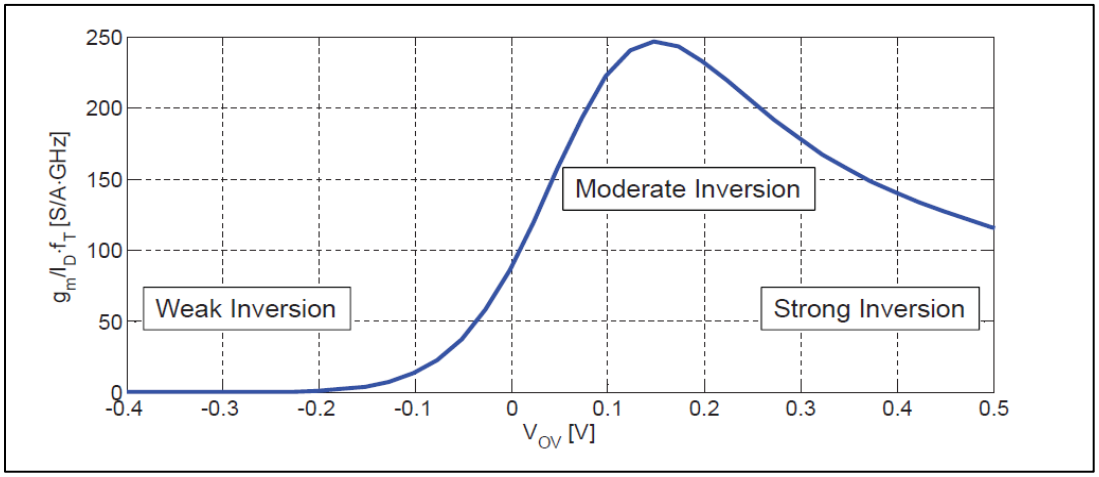


Fig. 4.1: g_m/I_D x f_T vs. V_{OV} [2]

As CMOS technology continues to scale to deep sub-micron minimum feature sizes with ever smaller power supply voltages, the minimum DC bias voltage headroom requirements for the classical two-stage OTA of Fig. 3.1 render it unattractive. Specifically, the DC bias circuitry requires $V_{DD} \geq (V_{GS\text{-}M10} + V_{GS\text{-}M11} + V_{SD\text{-}M9}) \approx (2V_{th} + 3V_{OV}) \approx 1.5$ V. This limitation is mitigated using the alternative PVT-insensitive Miller frequency compensation concept patented by Brehmer and Fisher in 1984 [18]. Although the idea was innovative, the proposed implementation in Fig. 1 of the patent is not correct because an NMOS k'-generator bias circuit was used with a PMOS common-source second gain stage—the k_n' and k_p' values do not cancel or track with PVT variations. The design shown in Fig. 4.2 with true PVT-insensitive Miller frequency compensation was described recently by Kuo, et al. [19], wherein the DC bias circuitry requires $V_{DD} \geq (V_{DS\text{-}M10} + V_{SG\text{-}M9}) \approx (V_{th} + 2V_{OV}) \approx 0.75$ V.

CMOS integrated circuits for biomedical sensors and other implanted devices not only need to be ultra-small for portability but also ultra-low power to ensure long lifetimes with physically small batteries and negligible heat damage to body tissue. Thus, weak inversion operation should be exploited when practical. The next section explores weak inversion operation for the M_1-M_2 input-pair in Fig. 3.1 or Fig. 4.2.

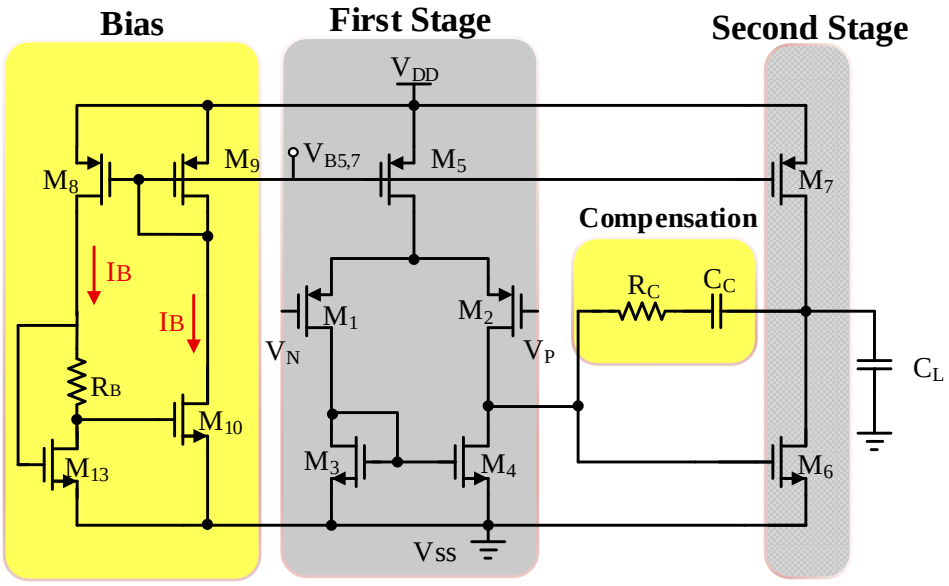


Fig. 4.2: CMOS two-stage OTA with low-voltage PVT-insensitive Miller frequency compensation (Fig. 3 in [19])

4.2 Comparison of g_m/I_D-based Designs as the Input Pair of the CMOS Two-stage OTA is Biased from Moderate to Weak Inversion

To investigate the design space for weak inversion operation, which is highly desirable for ultra-low-power biomedical applications, for example, g_m/I_D-based techniques are extended to redesign the two-stage OTA of Fig. 3.1 with the PMOS input-pair, M_1-M_2, biased from moderate to weak inversion with V_{OV} = 0.1 V, 0.05 V, 0 V, -0.05 V, -0.1 V, -0.15 V and -0.2 V, respectively As described above, all other devices are biased at V_{OV} = 0.15 V for all seven cases. As V_{OV} of M_1-M_2 is decreased, the corresponding PMOS g_m/I_D value increases monotonically for a given L as illustrated in Fig. 4.3. (Note: Same as Fig. 3.7 but with V_{OV} = 0.15 V). This study quantifies the tradeoff between power dissipation and chip area as V_{OV} for the input stage is varied from moderate to weak inversion.

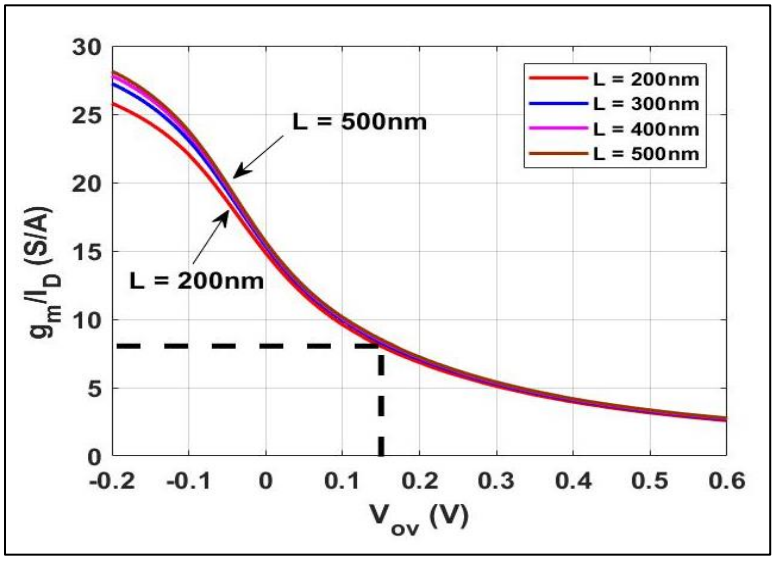


Fig. 4.3: PMOS g_m/I_D vs. V_{OV}. For all PMOS devices except the input pair, M_1-M_2, V_{OV} = 0.15 V and g_m/I_D = 8 S/A

The PMOS lengths (L) are determined from the voltage gain specification. As shown in Fig. 4.4 (Note: Same as Figs. 3.7 and 3.10 but with V_{OV} = 0.15 V), higher g_m/I_D values correspond to higher voltage gains for a given L. For L = 300 nm, for example, the voltage gain is minimum at g_m/I_D = 10 S/A where V_{OV} = 0.1 V and maximum at g_m/I_D = 27 S/A where V_{OV} = -0.2 V. Thus, the same lengths (L = 300 nm) are used for the PMOS input pair, M_1-M_2, for all seven designs. The target design

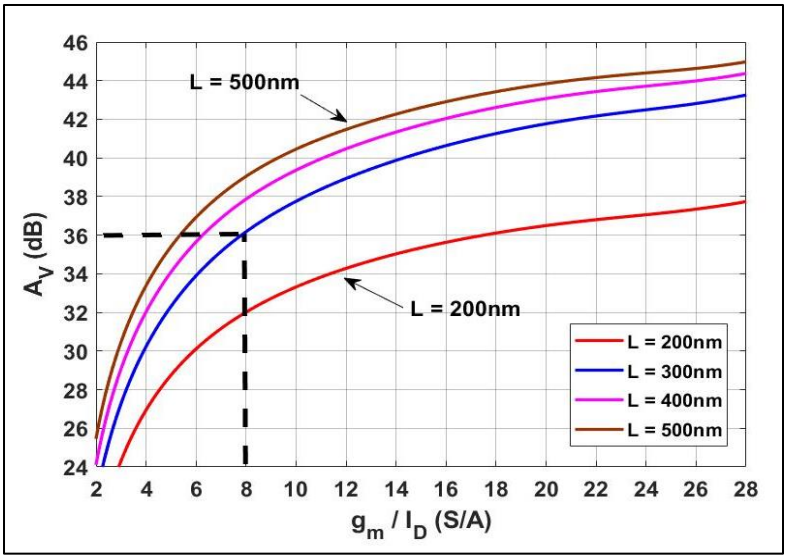


Fig. 4.4: Intrinsic PMOS small-signal voltage gain, A_{IP}, vs. g_m/I_D. A_{IP} = 36 dB for L= 300 nm and g_m/I_D = 8 S/A (Fig. 9 in [15])

specifications are identical to those in Table. 3.1. The NMOS devices with V_{OV} = 0.15 V correspond to g_m/I_D = 9 S/A as shown in Fig. 4.5 (Note: Same as Figs. 3.6 and 3.9 but with V_{OV} = 0.15 V).

The design flow follows the steps outlined in Chapter 3. Step 3 suggests $f_t \simeq$ 1.2 GHz for PM = 70°. Thus, from Fig. 4.6 (Note: Same as Fig. 3.5 but with V_{OV} = 0.15 V), the NMOS lengths should be less than 900 nm. Considering design margins, the NMOS transistors are designed with L = 700 nm in all seven designs.

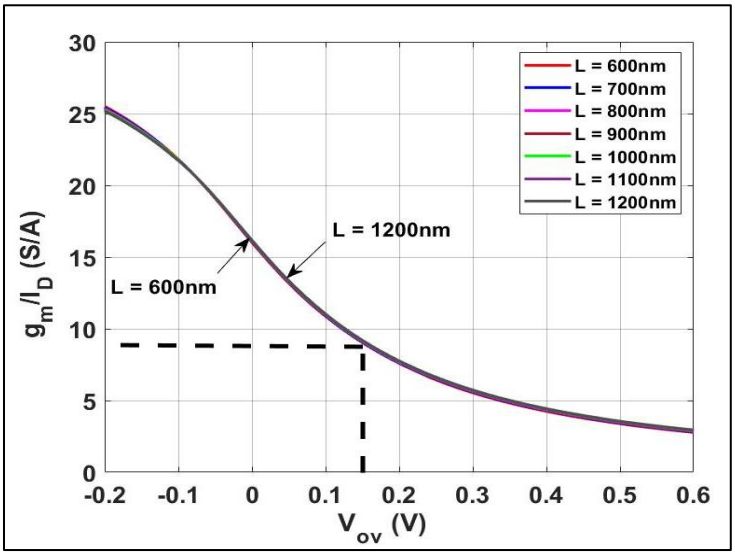


Fig. 4.5: NMOS g_m/I_D vs. V_{OV}. For all NMOSFETs, V_{OV} = 0.15 V and g_m/I_D = 9 S/A

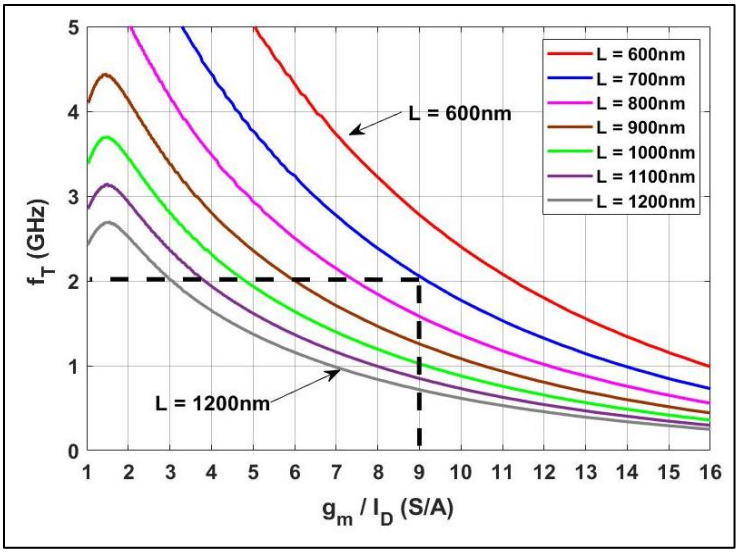


Fig. 4.6: NMOS f_T vs. g_m/I_D for a range of L values. $f_T \simeq 2$ GHz at $g_m/I_D = 9$ S/A and L = 700 nm

After the initial selection of the NMOS and PMOS device lengths, the total low-frequency voltage gain should be rechecked. First, from Fig. 4.4, A_{IP} = 36 dB, and from Fig. 3.9 with g_m/I_D = 9 S/A and L = 700 nm A_{IN} = 43 dB. Next, A_V is computed using (3.8) and (3.9) for the worst-case condition with V_{OV} = 0.1 V for the NMOS input pair, M_1-M_2. Hence, the corresponding small-signal voltage gain is $A_V = A_1A_2$ = (54.6 V/V)(45.1 V/V) = 2462.2 V/V = 67.8 dB, which is comfortably above the A_V = 60 dB target specification. However, the target unity-gain phase margin specification (PM = 60°) is unmet by surprisingly large amounts for deep subthreshold operation of M_1-M_2 with V_{OV} = -0.1 V, -0.15 V and -0.2 V. The UGBW target was also missed at V_{OV} = -0.2 V. The practical tradeoff between power dissipation versus closed-loop stability is detailed below.

4.3 *SPICE* Simulation Results for the g_m/I_D-based Designs as the Input Pair of the CMOS Two-stage OTA is Biased from Moderate to Weak Inversion

The step-by-step g_m/I_D design flow detailed in Chapter 3 was repeated to synthesize the seven two-stage OTA designs with PMOS input-pair V_{OV} values ranging from 0.1 V (moderate inversion) to -0.2 V (deep weak inversion). *SPICE* simulations were then used to determine the typical (TT2) performance parameters as summarized in Table 4.1.

Case	I	II	II	IV	V	VI	VII
V_{OV} of M_1-M_2 (V)	-0.2	-0.15	-0.1	-0.05	0	0.05	0.1
Gain (dB)	73	72.6	72	71	70	69	67
Unity-gain BW (MH_Z)	**29.3**	44	56.6	62	61.5	62	59
Phase Margin (°)	**14.6**	**33**	**55**	66	72.2	73.5	75.3
Input-referred noise @ 10 MHz (nV/√Hz)	10.6	10.6	10.7	10.8	11.4	11.7	13
First-stage Power Dissipation (mW)	0.06	0.07	0.07	0.09	0.1	0.13	0.15
Total Power Dissipation (mW)	0.24	0.25	0.25	0.26	0.28	0.31	0.33
C_C (pF)	1	1	1	1	1	1	1

Table 4.1. *SPICE* Simulation Results vs. V_{OV} of M_1-M_2

These astonishing results show extreme failures (highlighted in red) to meet the PM specifications for $V_{OV} < -0.1$ V. The main reason for the sharp reduction in PM is that the assumption used in Step 3 in Chapter 3 that $C_{gg\text{-}M6}$ is the dominant component of C_1 no longer holds as illustrated in Fig. 4.7. It is also clear from Fig. 4.7 that the power savings for $V_{OV} < -0.1$ V is relatively insignificant. More importantly, whereas $C_{dd\text{-}M4}$ and $C_{gg\text{-}M6}$ remain relatively constant as V_{OV} is decreased, $C_{dd\text{-}M1,2}$ becomes dominant for $V_{OV} < -0.1$ V with a concomitant dramatic increase in small-signal settling time as predicted from Fig. 3.8. The key conclusion is that if small-signal settling time of the OTA is a critical specification such as in some switched-capacitor circuit applications, for example, operation in deep weak inversion should be avoided.

Although Fig. 4.7 depicts a serious limitation for deep weak inversion OTA designs, some improvements in performance can be achieved by using non-standard layout techniques for critical devices. For example, $C_{dd\text{-}M1,2}$ comprises $C_{gd\text{-}M1,2}$ and $C_{db\text{-}M1,2}$. To a good approximation, $C_{gd\text{-}M1,2}$ is constant for all three regions of operation:

$$C_{gd\text{-}M1,2} = C_{overlap}\, W \tag{4.1}$$

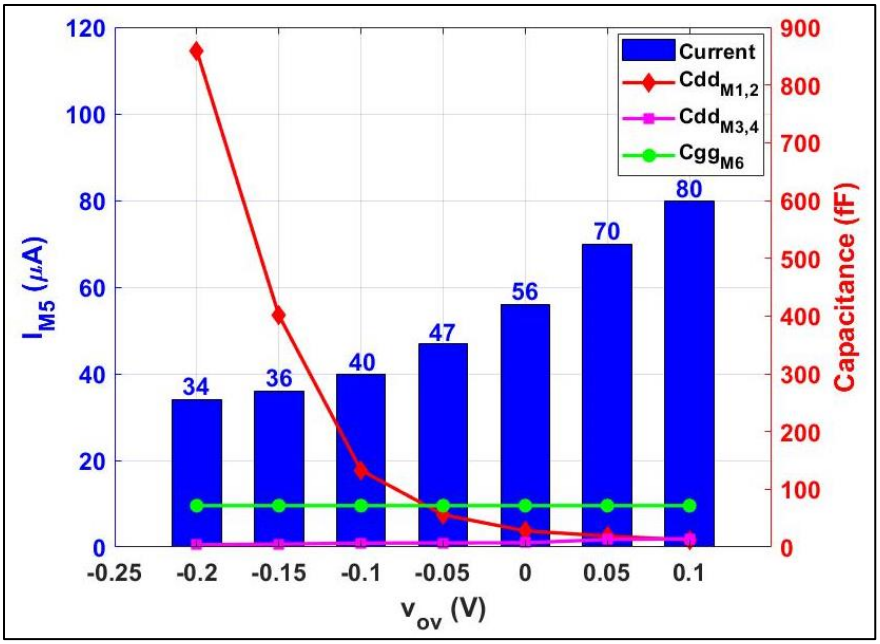


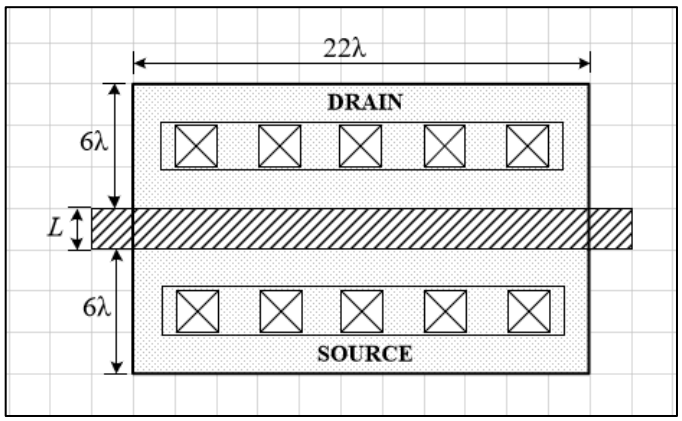
Fig. 4.7: DC bias current of the PMOS input pair and the parasitic capacitance components, $C_{dd\text{-}M1,2}$, $C_{dd\text{-}M3,4}$ and $C_{gg\text{-}M6}$, that comprise C_1 at the output of the first stage vs. V_{OV} values in strong, moderate and weak inversion [20]

whereas $C_{db\text{-}M1,2}$ is approximately given by [3]

$$C_{db\text{-}M1,2} = \frac{AD \cdot C_J}{(1+\frac{V_{DB}}{PB})^{MJ}} + \frac{PD \cdot C_{JSW}}{(1+\frac{V_{DB}}{PB})^{MJSW}} \tag{4.2}$$

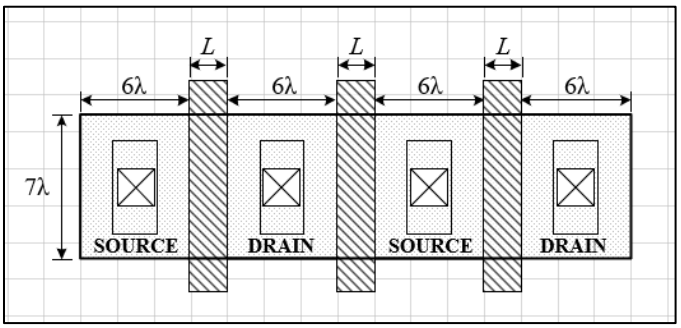
where AD and PD are the area and perimeter, respectively, of the drain diffusion. From (4.1), $C_{gd\text{-}M1,2}$ is proportional to the width of the PMOS input pair, which follows from the design value of V_{OV}. From (4.2), however, $C_{db\text{-}M1,2}$ depends on the AD and PD of the input transistors M_1-M_2. Thus, if AD and PD are reduced, not by design, but through layout, $C_{dd\text{-}M1,2}$ becomes a smaller contributor to the total parasitic capacitance, C_1. Based on this observation, the AD and PD values for three different M_1-M_2 transistor layouts with approximately equal W and L values are compared in Fig. 4.8 using universal λ layout rules with λ = 90 nm for a 180 nm CMOS process.

Although the area (AS) and perimeter of the source (PS) are both significantly increased compared to the conventional layout, the resulting large $C_{SS\text{-}M1,2}$ parasitic capacitance has virtually no effect on the performance of interest because it is connected to the common-source node of the input pair which is not in the signal path.



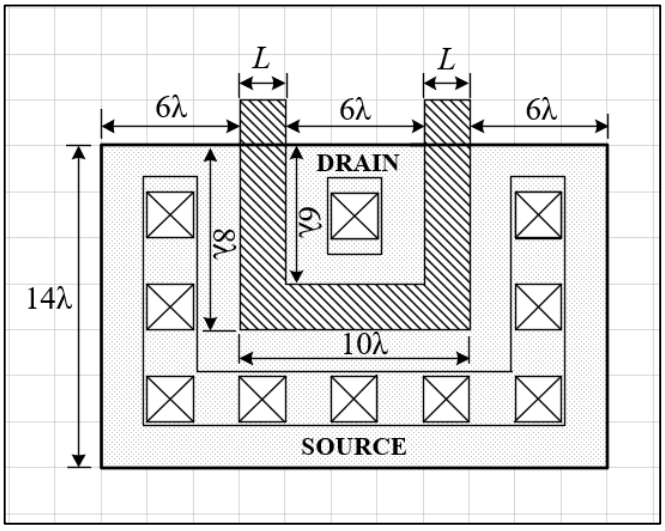
W = 22 λ and L = 2 λ; AD = 6 λ x 22 λ = 132 λ^2; PD = 2 x 6 λ + 22 λ = 34 λ

(a)



W = 7λ x 3 = 21 λ, L = 2 λ; AD = 2 x (6 λ x 7λ) = 84 λ^2; PD = 4 x 6 λ + 7 λ = 31 λ

(b)



W = 22 λ and L = 2 λ; AD = 6 λ x 6λ = 36 λ^2; PD = 6 λ

(c)

Fig. 4.8: Three different layouts: (a) Standard, (b) multi-finger, and (c) u-shaped

The standard, Fig. 4.8(a), and u-shaped, Fig. 4.8(c), layouts have W ≃ 22 λ, whereas W = 21 λ for the 3-finger layout in Fig. 4.8(b). L = 2 λ in all cases. Note that AD is reduced by nearly 4X from 132 λ^2 (standard) to 84 λ^2 (3-finger) to 36 λ^2 (u-shaped) across the three layouts. Similarly, PD is reduced by almost 6X from 34 λ (standard) to 31 λ (3-finger) to 6 λ (u-shaped). These simple, but non-standard, layout techniques do not cost any additional power dissipation or require any design modifications, but PM performance is improved significantly as shown in Table 4.2.

Table 4.2. Comparative PM Results for the Three Layouts (Fig. 4.8) of M_1-M_2

Vov of $M_{1\text{-}2}$ (V)	**-0.2**	**-0.15**	**-0.1**	**-0.05**	**0**	**0.05**	**0.1**
Gain (dB)	73	72.6	72	71	70	69	67
Unity-gain BW (MHz)	**29.3**	44	56.6	62	61.5	62	59
Phase margin (°)	**14.6**	**33**	**55**	66	72.2	73.5	75.3

STANDARD

Vov of $M_{1\text{-}2}$ (V)	**-0.2**	**-0.15**	**-0.1**	**-0.05**	**0**	**0.05**	**0.1**
Gain (dB)	73	72.7	72	71	70	68.4	67
Unity-gain BW (MHz)	33	47.5	58.6	63	60.6	61	58.3
Phase margin (°)	**14.5**	**36.5**	**58.7**	68.4	73.5	74.6	75.8

THREE-FINGER SEGMENTED

Vov of $M_{1\text{-}2}$ (V)	**-0.2**	**-0.15**	**-0.1**	**-0.05**	**0**	**0.05**	**0.1**
Gain (dB)	73	72.7	72	71	70	69	67
Unity-gain BW (MHz)	34.8	49	59.7	64.34	61.8	62.5	59
Phase margin (°)	**19**	**39.7**	**60.4**	69	73.6	74.5	75.8

U-SHAPED

Chapter 5: Conclusions and Future Work

As CMOS technology scales relentlessly to deep sub-micron minimum features sizes (e.g., < 10 nm) and uses three-dimensional *FINFET* devices, the conventional analog circuit design techniques based on simple square-law approximations that have persisted for more than four decades are no longer viable. More specifically, the discrepancies between expected performance parameter values in such designs and detailed *SPICE* simulations are approaching an order of magnitude.

The g_m/I_D-based design methodology, introduced by Silveira, et al. in 1996 [1], introduced a graphical design approach using *SPICE*-generated lookup tables intended to augment (and eventually replace) the conventional heuristic design approach based on increasingly more inaccurate square-law approximations. Only in recent years with the publication of a book on the subject by Jespers and Murmann in 2017 [2] has the approach gained in popularity in the industry.

One goal of this thesis was to present a clear and convincing g_m/I_D-based synthesis example of the most important CMOS two-stage OTA that has been in high-volume production around the world since 1980 [6]. That design uses a MOSFET operated in the triode region to realize the series frequency compensation resistance. Another major goal was to introduce and include a *SPICE*-generated lookup table for triode transistors into the g_m/I_D-based design flow. The performances of the synthesized circuit was in good agreement with *SPICE* simulations over the full practical range of CMOS process, voltage, and temperature variations.

It is well known that it is advantageous to operate some devices in weak inversion for ultra-low-power applications. The second major goal of this thesis was to investigate this second class of analog IC designs using the g_m/I_D-based design methodology for the same two-stage OTA. The results illuminated a severe tradeoff in power dissipation versus performance for these applications. Layout techniques were proposed that ease this tradeoff.

Future work will involve the extension of the g_m/I_D-based techniques to other classes of analog integrated circuits. One example is the synthesis of CMOS ring amplifiers that operate dynamically between strong, moderate and weak inversion. Another area of interest is the synthesis of CMOS wideband analog circuits.

Bibliography

[1] F. Silveira, D. Flandre, and P.G.A. Jespers, "A g_m/I_D based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA," *IEEE J. Solid-State Circuits,* vol. SC-31, pp. 1314-1319, Sep. 1996.

[2] P.G.A. Jespers and B. Murmann, *Systematic Design of Analog CMOS Circuits using Pre-Computed Lookup Tables*. Cambridge University Press, 2017.

[3] N. Paydavosi, et al., "BSIM 4.8.0 MOSFET Model User's Manual," Dept. of Electrical Engineering and Computer Sciences, Univ. of California, Berkeley, 2013.

[4] M.N. Sabry, H. Omran, and M. Dessouky, "Systematic design and optimization of an operational transconductance amplifier using the g_m/I_D design methodology," *Microelectronics Journal,* vol. 75, pp. 87-96, May 2018.

[5] T.B. Kumar, G.K. Sharma, A.K. Johar, D. Gupta, S.K. Kar, and D. Boolchandani, "Design automation of 5-T OTA using the g_m/I_D methodology," *IEEE Conf. on Information and Communication Technology,* 2019, pp. 1-5.

[6] W.C. Black, D.J. Allstot, and R.A. Reed, "A high-performance low-power CMOS channel filter," *IEEE J. Solid-State Circuits,* vol. SC-15, pp. 929-938, Dec. 1980.

[7] S. Clerc, T. Di Gilio, and A. Cathelin, *The Fourth Terminal: Benefits of Body-Biasing Techniques for FDSOI Circuits and Systems*, Springer, 2020.

[8] B.J. White, G.M. Jacobs, and G.F. Landsburg, "A monolithic dual tone multi-frequency receiver," *IEEE J. Solid-State Circuits,* vol. SC-14, pp. 991-997, Dec. 1979.

[9] B. Ahuja, "An improved frequency compensation technique for CMOS operational amplifiers," *IEEE J. Solid-State Circuits,* vol. SC-18, pp. 629-633, Dec. 1983.

[10] P.R. Gray and R.G. Meyer, "MOS operational amplifier design: A tutorial overview," *IEEE J. Solid-State Circuits,* vol. SC-17, pp. 969-982, Dec. 1982.

[11] G. Smarandoiu, D.A. Hodges, P.R. Gray, and G.F. Landsburg, “CMOS pulse-code-modulation voice codec,” *IEEE J. Solid-State Circuits,* vol. 13, pp. 504-510, Aug. 1978.

[12] D. Senderowicz, D.A. Hodges, and P.R. Gray, “High performance NMOS operational amplifier,” *IEEE J. Solid-State Circuits,* vol. SC-13, pp. 760-766, Dec. 1978.

[13] B.Y.T. Kamath, R.G. Meyer, and P.R. Gray, "Relationship between frequency response and settling time of operational amplifiers," *IEEE J. Solid-State Circuits,* vol. 9, pp. 347-352, Dec. 1974.

[14] National Semiconductor TP3040 PCM Monolithic Filter (1981). http://m.www.datasheets.pl/elementy_czynne/IC/TP3040.pdf

[15] C. Aueamnuay, A.V. Kayyil, N.B. Thota, P.K. Venkatachala, and D.J. Allstot, “g_m/I_D-based frequency compensation of CMOS two-stage operational amplifiers,” *IEEE Intl. Symp. on Circuits and Systems (ISCAS),* 2020.

[16] J. Mahattanakul and J. Chutichatuporn, “Design procedure for two-stage CMOS opamp with flexible noise-power balancing scheme,” *IEEE Trans, on Circuits and Systems,* vol. 52, pp. 1508-1514, Aug. 2005.

[17] H.C. Yang and D.J. Allstot, "Considerations for fast-settling operational amplifiers," *IEEE Trans. on Circuits and Systems,* vol. 37, pp. 326-334, March 1990.

[18] K.E. Brehmer and J.A. Fisher, “Compensated amplifier having pole-zero tracking,” U.S. Patent 4458212A, July 1984.

[19] Y.-W. Kuo, P.K. Ramakrishna, A.V. Kayyil, and D.J. Allstot, “Low-voltage tracking RC frequency compensation in two-stage operational amplifiers,” *IEEE Midwest Symp. on Circuits and Systems,* Aug. 2019, pp. 782-785.

[20] C. Aueamnuay, A.V. Kayyil, J. Liu, N.B. Thota, and D.J. Allstot, “g_m/I_D design considerations for subthreshold-based CMOS two-stage operational amplifiers,” *IEEE Intl. Symp. on Circuits and Systems (ISCAS),* 2020.